The Japanese Economic Review
The Journal of the Japanese Economic Association

The Japanese Economic Review doi: 10.1111/j.1468-5876.2009.00504.x
Vol. 61, No. 1, March 2010

THE EVOLUTION OF SPATIAL ECONOMICS: FROM THÜNEN TO THE NEW ECONOMIC GEOGRAPHY*

By MASAHISA FUJITA
Konan University

This paper presents a review of the evolution of spatial economics over the past two centuries. The focus is on the evolution of what I consider to be the most fundamental theory of spatial economics, i.e., *general location theory*. The paper starts with a review of Thünen (1826), and ends with a review of the New Economic Geography initiated by Paul Krugman in the early 1990s. It is shown that the study of general location theory has been successful at shedding light on many important features of actual spatial economies.
JEL Classification Numbers: F12, L13 and R12.

1. Introduction

Paul Krugman received the 2008 Nobel Prize in Economic Sciences for his seminal contribution to the development of New Trade Theory and New Economic Geography. This is great news for all scholars working in the field of spatial economics. In particular, as a person who has been studying spatial economics for my entire academic life, I am especially pleased with this news. Taking advantage of a great opportunity, this paper intends to review the evolution of spatial economics over the past two centuries, with special attention to the New Economic Geography (hereafter, NEG).

The field of economics that examines the questions of what types of economic activity take place where and why has traditionally been called *location theory* or *economic geography*. In addition, similar problems have been studied in several other branches of economics, each with specific context: *urban economics* focusing on cities; *international trade theory* treating national boundaries as impediments to the movement of goods and some factors of production; and *regional economics* relying on a somewhat artificial concept of regional boundaries. Formally speaking, *spatial economics* should include all branches of economics dealing with the analysis of economic processes and developments in geographical space. However, given space limitations, this paper focuses on the evolution of what I consider to be the most fundamental theory of spatial economics, that is, the "general theory of location and space-economy" using the terminology of Isard (1949), or for short, *general location theory*. As explained below, the NEG initiated by Paul Krugman in the early 1990s represents the newest wave in the development of general location theory.

According to Isard (1949, p. 505), "the general theory of location and space-economy is conceived as embracing the total spatial array of economic activities, with attention paid to the geographic distribution of inputs and outputs and the geographic variations in

* This paper is based on my presidential address delivered to the Japanese Economic Association Meeting at Kyoto University 7 June 2009. I am grateful to an anonymous referee, Marcus Berliant and Jacques Thisse for valuable comments on earlier versions of the paper. This research work has been supported by Grants Aid for Scientific Research A 21243021.

prices and costs." Or, in the words of Koopmans (1957, p. 153), in *general location models* "which are concerned with the location of all economic activities, this distribution [i.e., the geographical distribution of markets] becomes itself a variable" (bracketed explanation supplied by the present author). In summary, *a general location model aims, either in a descriptive or normative context, to explain the geographical distribution of all agents in a given location space, together with the associated spatial price system and trade pattern.* Depending on the context of a model, a "given location space" may represent, for example, a metropolitan area, a nation, a system of nations, or most ambitiously, the entire world. The potential importance of general location theory in spatial economics can be seen from the assertion by Ohlin (1933): "International trade theory cannot be understood except in relation to and as part of the general location theory, to which each of the mobility of goods and factors has equal relevance."

Few people would disagree with my opinion that the most challenging task in spatial economics is to develop general location models, which help us to understand important features of spatial economies (at various levels of spatial scale) in a unified manner. In developing such models, we have to confront new, highly complex problems brought about by the presence of space, in addition to all the difficulties of nonspatial general equilibrium models (or, multi-market models). For this reason, it is not surprising that the study of general location models, despite its long history, is still rather in its infancy, limited to several special classes of models. Nonetheless, it has been successful at shedding light on many important features of actual spatial economies such as: the formation of core-periphery spatial structures and income disparities within a country as well as among a system of nations; the "flying geese" pattern of industrial relocation within a country as well as among countries; the formation of various types of industrial agglomeration and specialized cities; the emergence of a hierarchical urban system in a country; and the formation of various types of specialized zones within cities. Furthermore, from the viewpoint of economic theory, general location models have the fundamental merit of enabling the synthesis of various old and new ideas developed by location theorists and economic geographers into a general equilibrium framework in a consistent and unified manner (Ottaviano and Thisse, 2005). This, in turn, helps to open new perspectives on spatial economics.

The history of spatial economics is very rich, but, on the other hand, it is somewhat perplexing. It is rich, indeed: starting with the monumental work of Thünen (1826), a variety of pioneering ideas have been developed periodically by great location theorists, geographers and economists such as Launhardt (1885), Marshall (1890), Weber (1909), Hotelling (1929), Ohlin (1933), Christaller (1933), Palander (1935), Kaldor (1935), Hoover (1936, 1937), Lösch (1940), and Isard (1949). Also, continuing in the second half of last century, many prominent economists have turned their attention to the subject at least in passing, as I shall explain later.

On the other hand, the history of spatial economics is somewhat perplexing. Despite its long and deep intellectual tradition, and despite the obvious importance of understanding the economic processes in the real world, spatial economics remained at the periphery of economic science until very recently. Even today, it is rare to find an economic text in which "space" is studied as an important subject. As suggested by Krugman (1995), this is perhaps due to the lack in the past of a unified framework, or of a comprehensive general location theory, that embraces both increasing returns and imperfect competition, the two basic ingredients of the formation of actual economic landscape. Furthermore, perhaps due to the lack in the past of comprehensive references on the subject, some economists who became interested in spatial economics recently seem to have developed the feeling that the field had been almost unexplored, ignoring the many valuable contributions introduced by early location theorists and economic geographers. This has sometimes caused unpleasant feelings among some scholars in different branches of spatial economics. If there is no shame in rediscovering science, there is a certain waste of resources (Thisse, 1996). The purpose of this review paper is therefore to put in perspective the evolution of general location theory over the past two centuries.

Moving backward historically, the NEG represents the most recent and most successful development of general location theory. As is well-known, the central topic of NEG has been how to explain the emergence of a *core-periphery structure* on a nationwide scale, or on an international scale. The hallmark of the NEG is a general equilibrium approach to the modelling of endogenous agglomeration forces generated through the three-way interactions among increasing returns, transport costs (broadly defined), and the movement of production factors. Thus, this paper discusses, of course, the development of the NEG since the early 1990s.

Jumping backward almost two centuries, the historic book *The Isolated State* (1826), by Johann Heinrich von Thünen, signified the birth of spatial economics. It also represents the oldest and, at the same time, the grandest attempt to develop a general location theory. Indeed, according to Paul Samuelson in his 1983 commemorative paper published on the occasion of the two-hundredth anniversary of Thünen's birth, "Among geographers and location theorists, Thünen is a founding God" (Samuelson, 1983, p. 1468). He then adds that, "Modern geographers claim Thünen. That is their right. But economists like me, who are not all that taken with location theory, hail Thünen as more than a location theorist. His theory is a theory of general equilibrium" (Samuelson, 1983, p. 1482).

As a founder of modern economics, Samuelson has naturally emphasized the originality of Thünen's contributions to general economic analysis, while somewhat playing down Thünen's legacy to location theory and economic geography. As a spatial economist, however, I would like to claim that Thünen's theory is a *theory of general equilibrium in space*, or for short, *a general location theory*.

As will be elaborated in Section 2, Thünen imagined an "isolated state" where a very large town is located at the centre of a homogeneous plain. He then attempted to determine simultaneously all variables of the economy through competitive markets of goods, labour and land, with a special focus on the land use pattern and land rent pattern in the agricultural hinterland. To achieve this grand objective, Thünen himself developed all the basic elements of modern competitive theory.

There is no doubt that Thünen's theory represents a great achievement in the history of spatial economics as well as in general economic theory. From the viewpoint of general location theory, however, a crucial question remains: in Thünen's entire theory, it is assumed a priori that all manufactured goods are to be supplied from the town. *What are, then, the agglomeration forces and economic mechanisms that keep all manufacturing activity in the town?*

As explained in the next section, it seems that Thünen himself was deeply concerned with this question. In retrospect, however, it was not possible to provide a satisfactory answer to this question at that time, for a satisfactory answer needed a *noncompetitive theory* of industrial agglomeration incorporating *increasing returns* in manufacturing production. The rest of the history of general location theory has evolved, and still has been evolving, around the consecutive challenges to finding a satisfactory answer to this question.

In full knowledge of the foundation that Thünen had laid, the next stage of general location theory centred on the development of *industrial location theory*, mostly by German scholars, in the late 19th century and the first half of the 20th century. With an explicit consideration of scale economies or indivisibilities in manufacturing production, they had developed industrial location theory together with *noncompetitive models of firms*. Given the difficulty of their task, it is understandable that the industrial location theory had been developed in a *partial equilibrium framework* in which the location of consumers is a priori given and fixed. In Section 4, I will focus on the pioneering theory of market areas developed by Lösch (1940), in which oligopolistic competition among firms producing a homogeneous product under increasing returns leads to the formation of hexagonal market areas.

The next stage of the development of general location theory, then, might be expected naturally to be the synthesis of Thünen's agricultural land use model based on competitive theory and industrial location theory based on oligopolistic competition. But that did not actually happen. Just when major contributions to industrial location theory appeared, next came the age in which neoclassical economics based on the competitive paradigm were at their zenith. In this period, between the early 1940s and the early 60s, the influence of neoclassical economics was so strong that little progress was made in developing new general location models based on noncompetitive theory.

In retrospect, however, this period served an important purpose for the future development of general location theory, involving an intensive soul search on the theoretical foundations of general location theory. That is, in this period, a long debate continued between two groups of economists about whether or not the neoclassical general equilibrium model based on perfect competition is comprehensive enough to fully reflect spatial dimensions. As will be elaborated in Section 4, on one side, the neoclassical general equilibrium school in the tradition of Walras, Pareto and Hicks maintained that a flexible application of the basic competitive theory can satisfactorily handle the problem of space. On the other side, the other school, led by Isard and supported by Koopmans and several others, asserted that in order to capture the essential impact of space on the distribution of economic activities, new models were needed that were fundamentally different from those found in standard general equilibrium theory based on perfect competition. In the protracted debate concerning the fundamental limitation of the competitive paradigm in spatial economics, the definitive answer, to my mind, was given by Starrett (1978), which essentially validated Isard's view, as will be explained in Section 4.

In the next stage, following the soul search regarding the foundations of general location theory, several successful attempts were made to formulate general location models in the context of urban morphology. Theoretically speaking, these urban models served as precursors to the NEG that appeared in the early 1990s.

Reflecting on what has been mentioned above, the rest of the paper is organized as follows: in Section 2, Thünen's theory of *The Isolated State* is discussed. Section 3 presents the discussion of industrial location theory developed in the first half of the last century, with a focus on Lösch (1940). Section 4 presents first the debates on the theoretical foundations of spatial economics between the neoclassical school and location theorists. Then, the *spatial impossibility theorem* of Starrett (1978) is introduced, and its implications are discussed. In Section 5, general location models of urban morphology are reviewed. Section 6 presents the development of the NEG. Finally, the paper concludes with the discussion of possible future directions for spatial economics.

2. Thünen's *The Isolated State*: The birth of spatial economics

If we want an example of how a great economist can use stark simplifications to get at the essence of an issue, it would be hard to beat the opening paragraph of *The Isolated State* published in 1826:

> IMAGINE a very large town, at the centre of fertile plain which is crossed by no navigable river or canal. Throughout the plain the soil is capable of cultivation and of the same fertility. Far from the town, the plain turns into an uncultivated wilderness which cuts off all communication between this State and the outside world.
>
> There are no other towns on the plain. The central town must therefore supply the rural areas with all manufactured products, and in return it will obtain all its provisions from the surrounding countryside.
>
> The mines that provide the State with salt and metals are near the central town which, as it is the only one, we shall in future call simply "the Town."[1]

From this beginning, as is well-known, Thünen developed his classic model of the joint determination of land use and land rent in the agricultural hinterland surrounding the Town. He supposed that crops differ in both yield per acre and their transport costs and allowed for the possibility that each crop could be produced with different intensities of cultivation. And he asked: How will the land be allocated if there is free competition among farmers and landowners with each individual acting according to his perceived self-interest?

Thünen showed that competition among the farmers will lead to a gradient of land rents that declines from a maximum at the Town to zero at the outermost limit of cultivation. Each farmer is faced with a trade-off between land rents and transport costs. As transport costs and yield differ among crops, a pattern of concentric rings of production will result. In equilibrium, the land-rent gradient must be such as to induce farmers to grow just enough of each crop to meet the demand, and this condition, together with the condition that rents be zero for the outermost farmer, fully determines the outcome.[2]

Figure 1 illustrates schematically the typical outcome of a Thünen model in the case of three agricultural goods, or crops. In the upper part of the figure, each straight line represents the *bid rent curve* of the corresponding crop,[3] the rent that farmers raising that crop would be willing to pay at each distance from the Town. As the competition among farmers makes the entire surplus from growing a crop at any distance accrue to landowners, each bid rent curve is equal to the surplus (net of all costs except land rent) per unit of land from growing that crop at each distance from the Town. The heavy line, the upper envelope of the bit rent curves, defines the *market rent curve* in equilibrium. As each location is occupied by the crop that offers the highest bid rent, concentric rings of cultivation emerge, as shown in the lower part of the figure.

---

[1] This quotation is from the English translation of Thünen (1826) by Waltenberg (1966, p. 7).

[2] For early contributions to a systematic treatment of Thünen's ideas in fully mathematical form, see Launhardt (1885, ch. 30), Lösch (1940, ch. 5) and Dunn (1954).

[3] The term *bid rent curve* was first introduced by Alonso (1964) in order to distinguish it from the market (or, realized) rent curve. In Thünen (1826), it is simply called *the surplus*.

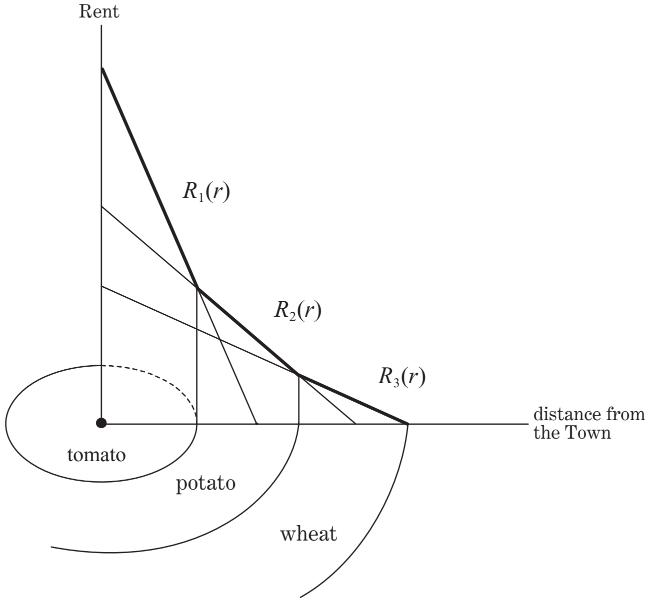

FIGURE 1. The land rent profile and Thünen rings with three crops

Thünen's model may now seem quite simple and obvious, but it actually embodies ingenious and profound analyses of the spatial economy (Fujita *et al.*, 1999, pp. 16–18). In particular, it is a striking example of the power of economic modelling to generate unexpected results. After all, determining which crops to grow where is not that easy. By allocating an acre of land near the Town to some crop, you indirectly affect the costs of delivering all other crops, because you force them to be grown further away. Furthermore, in Thünen's original model in *The Isolated State*, the wage of farmers at each distance from the Town is to be endogenously determined such that the utility of farmers, who consume crops grown in the field as well as goods manufactured at the Town, is the same everywhere.[4] Even in the case where there is no substitution of inputs in crop production, it is by no means trivial to determine which crops will be grown where. Yet Thünen's analysis shows us that there is a clear answer to this question: the spontaneous emergence of a concentric ring pattern. Indeed, the concentric rings will emerge even if no farmer knows what anyone else is growing, so that nobody is aware that the rings are there. Moreover, it turns out that this unplanned outcome is socially efficient.[5] It represents the working of Adam Smith's invisible hand at its best.

[4] Actually, in the first part of his analysis, Thünen assumed implicitly that wage rate was the same everywhere. In a later part, the wage rate becomes the variable at each location. This aspect of wage determination of farmers has been completely neglected in almost all subsequent variations of Thünen's model except in the recent general equilibrium analyses by Samuelson (1983) and Nerlove and Sadka (1991).

[5] That is, the total combined cost (excluding land rent) of producing and shipping the equilibrium consumption of crops at the Town is minimized.

When Thünen was just 20 years old, in his first paper, *Description of Agriculture in the Village of Gross-Flottbeck*, he had already glimpsed the idea of the so-called *Thünen rings*, an eternal gem of human intellect.[6] Thünen, however, was not an ordinary armchair scholar. Quite the opposite: he never occupied an academic position in his life. In fact, the initial idea of young Thünen had evolved to a grand theory of general equilibrium while he was working on his own Tellow estate (near Rostock in Germany), engaging in ceaseless agricultural improvement on his land. Thünen was satisfied with his abstract model only after undertaking laborious investigations of costs and returns on his Tellow estate over ten years, and then confirming that the collected data fitted directly into his model. It's no wonder that Schumpeter (1954, p. 466) called Thünen "one of the patron saints of econometrics." In short, Thünen's timeless model of agricultural land use and rent was "cultivated on land" literally while he was working as a farmer.

According to Samuelson (1983, p. 1468), Thünen "not only created *marginalism* and *managerial economics*, but also elaborated one of the first models of *general equilibrium* and did so in terms of realistic *econometric* parameters" (emphases by the original author). Specifically, Samuelson asserts that Thünen's model has in it elements of all of the following systems:

1 The Ricardo–Torrens theory of comparative advantage.
2 The Malthus–West–Ricardo theory of rent.
3 The Heckscher–Ohlin and Stolper–Samuelson theory of factors-and-goods pricing.
4 The Marx–Dmitriev–Leontief–Sraffa system of input-output. (p. 1481)

This is a praise so grand that it could be bestowed upon no other economist in history, except possibly Adam Smith. Indeed, Samuelson asserts that "Thünen belongs in the Pantheon with Léon Walras, John Stuart Mill, and Adam Smith" (p. 1482), applauding Thünen's work as a magnificent edifice of *general equilibrium* that contains all the basic elements of modern *competitive theory*.

This is not yet the end of the story, however. Economic geographers and location theorists always refer to Thünen's work in the context of the agricultural land use theory, but hardly in the context of agglomeration economies or city formation. Thus, one would be surprised to read Section 2.6 in Part II of Thünen (1826/1966), which contains the extracts of posthumous papers on location theory written between 1826 and 1842, and edited by Hermann Schumachen in 1863. Investigating whether industrial firms are better off located in major towns (especially in the capital), Thünen first asks for the reasons against the location of industries in the capital, or the centrifugal forces (using the terminology of the new economic geography) that operate against the establishment of a manufacturing industry in a metropolis. Although there is no room here to provide details, suffice it to say that Thünen's treatise is surprisingly comprehensive, including the impact of high land rents and high food prices on monetary wages in large cities.

Thünen also investigates in depth the centripetal forces of industrial agglomeration. His systematic inquiry discovers that the following factors, on the other hand, favour the location of industries in large towns:

[6] For this point, refer to the Introduction by Peter Hall in the English translation of Thünen (1826) by Waltenberg (1966, p. xiii).

> (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 economics 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. (Thünen, 1826/1966, pp. 287–90)

Although Thünen wrote these notes 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. In particular, observe that the combination of Thünen's agglomeration factors 1, 2 and 4 closely resembles the "basic story" in Krugman (1991b) on the emergence of a core-periphery structure. Furthermore, if we combine these factors with the last one (7), which is about interindustry linkages and technological spillovers, we receive another fundamental explanation for the emergence of industrial agglomeration.

We can, thus, conjecture that if Thünen's original theory of a monocentric spatial economy (based on perfect competition) were unified with his equally pioneering thinking about industrial agglomeration above, then it would become a typical model of general location theory. This, however, amounts to the equivalent of asking Leonardo da Vinci, upon one's first view of his amazing "flying machines" drawn in the late 15th century, to invent a real airplane back then. As we all know, it was only after the invention of powerful combustion engines in the early 20th century that the Wright brothers were able to actually and physically demonstrate the first successful flight in an engine-powered aircraft. Likewise, it was only after the appearance of the Dixit–Stiglitz model of monopolistic competition in 1977 that the development of general location theory gathered momentum. That is, the monopolistic competition model of Dixit–Stiglitz provided general location theory with a powerful combustion engine.

Before closing this section, I would like to note that despite his monumental contribution to economic thought, Thünen's ideas languished for more than a century without attracting widespread attention. Although it is not entirely clear why this happened, according to Isard (1949), this reflects the *Anglo-Saxon Bias* against spatial analysis, following Marshall's early opinion on "the influence of time being more fundamental than that of space" (Marshall, 1890/1920, chap. XV). In fact, given that even today there is no complete translation of Thünen (1826) into English, his work has not been easily accessible to non-German speaking people. Anyway, it is certainly true that, "The science of economics has suffered from the relative neglect of his methods during the nineteenth [*sic*] and early twentieth centuries" (Isard, 1949, p. 480).

3. Noncompetitive models of industrial location

Given that the work of Thünen (1826) was developed at the very beginning of the Industrial Revolution in Germany, its focus was naturally on agricultural activity. Meanwhile, the economic importance of manufacturing increased steadily over the second half of the 19th century, which naturally refocused the attention of economists to the location of industry. In the first half of the 20th century, industrial location theory was advanced significantly, mostly by Germans (more precisely, by German-writing scholars). In addition to many informal discussions on industrial agglomeration, several pioneering formal analyses of industrial location appeared. These formal analyses were presented in a *partial equilibrium framework*: under the given location (or spatial distribution) of consumers/markets, locations of plants/firms were studied by considering their indivisibility or scale economies. This is understandable, for the analytical tools of noncompetitive markets were quite limited in that period. Furthermore, similar analyses were often presented independently. This is partly because of the asymmetric information flow between the German-writing scholars and the English-writing academic world; the academic literature in English tended to be well-known throughout the world, whereas that in German mostly remained unknown outside of the German-writing world. Also, following Thünen, Germany remained the centre of academic study on spatial economics until Walter Isard started regional science in the USA in the early 1950s.

To my knowledge, the first systematic study of the so called *market area analyses* was presented by Launhardt (1885).[7] Given the fixed locations of two firms in a two-dimensional uniform space, Launhardt studied very systematically the boundary of the market areas of the two firms, A and B, that provided the same product at fixed mill prices, P_A and P_B, respectively. The transport rates associated with the two firms are not necessarily the same. Furthermore, the transport cost is not necessarily proportional to the distance from each firm.

More surprisingly, Launhardt (1885) also studied the *spatial price policy* of two oligopolistic firms, which turned out to be a special case of the more general analysis presented by Hotelling (1929) much later. Launhardt assumed that two firms, A and B, locate at each end of a linear segment and supply the same product. Consumers are distributed uniformly on the line, and each consumer purchases one unit of the product regardless of the price. Assuming that each firm chooses the profit-maximizing mill price given the mill price of their competitor, Launhardt obtained the equilibrium mill prices of the two firms. The result turned out to be the same, of course, as that of Hotelling (1929) when conditions are the same.[8] Unlike Hotelling (1929), however, Launhardt did not study the joint determination of location and mill price by the two oligopolistic firms. Still, the work of Launhardt (1885) was surprisingly advanced at his time. Unfortunately, however, the work of Launhardt (1885) was little-known until (and even after) Tord Palander (1935) introduced Launhardt's work together with Hotelling's model.

In England, meanwhile, Marshall (1890) presented his celebrated study on industrial agglomeration, in which he examined systematically the reasons for the concentration of

---

7 This part of my discussion of Launhardt (1885) is based on the English translation of Launhardt's work. See also Ekelund and Hébert (1999).

8 That is, when two firms locate at each end-point of the linear market, and transport cost and marginal production cost are the same for the two firms.

specialized industries in particular localities. According to Marshall (1890/1920, chap. X), externalities are crucial in the formation of economic agglomerations and generate something like a lock-in effect:

> 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 neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas (p. 225).

Despite its vagueness, the concept of so-called *Marshallian externalities* has been much used in the economics and regional science literature devoted to the location of economic activities because it captures the idea that an agglomeration is the outcome of a "snowball effect" in which a growing number of agents want to congregate to benefit from a larger diversity of activities and a higher degree of specialization. Marshall, however, did not provide the microeconomic mechanisms behind such externalities, which have not been fully developed even today.

In Germany, then, Weber (1909) developed a unique theory of industrial location, which greatly influenced the later development of industrial location theory in economic geography and regional science. He considered the location of a manufacturing plant that uses fixed amounts of several localized-inputs per unit output. The location of the market for the product, and that of each localized-input is exogenously given and fixed. Considering a uniform plain, he studied the optimal location of the plant, which minimizes the total transport cost per unit output. His analysis included the celebrated case of a *location-triangle* (formed by the market, and the supply-points of two inputs). He also showed that the location of a plant having the minimum transport cost can be obtained by a simple mechanical model.[9] Weber then extended his transport cost-oriented model to a more general situation in which labour cost also varies with location. In addition, independently of Marshall (1890), Weber extensively discussed the reasons for and mechanisms underlying the local agglomeration of industry (Weber, 1909, chap. V).

Undoubtedly, Weber (1909) presented a pioneering theory of industrial location. In retrospect, however, his theory was completely devoid of price analysis and market structure. That is, his theory did not consider any endogenous determination of the prices of inputs and outputs. Partly for this reason, his industrial location theory was not sufficiently appreciated by economists in later periods. Again, however, this represents another incident of the historical constraints on economic analysis. That is, at the time of Weber (1909), neither the noncompetitive theory of markets nor game theoretic approach to interactive behaviour was well-developed.

Meanwhile, models of noncompetitive firms had been developed successively during the 1920s and 1930s, mostly outside Germany, by pioneering economists such as Hotelling (1929), Chamberlin (1933), Robinson (1933) and Kaldor (1935). These noncompetitive economic models, in turn, induced the development of new industrial location theory

---

9 Palander (1935) shows that the same problem was already studied by Launhardt (1885), in which the same mechanical model was presented.

based on the noncompetitive behaviour of firms. In particular, the celebrated work of Hotelling (1929) on spatial competition encouraged Tord Palander (1935), a Swedish economist, to reconsider Weber's industrial location theory in the more general framework of spatial price theory. It is also surprising to know that Palander (1935) critically examined the work of Hotelling (1929), and derived different conclusions from his own analysis of the same Hotelling problem (with two firms, selling the same product, on a linear market).[10] In particular, contrary to Hotelling's assertion that spatial competition for consumers distributed on a line segment would make both firms tend to agglomerate at the centre of the market, Palander's own analysis (based on an informal two-stage game) indicates that competitors tend to keep a distance from each other in order to avoid severe price-competition. This new result by Palander suggested that a new noncompetitive theory was necessary in order to explain the phenomenon of industrial agglomeration. Unfortunately, however, the work of Palander (1935) was recognized by few scholars outside the German-speaking world because it was written in German.

Next, we turn to the study of August Lösch (1940), another great location theorist in Germany. In contrast to the work of Palander (1935), Lösch seems to have been influenced more by the monopolistic competition theory of Chamberlin (1933) (than the spatial competition theory of Hotelling) as well as by the general equilibrium approach by Léon Walras, Wassily Leontief, and Ohlin (1933). His monumental book, *Die räumliche Ordnung der Wirtschaft* [*The Economics of Location*], was published in 1940 when he was just 34 years old. As the title of the book suggests, Lösch clearly aimed for the development of a general location theory in this book.

After an intensive review of the development of spatial economics from Thünen (1826) to Chamberlin (1933) and Ohlin (1933), in Chapter 8, he presents a system of equations describing a full general equilibrium of all locations and prices. Being unsatisfied with the mere confirmation of having as many equations as unknowns, however, he next sets a more limited but more exciting task of examining the concrete spatial structure of the market area in a simple, abstract "economic region." To achieve this objective, in the tradition of Thünen, he uses stark simplifications to get at the essence of the issue, with the following opening paragraph of Chapter 9:

> Among all the factors that can create an economic region we shall select the economic. 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.
>
> In the following derivation we start from radical assumptions in order that no spatial differences may lie concealed in what we assume: that economic raw materials are evenly and adequately distributed over a wide plain. Our area shall be homogeneous in every other respect as well, and contain nothing but self-sufficient farms that are regularly distributed. How can this starting point lead to spatial differences?

In this context, first Lösch focuses on one industry in which each independent firm provides the same product (beer) using a single plant (brewery). As the operation of

---

[10] Palander found that when the two firms were located too close to each other (and hence each had a large market in its hinterland), there was no price equilibrium. Surprisingly, essentially the same conclusion was obtained almost half a century later by d'Aspremont *et al.* (1979), through a rigorous analysis based on noncooperative game theory.

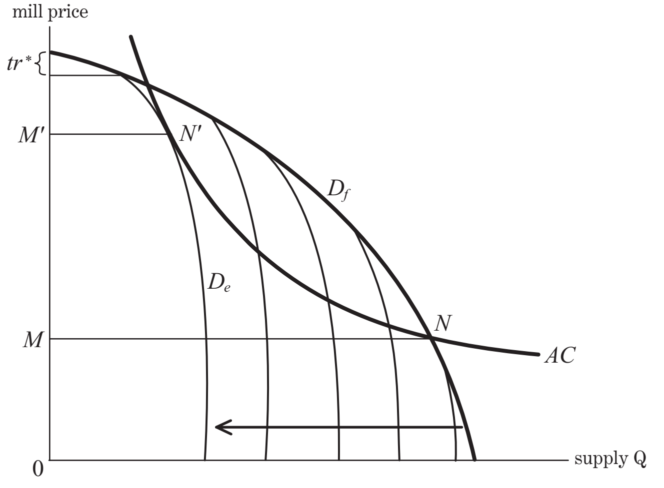

FIGURE 2. Shifting demand curves for each firm in association with the shrinking market areas

brewery is subject to scale economies, the average cost curve, *AE*, of each brewery is decreasing with the output (at least initially) as depicted in Figure 2.[11] All farmers (consumers) are identical and continuously distributed over the unlimited plain at a uniform density. Each consumer's demand for beer is exogenously given by a decreasing function of (delivered) price, which eventually becomes zero at a certain reservation price.

Consider one brewery, and suppose that it is *free of competition*. Assume also that transport cost per unit of beer is proportional to distance. Then, we can obtain the total demand for this brewery at each given mill price *P*, which is represented, in Figure 2, by the demand curve D_f in the case of the *free of competition*. As the average cost curve *AC* intersects the demand curve D_f at point *N*, this brewery can operate with a positive profit in the case of the free of competition.

Now, suppose that breweries are uniformly distributed over the plain such that each pair of neighbouring breweries has the same distance. Suppose further that the distance between each pair of neighbours is so large that when each firm sets its mill price at point *M* in Figure 2, the corresponding market area, which is a circle, does not overlap with any other circle. This situation is represented by Diagram (a) in Figure 3. In this situation, each firm can certainly get a positive profit by choosing the optimal mill price.

Next, we further decrease the distance between adjacent breweries so that the market areas of each pair of neighbouring breweries are just tangent to each other, as depicted by Diagram (b) in Figure 3. In this situation, each brewery is still essentially free of competition, and hence each can get a positive profit under the same optimal price. However, a small change occurs in the demand curve. That is, if each brewery charges the same mill price, which is lower than *M*, then the competition occurs, and hence the demand curve below the point *N* in Figure 2 shifts to the left.[12]

[11] In the following analysis, prices of all inputs are assumed to be exogenously given and fixed, and hence the average cost curve is also given.

[12] When studying the market competition among firms, Lösch implicitly employed the assumption of "price matching": when a firm chooses a certain mill price, all adjacent firms are assumed to choose the

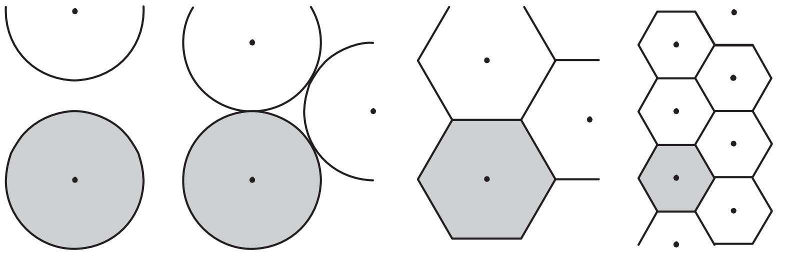

FIGURE 3. Development of market areas from the large circle to the final small hexagon (adopted from Lösch, 1940, fig. 23)

As we further increase the density of breweries, the corresponding demand curve of each brewery shifts further to the left, and the market area of each brewery (at the optimal mill price) becomes a regular hexagon as depicted in Diagram (c) of Figure 3. Continuing further, as represented by Diagram (d) in Figure 3 and the corresponding demand curve, D_e in Figure 2, when the density of breweries becomes sufficiently high, the demand curve of each brewery is just tangent to the average cost curve. Thus, at point N' in Figure 2, we have that *the marginal cost = the average cost = the marginal revenue* for each brewery, and the equilibrium is reached; each brewery has the same hexagonal market area, and the profit of each firm is just zero. This result represents the so-called *Löschian spatial equilibrium* of oligopolistic industry. This basic model of oligopolistic spatial equilibrium for a single industry represented a great contribution to the further development of general location theory in the following years.

In Chapter 11 of Lösch (1940), he attempted to extend this basic model by introducing several industries, and examined the system of market networks. From the viewpoint of economic theory, however, this part of Lösch's study is less satisfactory because no interdependence between markets of different goods was explicitly considered in terms of microeconomics. Given the complexity of the problem, however, it is understandable that Lösch was unable to provide a satisfactory economic analysis of spatial markets with multiple industries. Indeed, even today, there is no satisfactory model of general location analysis with multiple oligopolistic industries.

In connection with Lösch's analysis of market networks, I should mention the grand idea of *central place theory* originated by Christaller (1933). In considering an ideal system of supplying a dispersed population with "central goods" or public services, he developed a conceptual theory of central places. As in Lösch (1940), Christaller imagined a featureless plain, inhabited by an evenly spread population of farmers. They are to be served with a large number of central goods, which are ordered into groups in terms of their "*market ranges*." The market range of a central good is supposed to be determined by its transport cost and the degree of scale economies in production. Christaller argued, and produced evidence to support the theory that central places, or cities, form a *hierarchy*: the city at the top would produce all groups of urban goods and lower-order cities successively fewer groups of urban goods. And thus every textbook on location

same mill price. A more precise and detailed analysis of the interdependency between firms is provided by Beckmann (1972).

theory contains a picture of an idealized central place system in which a hierarchy of central places occupies a set of *nested hexagons*.

The basic ideas of central place theory seem powerfully intuitive. Indeed, the hierarchy image is so natural that it is hard to avoid describing things that way. Unfortunately, however, as soon as one begins to think hard about central place theory, one realizes that it does not quite hang together as an economic model. In economic modelling we try to show how a phenomenon *emerges* from the interaction of decisions by individual families or firms; the most satisfying models are those in which the emergent behaviour is most surprising given the "micro motives" of the players. What is therefore disappointing about central place theory is that it does not give any account along these lines. Christaller suggested the plausibility of a hierarchical structure; he did not give an account of how individual actions would produce such a hierarchy (or even sustain one once it had been somehow created). This is, however, understandable, for a satisfactory economic theory of central place theory means the development of a general location model of the entire urban system in an economy. The microeconomic foundations of the central place theory have not been fully developed even today.[13]

4. Soul searching on the foundations of spatial economics and the Spatial Impossibility Theorem

Although many economists would not have noticed, the debate about whether or not the neoclassical general equilibrium model based on perfect competition is comprehensive enough to fully reflect spatial dimensions has a long, interesting history.

On one side, the neoclassical general equilibrium school in the tradition of Walras, Pareto, and Hicks maintained that the problem of space can be handled simply by defining each commodity by its location as well as by its physical characteristics, and hence once we have made a clever indexing of commodities, we can essentially forget space in economic theory. As is well-known, this is the way Arrow and Debreu (1954) treated space (and time) in their seminal article.

On the other side, from the standpoint of the alternative view, supported by Kaldor, Lösch, Isard, Koopmans and several others, the problem is not that simple. To capture the essential impact of space on the distribution of economic activities, new models are needed that are fundamentally different from those found in standard general equilibrium based on perfect competition. In particular, more than half a century ago, Isard asserted forcefully his view in his seminal paper. The following insight is especially notable:

> Interdependence theory has an underlying premise the principle of pure competition. Yet, in no sense at all, can the traditional interpretation of this premise hold when we introduce space and thus transport costs into the analysis. If the various places in a region under consideration are treated as different markets (corresponding in this way to the varying local prices resulting from transport costs between these places) then the necessary condition of a large number of buyers and sellers for each commodity and factor at each market cannot be

[13] It may be noted, however, that Fujita, Krugman and Mori (1999) present a monopolistic competition model of urban system in which the economy tends to evolve a hierarchical structure reminiscent of Christaller (1933).

> fulfilled. If the region itself is viewed as one market, one could interpret the different prices ruling for a given commodity at the various places within the region (1) as signifying nonuniformity of product, or better yet, (2) as signifying a uniform product in a persistently imperfect market where individuals are in monopoly situations in accordance with the advantages of their respective positions. Neither case could be regarded as pure competition. (Isard, 1949, p. 505)

Once it is recognized that the competitive equilibrium paradigm cannot be the right foundation for the space-economy, what theory is conceivable? The following is Isard's second insight specifying a general theory of monopolistic competition as the alternative:

> And in a sense, too, because of the monopoly elements which are almost invariably present in spatial relations, a broadly defined general theory of monopolistic competition can be conceived as identical with the general theory of location and space-economy. (Isard, 1949, pp. 505–506)

It is important to note that when Isard refers here to "general theory of monopolistic competition," he had in mind not only the spatial oligopolistic competition theory along the lines of Hotelling (1929), Kaldor (1935), and Lösch (1940), but also Chamberlin's *The Theory of Monopolistic Competition* published in 1933. This can be seen clearly in the following statement: "Progress along Chamberlinian lines, however, is a sine qua non for developing further the theory of the space economy, especially in its welfare aspects" (Isard, 1949, p. 501). This is a remarkable insight, as if Isard could foresee the outbreak in the late 1980s of spatial economic models based on the Chamberlinian monopolistic competition approach.

Furthermore, Koopmans claimed in his *Three Essays on the State of Economic Science* that the vital effects of space become evident when our concern is the location of several economic activities and, hence, when the spatial distribution of activities itself becomes a variable. In this respect, Koopmans (1957, p. 154) maintained that "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."

A crucial theoretical result suggesting the fundamental limitation of the competitive paradigm in space-economy was first presented by Koopmans and Beckmann (1957) in the context of a *quadratic assignment problem*, which is defined by the following set of assumptions: each firm is indivisible, and the amount of land available at each location is such that a single firm can set up there. Hence every firm must be assigned to a single location, and every location can accommodate only one firm. Each firm produces a fixed amount of goods and uses one unit of land as well as fixed amounts of the goods produced by the others. Suppose further that the technology used by each firm is not affected by the chosen location. Finally, shipping a good from a location to another location involves a positive cost.

Although the issue addressed by Koopmans and Beckmann was the possibility of sustaining the optimal assignment through a competitive price system, we can see that any feasible assignment cannot be decentralized through a competitive price system. To illustrate the nature of the difficulties encountered, we restrict ourselves to the case of two firms, denoted $i = 1, 2$, and two locations, denoted $r = A, B$. Without loss of generality, we assume that firm 1 is assigned to A and firm 2 to B (refer to Figure 4). Firm i produces q_i units of good i and purchases q_j units of good j from the other firm j regardless of its own location. Firm i also receives a revenue $a_i > 0$ from other activities with the rest of the world, which does not depend on its location. Finally, each good i can be shipped from its place of production to the other location by a carrier at a given cost $t_i > 0$.

To study the sustainability of this assignment, we follow the suggestion of Arrow and Debreu by considering the same good at locations A and B as two different commodities, each with its own price. Let p_{ir} be the price of good i at location r and R_r be the rent to be paid by a firm for using one unit of land at location r. Firm 1's profit at location A is defined as follows:

$$\pi_{1A} = a_1 + p_{1A}q_1 - p_{2A}q_2 - R_A.$$

Likewise, Firm 2's profit at location B is given by

$$\pi_{2B} = a_2 + p_{2B}q_2 - p_{1B}q_1 - R_B.$$

If this price system sustains the foregoing configuration, then, as shown by Samuelson (1952), the equilibrium prices p_{ir} must satisfy the following conditions:

$$p_{1B} = p_{1A} + t_1 > p_{1A} \tag{1}$$

$$p_{2A} = p_{2B} + t_2 > p_{2B}. \tag{2}$$

In other words, the price of good 1 (2) in location $B(A)$ is equal to its price in location $A(B)$ plus the corresponding transport cost $t_1(t_2)$. This implies, as shown in Figure 4, that *for firm* 1 *located at A, the price of its output* (q_1) *is higher, and the price of its input* (q_2) *is lower at the other location, B; the converse is true for firm* 2 *located at B*. In this situation, if the land rent is the same at two locations (i.e., $R_1 = R_2$), then each firm (that takes all prices as given) desires to move to the other location. If $R_1 > R_2$ ($R_1 < R_2$), then at least firm 1 (firm 2) will desire to move to the other location. Hence, we can conclude that this location assignment cannot be sustained by any competitive price system.

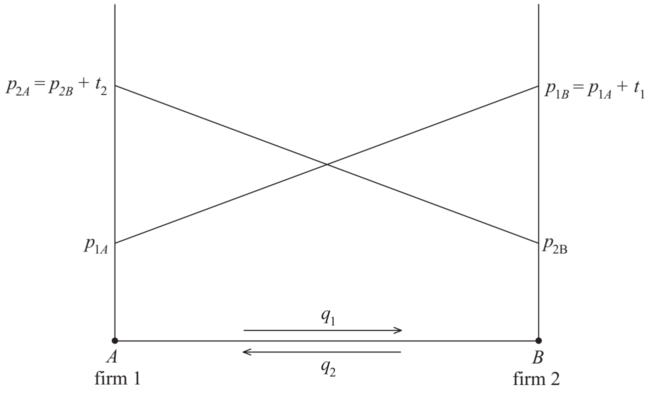

FIGURE 4. The price gradients supporting the trade induce firms to change their locations

To gain more insights about the nature and magnitude of difficulties, let us calculate the incentive for each firm to change its present location to the other location. First, if firm 1 sets up at location *B*, its profit is

$$\pi_{1B} = a_1 + p_{1B}q_1 - p_{2B}q_2 - R_B.$$

Hence, using (1) and (2), we can readily verify that

$$\pi_{1B} - \pi_{1A} = t_1q_1 + t_2q_2 + R_A - R_B, \tag{3}$$

which represents the incentive for firm 1 to move from location *A* to location *B*. Likewise, if firm 2 sets up at location *A*, its profit is

$$\pi_{2A} = a_2 + p_{2A}q_2 - p_{1A}q_1 - R_A.$$

Hence the incentive for firm 2 to move from location *B* to location *A* is given by

$$\pi_{2A} - \pi_{2B} = t_2q_2 + t_1q_1 + R_B - R_A. \tag{4}$$

Summing expressions (3) and (4) yields:

$$(\pi_{1B} - \pi_{1A}) + (\pi_{2A} - \pi_{2B}) = 2(t_1q_1 + t_2q_2).$$

This means that the sum of the firms' incentives to move is equal to twice the total transport costs. Because this sum must be nonpositive for a competitive equilibrium to exist, transportation cannot occur in such an equilibrium. This suggests that competitive pricing and positive transport costs are incompatible in a homogeneous spatial economy, for the incentive to change location is of the same order of magnitude as transport costs.

Koopmans and Beckmann conjectured that this negative result also holds in the general M-plant case, the proof of which was given much later by Heffley (1982). In particular, it can be shown that under any feasible assignment, the following relation holds:[14]

the aggregate net incentive of M firms to move to M locations
= M times the total transport costs associated with the original assignment. (5)

This implies that, under any given set of competitive prices, any feasible assignment is very far from equilibrium as long as the transport costs are positive.

In the protracted debate concerning the usefulness and/or limitation of competitive paradigm in spatial economics, the definitive answer was given by Starrett (1978), who shows that the negative result obtained by Koopmans and Beckmann (1957) holds in a perfectly general setting as long as the space is assumed to be homogeneous. Here, space is said to be *homogeneous* when all immobile resources are evenly distributed over the location space, and the utility functions of consumers and technology of firms are independent of location. In other words, consumers and producers have no intrinsic preferences for one place over others. Note that in order to determine whether a spatial model is able to explain the endogenous formation of economic agglomerations, the best approach is to consider the case of a homogeneous space in which economic agents are

---

[14] For a proof, see Fujita and Thisse (2002, Appendix to chap. 2).

free to choose their locations. For, if any concentration of economic activities is to occur, it has to be due to endogenous economic forces. As in Koopmans and Beckmann (1957), Starrett calculates the aggregate net incentive of all agents to move over all possibilities and obtains the following result for a homogeneous *M*-region economy, which is an extension of relation (5) to a general equilibrium context:

the aggregate net incentive to move = M times the total quasi-revenue of the transport sector under the original allocation. (6)

Here, "quasi-revenue" means that transport inputs are being evaluated at average prices over all regions.[15] Thus, we can conclude as follows:

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 and preferences are locally nonsatiated, then there is no competitive equilibrium involving costly transportation.*[16]

This is clearly a surprising result that requires more explanation. In a two-region economy, for example, when both regions are not in autarky, one should keep in mind that the price system must perform two different jobs simultaneously: (1) to support trade between regions (while clearing the markets in each region), and (2) to prevent firms and households from relocating. The spatial impossibility theorem says that, in the case of a homogeneous space, *it is impossible to hit two birds with one stone*: the price gradients supporting trade bear wrong signals from the viewpoint of locational stability. Indeed, if a set of goods is exported from *A* to *B*, then the associated positive price gradients induce producers located in region *A* (who seek a higher revenue) to relocate to region *B*, whereas region *B*'s buyers (who seek lower prices) want to relocate to *A*. Likewise, the export of another set of goods from *B* to *A* encourages such "cross-relocation." The land rent differential between the two regions can discourage the relocation in one direction only. Hence, as long as trade occurs at positive costs, some agents always want to relocate.[17]

The logical implications of this theorem are deep and far reaching. On the one hand, the Spatial Impossibility Theorem illuminates the fundamental limitation of the competitive paradigm in the study of spatial economies. Indeed, the Spatial Impossibility Theorem implies that when space is homogeneous and transportation is costly, the only possible competitive equilibrium is the so-called *backyard capitalism* in which every location operates as an autarky. In turn, this is possible only when production activities are perfectly divisible. This clearly shows the limits of the competitive paradigm for studying the main features of actual spatial economies. In other words, in the historical debate on the essence of spatial economies, Isard was right.

---

[15] Equation (6) above follows from equation (4.6) in Starrett (1978, p. 32) and using the assumption of uniformity of space.

[16] Fujita (1990, p. 213) dubbed this result by Starrett (1978) the Spatial Impossibility Theorem because it shows the inability of competitive models to explain the endogenous formation of economic agglomeration.

[17] For an intuitive explanation of this result using a standard diagrammatic approach, see chapter 2 in Fujita and Thisse (2002), which shows that the fundamental reason for the Spatial Impossibility Theorem is the *nonconvexity of the set of feasible allocations* caused by the existence of positive transport costs and the fact that agents have an address in space.

Note also that this theorem does not make any convexity assumption regarding preferences and technologies. Thus, this theorem is applicable, for example, to the general location models of Thünen's economy developed by Samuelson (1983), and by Nerlove and Sadka (1991), in a competitive framework. In their models, the wage rate is adjusted over the isolated economy such that no worker/farmer has a positive incentive to change his location. Thus, Equation (6) above implies that the aggregate net incentive of manufacturing firms (concentrated in the city) to move to the agricultural hinterland is very large. Therefore, unless some other unexplained strong forces exist to retain all firms in the city, Thünen's monocentric economy cannot be in competitive equilibrium.

On the other hand, the Spatial Impossibility Theorem is also very useful in delineating possible types of interesting general location models. That is, the Spatial Impossibility Theorem tells us that if we want to understand interesting phenomena about the spatial distribution of economic activities, in particular the formation of major economic agglomerations as well as regional specialization and trade, we must make at least one of the following three assumptions:

1 Space is heterogeneous.
2 Externalities in production and consumption exist.
3 Markets are imperfectly competitive.

In reality, of course, economic space is the outcome resulting from different combinations of these three agglomeration forces. However, it is convenient here to consider and evaluate their effects separately.

A Comparative advantage models. The heterogeneity of space introduces an uneven distribution of immobile resources (such as mineral deposits or some production factors) and amenities (climate), as well as the existence of transport nodes (ports, transshipment points) or trading places. This approach, while retaining the assumption of constant returns and perfect competition, yields comparative advantage among locations and gives rise to interregional and intercity trade.
B Externality models. Unlike models of comparative advantage, the basic forces for spatial agglomeration and trade are generated endogenously through nonmarket interactions among firms and/or households (knowledge spillovers, business communications, and social interactions). Again, this approach allows us to appeal to the constant return/perfect competition paradigm.
C Imperfect competition models. Firms are no longer price-takers, thus making their price policy dependent on the spatial distribution of consumers and firms. This generates some form of direct interdependence between firms and households that may produce agglomerations. However, it is useful to distinguish two types of approaches, namely:
  C1 Monopolistic competition. This leads to a departure from the competitive model and allows for firms to be price-setters and to produce differentiated goods under increasing returns; however, strategic interactions are weak because one assumes a continuum of firms.
  C2 Oligopolistic competition. Here, we face the integer aspect of location explicitly. That is, we assume a finite number of large agents (firms, local governments, and land developers) who interact strategically by accounting for their market power.

The implications of the modelling strategy selected are important. For example, models under A, B, and C1 permit the use of a continuous density approach that seems to be in line with what geographers do. By contrast, under C2, it is critical to know "who is where" and with whom the corresponding agent interacts. In addition, if we focus on class A models based on the heterogeneity of space, the market outcome is efficient. On the other hand, because the other two approaches involve market failures, the market outcome is likely to be inefficient.

The development of general location models, so far, naturally followed "the path of the minimum resistance." First, class A models were developed mostly in urban economics, such as Schweizer and Varaiya (1976, 1977), Schweizer *et al.* (1976) and Schweizer (1978). Then, class B models were developed, again in urban economics, as we shall explain in Section 5. Most recently, class C1 models have been developed extensively in the NEG as we shall explain in Section 6. On the other hand, due to technical difficulties, most of the class C2 models developed so far are partial equilibrium models, leaving the advancement of class C2 general location models mostly for the future.[18] The unification of the different classes of models and the eventual "balanced growth" of all four classes of models are highly anticipated in the future.

5. Urban morphology

Modern urban land use theory, which forms the core of urban economics, is essentially a revival of Thünen's theory (1826) of agricultural land use. Despite its monumental contribution to scientific thought, Thünen's theory languished for more than a century without attracting the widespread attention of economists. During that time, cities grew extensively and eventually outpaced the traditional concepts of urban design. The resulting rise in urban problems since the late 1950s has manifested an urgent need for a comprehensive theory of modern urban systems and, in particular, has helped to refocus the attention of location theorists and economists on the seminal work of Thünen. Following the pioneering work of Isard (1956), Beckmann (1957) and Wingo (1961), Alonso (1964) succeeded in generalizing Thünen's central concept of *bid rent curves* to an urban context.

As is well-known, the land use model of the *monocentric city* originated by Alonso (1964) is formally equivalent to the land use model of a monocentric economy constructed by Thünen (1826), in which the "isolated city" is replaced with the central business district (the CBD), whereas the agricultural land surrounding the city is replaced with residential land. Thus, in traditional monocentric urban models, the city is a priori assumed to be monocentric (that is, all production activities of the city are supposed to take place at the given centre, or CBD, which equally is supposed to be surrounded by the residential area of workers who commute to the city). This monocentric model of cities provided modern urban economics with a theoretical foundation, inspiring a great deal of theoretical and empirical work.[19] The main reason for this success is that the model is

[18] For an overview of class C2 models, see Anderson *et al.* (1992), and Fujita and Thisse (2002).

[19] For an overview of urban economic theory based on monocentric city models, see, for example, Fujita (1989). For more recent surveys of studies on urban spatial structure, see Anas, Arnott and Small (1998), Baumont and Huriot (2000), and Fujita and Thisse (2002).

compatible with the competitive paradigm. However, from the standpoint of both theory and reality, the monocentric urban model has drawbacks. From the standpoint of theoretical completeness, the monocentricity or non-monocentricity of a city is to be explained within the framework of the model, rather than being assumed a priori. Furthermore, from the standpoint of reality, many empirical studies of cities such as Mills (1972a,b) and Odland (1978) have shown the pervasive tendency of increasing decentralization of both firms and households and the consequent decline of the role of the CBD as the single focus of employment.

In response to the need for a more general theory of urban spaces, the development of general location models of urban morphology started in the early 1970s. Whether in a descriptive or normative context, such a model aims to explain the geographical distribution of all agents in a given urban area, without the a priori assumption of any centre. All early descriptive models belong, not surprisingly, to class B, in which agglomeration forces are generated through *nonmarket interaction* (externalities) such as social interactions and business communications. The development of class C1 models based on monopolistic competition started only in the late 1980s.

To the best of my knowledge, Solow and Vickrey (1971) represent the first economic model in which the land use pattern in an urban area is determined without the a priori assumption of any centre. They examined, somewhat surprisingly, the optimal and equilibrium allocations of business area between the generators of traffic (business firms) and the conveyor of traffic (roads). Each unit of business area is assumed to generate a fixed number of trips to every other unit. The potential business area is treated essentially as one-dimensional (*a long narrow city*). In the second half of the 1970s, the development of what is termed the general location model of cities advanced in earnest. Beckmann (1976) represents the first general location model of residential land use, in which the need for social communications generates agglomeration forces whereas, on the other hand, the desire for the consumption of more land in an area with lower land rents works as the dispersion force. When the residential land market of the city is in equilibrium, *the city exhibits a bell-shaped population distribution supported by a similarly shaped land rent curve*. Thus, the natural gregariousness of human beings leads to the spatial concentration of people within a compact area such that the density of population decreases with the distance from the endogenous centre.

Soon after Beckmann (1976), in two independent articles, Borukhov and Hochman (1977) and O'Hara (1977) have shown how the need for face-to-face communications may induce (office) firms to congregate and to form the CBD, even though clustering results in higher land rents. Furthermore, Borukhov and Hochman (1977) compare the socially optimal distribution of firms with the equilibrium, and show that *the optimal distribution is more concentrated near the centre than the equilibrium one*. This result may come as a surprise because the conventional wisdom is that market cities are too crowded near the centre. The reason for this surprising result is that individuals do not account for the *locational externality* they generate. That is, although the location of each individual directly affects the travel costs of all others who interact with that individual, each individual considers only his/her own travel costs in making his/her locational decision. Social optimality requires that individuals are to internalize such location externalities, leading to more concentration (than the equilibrium) of individuals near the centre, thereby reducing interaction costs for others.

The same conclusion that the optimal distribution is more concentrated than the equilibrium has been obtained in many subsequent studies based on class B models. It must be noted, however, that all such studies consider only positive externalities, neglecting possible negative externalities such as traffic congestion and environmental degradation. Thus, if both positive and negative externalities are considered, it is not a priori obvious whether the optimal distribution is to be more concentrated or more dispersed than the equilibrium.

All the models of urban morphology so far consider separately either consumers' agglomeration or firms' agglomeration. Hence, it is natural to study next how the interaction between both types of agents shapes the spatial structure of the entire city. The first such study of a two-sector model of urban general location has been presented independently in two papers, namely, Ogawa and Fujita (1980) and Imai (1982). Whereas Ogawa and Fujita (1980) focus on the equilibrium spatial structure, and Imai (1982) on the other hand concentrates on the optimal one, nevertheless both present essentially the same model. In their model, using the framework of Solow's long narrow city, the agglomeration force is generated through business interactions among firms (neglecting social interactions among households for simplicity), whereas the dispersion force arises from the commuting costs of workers. That is, all other things being equal, each firm prefers to locate close to other firms. On the other hand, the clustering of many firms into a single area increases the commuting distances for workers, leading to higher wages and land rent in the area surrounding the cluster. Consequently, the equilibrium distribution of firms and households/workers is determined as the balance between these two opposing forces.

In the model of Ogawa, Fujita and Imai, it is further assumed that each (business) firm interacts equiprobably with every other firm. The interaction cost between any two firms is proportional to the distance between them, while the commuting cost of each worker is proportional to the commuting distance. When commuting costs are high in comparison with the interaction costs among firms, the equilibrium involves a complete mix of business and residential activities. As the commuting cost falls, two business districts, which are themselves flanked by residential areas, are formed around the integrated section that shrinks. Eventually, *when commuting costs are low enough, the city becomes monocentric with the emergence of a single business district surrounded by two residential sections.* This seems to accord with what we have observed since the beginning of the technological revolution in transportation. Imai (1982) examines the optimal configuration of the city in the same model, and shows that, as in the preceding one-sector consumer models, firms tend to be less agglomerated in equilibrium than they should be at the optimum.

Although the city is essentially monocentric in the context of the linear interaction field described above, Fujita and Ogawa (1982) show that the possible equilibrium configurations of the city are much more complex in the case of a *nonlinear interaction field* in which the benefit of interactions between each pair of firms decreases negative-exponentially with the distance between them. In addition to the configurations just mentioned, certain combinations of the commuting cost parameter and the firms' interaction cost parameter may lead to a duocentric configuration, or to a configuration involving a primary business centre and two secondary business centres, or to three more or less identical centres. In the last three cases, the equilibrium configuration may also be viewed as describing a system of two or three cities. Furthermore, *multiple equilibria* may arise under the same set of parameters. Thus, when some parameters change slightly, the transition from one equilibrium to another may be catastrophic. All these results show how nonlinearities in the interaction field may lead to a set of vastly different outcomes and, by the same token, may explain why it is hard to make reliable predictions about urban development.

Thus far, a firm has been considered as a single-unit entity. In contrast, Ota and Fujita (1993) present a general location model of a city with multi-unit firms, where each firm consists of a front-unit (e.g. business office) and back-unit (e.g. plant or back-office). Each front-unit interacts with all other front-units for the purpose of business communications, while each back-unit exchanges information or management services only with the front-unit of the same firm. Each firm chooses the location of its front-unit and back-unit optimally. In this context, they show that the advancement of intrafirm communications technologies leads to the increasing segregation of back units from front units. Eventually, when intrafirm communication costs become sufficiently low, the most typical configuration involves the agglomeration of front units at the city centre surrounded by a residential area, whereas back units are established at the outskirts of the city together with their employees. In other words, a primary labour market emerges at the city centre (e.g. Manhattan), whereas secondary labour markets are created in the far suburbs. This result represents rather well the spatial configuration of large cities today such as New York, Tokyo, and Hong Kong with its hinterland in China.

Finally, in all models introduced thus far in this section, the agglomeration forces in the city are supposed to arise from nonmarket interactions (technological externalities) among agents. In contrast, Fujita (1988) represents the first general location model of class C1, which demonstrates that pure market interactions alone can generate spatial agglomeration of economic activities.[20] To this end, a spatial version of the Chamberlinian monopolistic competition model is developed, in which a continuum of firms (shops) supply a continuum of differentiated goods to homogeneous households in a linear city. Firms no longer assume that they can sell whatever they want at given market prices. Instead, each firm is aware that its optimal choice (location and price) depends on the demand for the variety it supplies. This demand itself rests on the spatial consumer distribution, thus showing how firms' choices are directly affected by consumers' choices. In turn, the optimal choice of a consumer (location and consumption) depends on the entire firm distribution. This is so because firms sell a differentiated product and consumers like variety; hence, their purchases are distributed across locations and the distribution of their shopping trips varies with the number of varieties available at each location. This creates mutual attractions between the two distributions: other things being equal, the centre of the distribution of consumers is the best location for each firm, whereas the centre of the distribution of firms is the best location for each consumer. Hence, this model can provide an explanation for the formation of the downtown or commercial areas in a city. Although this model focuses on the monopolistic competition on the product market, Fujita (1990) shows that the same principles can be applied to study how the availability of differentiated intermediate goods (such as producer services and financial services) may induce the agglomeration of exporting firms (such as high-tech firms or the headquarters of multinational firms) and the suppliers of intermediate goods, together forming the CBD.[21] These monopolistic competition

[20] See also Papageorgiou and Thisse (1985).

[21] In Abdel-Rahman and Fujita (1990), the role of differentiated intermediate goods in the formation of the CBD is studied under the assumption that all producers of such goods locate together in the CBD. The purpose of Fujita (1990) is to make this center endogenous.

models of urban morphology foreshadow the birth of the new economic geography based on monopolistic competition models in the 1990s, which is to be discussed in the next section.

6. The New Economic Geography

As noted previously, mainstream economists have paid, until very recently, remarkably little attention to economic geography, that is to say, the location of economic activity. However, since the early 1990s, there has been a renaissance of theoretical and empirical work on the spatial aspects of the economy. Among others, the pioneering work of Krugman (1991a,b) has triggered a new flow of interesting contributions to economic geography, culminating in the book by Fujita *et al.* (1999a). This "New Economic Geography (NEG)" has quickly emerged as one of the most exciting areas of contemporary economics.

The defining issue of the new economic geography is essentially the same as that of general location theory, i.e., how to explain the formation of a large variety of economic agglomeration in geographical space. The hallmark of the new economic geography is, however, the presentation of a unified approach—one that emphasizes the three-way interaction among increasing returns, transport costs (broadly defined), and the movement of production factors—in which *general equilibrium modelling* is combined with non-linear dynamics and the evolutionary approach to equilibrium selection. For example, Fujita *et al.* (1999a) demonstrate how the same modelling architecture can be used to explain the emergence and/or self-organization of either a city, a system of cities within a country, a core-periphery regional structure in a country, or industrial specialization and trade among countries.

All early models in the NEG belong to the general location models in class C1, using the modelling framework based on the Dixit–Stiglitz model of monopolistic competition. That is, in these models, agglomeration forces arise solely from pecuniary externalities through linkage effects among consumers and industries, neglecting technological externalities (such as information spillovers). Such a narrow focus of the early literature in the NEG was designed in order to establish a firm micro-foundation of the spatial economy based on modern tools of economic theory. It does not mean that the new economic geography is limited to such a narrow range of models and issues. Actually, its framework is widely open to further development as discussed in Fujita and Krugman (2004). Therefore, the main contribution of the NEG lies not in presenting specific forms of models but in developing a common vehicle equipped with a common "grammar" for studying a wide range of new issues.

Given that the state of the art in the literature of the NEG is well presented elsewhere,[22] selected pioneering contributions are briefly discussed below. For convenience, the NEG models below are divided into three groups: *regional models* (in which the economy is divided into a discrete set of regions), *international models* (in which the economy consists of a discrete set of nations), and *urban-system models* (in which the economy consists of a system of cities and their hinterlands).

[22] See, among others, Fujita *et al.* (1999), Fujita and Thisse (2002), Baldwin *et al.* (2003), Fujita and Krugman (2004), and for the most recent one, Krugman (2009).

6.1 Regional models

The seminal work of Krugman (1991a) has essentially defined the modern sphere of the NEG. The *core-periphery model* introduced in that paper has served like the $2 \times 2 \times 2$ models of textbook trade theory, providing a basic framework for the new economic geography—a general equilibrium framework that illustrates how the interactions among increasing returns at the level of the firm, transport costs and factor mobility can cause spatial economic structure to emerge and change.

There are two regions, two production sectors (agriculture and manufacturing), and two types of labour (farmers and workers). The manufacturing sector produces a continuum of varieties of a horizontally differentiated product; each variety is produced by a separate firm with scale economies, using labour as the only input. The agriculture sector produces a homogeneous good under constant returns, using farmers as the only input. Workers are freely mobile between regions, whereas farmers are immobile, distributed equally between the two regions. Finally, the agricultural good is costlessly traded between regions, whereas the interregional trade of manufactured goods involves a positive transport cost (in an iceberg form).

In this model, the immobility of farmers is a centrifugal force because they consume both types of goods. The centripetal force is generated through a circular causation of *forward linkages* (the incentive of workers to be close to the producers of consumer goods) and *backward linkages* (the incentive for producers to concentrate where the market is larger). If forward and backward linkages are strong enough to overcome the centrifugal force generated by immobile farmers, the economy will end up with a *core-periphery pattern* in which all manufacturing is concentrated in one region. It is shown that the core-periphery pattern is likely to occur in three cases: first, when the transport cost of the manufactures is low enough; second, when varieties are sufficiently differentiated; or third, when the expenditure on manufactured goods is large enough. Agglomeration need not occur, of course. However, a small change in critical parameters can "tip" the economy, from one in which two regions are symmetric and equal to one in which tiny initial advantages cumulate, turning one region into an industrial core and the other into a de-industrialized periphery. In other words, the dynamics of the model economy are subject to *catastrophic bifurcations*.[23]

6.2 International models

In the core-periphery models of Krugman (1991a), the centripetal force is generated through circular causation using forward and backward linkages between firms (producing differentiated consumer goods) and workers (who, in their other capacity as consumers, have a two-fold nature). For such a circular causation leading to agglomeration, the mobility of workers among locations/regions is indispensable. In contrast, Venables (1996) introduced two imperfectly competitive industries that are vertically linked through an input-output structure. This model has opened a new way to explaining the international concentration/specialization of industries (and the resulting international trade) through forward and backward linkages between the vertically linked industries in the absence of labour mobility.

---

[23] Observe, however, that the transition becomes smooth when workers are heterogeneous (Tabuchi and Thisse, 2002).

Krugman and Venables (1995) have introduced an international core-periphery model, which is slightly different from Venables (1996). Although the latter model involves two vertically linked industries, this model by Krugman and Venables involves a single monopolistically competitive manufacturing industry, which produces goods used *both for final consumption and as intermediates*. Intermediate usage creates cost and demand linkages between firms and a tendency for manufacturing agglomeration, again in the absence of labour mobility. Puga and Venables (1996) have further generalized this model by introducing multiple industries, each producing goods used both for final consumption and as intermediates to other industries. They demonstrate that this model with a general input-output structure can explain the spread of industries from country to country as an international region grows, resembling the phenomenon of the so-called "flying geese" pattern of the international spread of industries that occurred in East Asia over the second half of the last century.

All the models in the NEG introduced thus far are based on the same two crucial assumptions used by Krugman (1991a), i.e., the constant elasticity of substitution utility function and iceberg transport cost. Although these two assumptions lead to a convenient setting in which demands have a constant elasticity, they are nonetheless very specific. Thus, it is very important to know how the main results of the core-periphery model by Krugman (1991a) depend on such specificities of the framework. In this respect, Ottaviano, Tabuchi and Thisse (2002) represent an important contribution to the literature, in which they revisit the core-periphery model using an alternative framework that involves downward-sloping linear demands and linear transport cost measured in terms of the numéraire. They show that although the conclusions are not exactly the same as those derived by Krugman (1991a), this alternative model, called the *linear model*, also yields a core-periphery structure once transport costs are sufficiently low. Therefore, the core-periphery structure seems to be robust against very different formulations of preferences and transport technologies. Furthermore, the linear model has the advantage of yielding analytical solutions, and thus makes the study of welfare properties of the core-periphery structure easier. Most importantly, the linear model permits the study of different spatial price policies.[24]

Finally, Baldwin (1999) presents a dynamic model of agglomeration which is based on endogenous capital and perfect foresight. The production of any variety of differentiated product is assumed to require one unit of capital (which constitutes a fixed cost), whereas capital is produced in the competitive sector under constant returns using labour only. In this context, it is shown that, even without the mobility of labour and capital, the core-periphery structure can arise from the "expenditure-shifting" effect through capital formation (i.e., entry of new firms) in one region and capital decumulation (i.e., firm deaths) in the other region. The growth of the economy eventually stops in this model. However, if we interpret each unit of capital as representing the patent for a differentiated product, the framework of the model can be readily expanded to allow for endogenous growth. Indeed, such an extension is achieved in Baldwin *et al.* (2001) to study links between income divergence, industrialization and growth takeoffs.[25]

---

[24] As is well known, in the Krugman (1991a) model, mill pricing and discriminatory pricing yield the same result.

[25] Fujita and Thisse (2003) present a further extension of the Baldwin (1999) model to study the endogenous growth of a spatial economy under the presence of labor mobility.

6.3 Urban-system models

This subsection discusses some of the seminal papers on *urban systems* in the context of the NEG. In Section 5, we have been concerned with the internal spatial structure of a city or metropolis. In contrast, a model of an urban system concerns the distribution of sizes, types and locations of cities in an economy. Actually, in the past, economic analysis of urban systems has been conducted using two distinct frameworks. The traditional theory of an urban system, initiated by Vernon Henderson (1974, 1977), examines the distribution of sizes and types of cities, without concern for their locations. In other words, cities were treated as "floating islands." In contrast, the location of cities is a major concern in recent models of urban systems developed in the NEG.

The NEG offers two distinct approaches to the study of urban systems in space. One approach to the evolution of spatial structure has been offered by Krugman (1996). Keeping the basic framework of the core-periphery model in Krugman (1991a), he now imagines an economy consisting of a continuum of locations (or small regions) spread evenly around the circumference of a circle, where each location is endowed with the same number (density) of immobile farmers. In this context, using Turing's (1952) classic approach for morphogenesis in biology, he shows how the initial "Flat Earth" self-organizes, after a small random perturbation, into a regular central place system in which cities or locations of manufacturing are distributed evenly around the circle. The same approach has been used in Krugman and Venables (1997) to study the self-organization of spatial structure in a "seamless world" in which a continuum of small countries spreads evenly around the circumference of a circle.

An alternative, perhaps more directly realistic approach, has been followed in a series of papers by Fujita and Krugman (1995), Fujita and Mori (1997), Mori (1997), and Fujita *et al.* (1999b). They change the basic $2 \times 2 \times 2$ model of core-periphery. Instead of two regions, the location space is now described by the real line along which land is distributed uniformly. All workers in the economy are now assumed to be identical and free to choose their location and occupation. The agricultural good is now produced using both land and labour. Finally, transport costs are assumed to be positive for both the agricultural and industrial goods. In this model, only the agricultural land is immobile, which is the source of the centrifugal forces. In this context, Fujita and Krugman (1995) start with Thünen's "isolated state" (a city, defined as a concentration of manufacturing, surrounded by an agricultural hinterland), and obtain conditions under which no firm wishes to "defect" from the city to the agricultural hinterland.[26] It is shown, in particular, that the isolated state with a single city is an equilibrium when the population of the economy is small enough. Then, in the context of the same model, Fujita and Mori (1997) increase gradually the population of the economy as a whole. When the outer reaches of the hinterland eventually become sufficiently far from the centre, it becomes worthwhile for some manufacturing to "defect", giving rise to a new city. Further population growth gives rise to still more cities, and so on.

Key to this approach is the recognition that the attractiveness of any given location for manufacturing can be represented by an index of "market potential" derived from the underlying economics (Krugman, 1993; but the idea of market potential goes back to Harris, 1954, and this new work can be regarded as a justification of that approach). The

[26] To our knowledge, Fujita and Krugman (1995) represent the first version of the von Thünen model that simultaneously derives the existence of the central city and the pattern of land use.

process of change in the economy can then be regarded as involving a sort of co-evolution in which market potential determines where economic activity locates, and the shifting of that activity in turn redraws the map of market potential.

Like the Turing approach, this city-evolution approach ends up suggesting that despite the multiplicity of equilibria, there should be some predictable regularities in spatial structure. Once the number of cities has become sufficiently large, the size of and distance between cities tend to settle down at a roughly constant level determined by the relative strength of centripetal and centrifugal forces, providing some justification for the central place theory of Lösch (1940/1954). Fujita *et al.* (1999b) show further that when there are multiple industries that differ in terms of scale economies and/or transport costs, the economy tends to evolve a hierarchical structure reminiscent of Christaller (1933). Mori (1997) also shows that falling transport costs may eventually lead to the formation of a *megalopolis*, consisting of large core cities that are connected by an industrial belt (i.e., a continuum of small manufacturing cities). Thus, this line of work provides a link back to some of the older traditions in location theory and economic geography.

In all models of new economic geography thus far introduced, urban systems have been studied separately from international trade. Krugman and Livas Elizondo (1996) provide an important link between the two topics, opening a way to study the impact of external trade on internal geography.

7. Conclusion

We have reviewed the evolution of spatial economics from Thünen to the NEG with a focus on the development of general location theory. As we have seen, the NEG represents the newest and most successful development of general location theory. However, of course, the NEG does not represent the final phase of the evolution of general location theory.

In particular, in most models of the NEG developed so far, agglomeration forces arise solely from pecuniary externalities through linkage effects among consumers and industries, neglecting all other possible sources of agglomeration economies such as knowledge externalities and information spillovers. This has led to the opinion that the theories of the NEG have been too narrowly focused, ignoring as much of the reality as old trade theory. I fully understand the concern. But, first, let me defend our position. It is true that the theoretical framework of the NEG has been very narrowly focused. But, it was a deliberate choice. That is, such a narrow focus of the NEG was designed in order to establish a sound micro-foundation for geographical economics based on modern tools of economic theory. It does not necessarily mean that the NEG is limited to such a narrow range of models and issues. On the contrary, its framework is open wide to further development. Indeed, recently many such possibilities are being explored vigorously by a multitude of young scholars, as some of them have been reviewed in this paper.

Having said that, however, there still remains much room for further development of the NEG. In particular, there remains one type of agglomeration force for which micro-foundations have seen little development so far; that is, the linkages among people through the creation and transfer of knowledge, or in short, the K-linkages. (Hereafter, "knowledge" is defined broadly to include ideas and information.)

Traditionally, K-linkage effects have been called either "knowledge spillovers" or "knowledge externalities". However, the term "spillovers" tends to have a connotation of passive effects. And the term "externalities" tends to imply too many different things at once. So, in the remaining discussion, instead of knowledge spillovers or externalities, we use the term K-linkages in order to emphasize that they represent the agglomeration forces resulting from the activities related to both the "creation of knowledge" and the "transfer of knowledge" or "learning" (either in an active way or a passive way). In contrast to the K-linkages, the traditional linkages through production and transactions of (traditional) goods and services may be called the "E-linkages" (where "E" represents the economic activities in traditional economics).

Using such a terminology, we may imagine that the agglomeration forces in the real world arise from the dual effects of E-linkages and K-linkages. In this context, we conjecture that the role of K-linkages has been becoming increasingly more dominant recently. Yet, developing the micro-foundations of K-linkages seems to be the most challenging task, largely left for young scholars of the future.

We hastily add that there has been a great deal of conceptual studies of knowledge externalities/spillovers in a spatial context, starting with Marshall (1890), and including more recent pioneering work such as Jacobs (1969), Lucas (2001) in an urban context, and Porter (1998) in the context of industrial clusters.[27] Yet, it would be fair to say that there is a lot of room left for advancing the micro-foundations of K-linkages in space. Particularly, in developing the micro-foundations of K-linkages, "creation of knowledge" must be clearly distinguished from "transfer of knowledge" or "learning". Furthermore, for the creation of new ideas, cooperation among heterogeneous people is essential. Yet, through communication and migration, the degree of the heterogeneity of people in a region changes over time. Thus, the nature of K-linkages is essentially dynamic, and hence their full-fledged treatment requires a dynamic framework.

Recently, several pioneering works have appeared on the dynamic models of K-linkages, although they are mostly nonspatial. Among others, Jovanovic and Rob (1989, 1990), Jovanovic and Nyarko (1996), Auerswald *et al.* (2000), and Keely (2003) present micro-models of K-linkages. These studies are mostly concerned with the dynamics of the vertical differentiation of knowledge. In contrast, Berliant *et al.* (2003) and Berliant and Fujita (2008, 2009) are concerned with the dynamics of heterogeneous knowledge differentiation through the cooperative processes of knowledge creation and transfer.

Building upon such pioneering works, it is hoped that micro-foundations of K-linkages in space will be developed in the near future. Then, we may be able to develop a comprehensive theory of spatial economics, which integrates fully the dual effects of E-linkages and K-linkages in space. Using such a generalized framework, for example, we may be able to explore the economic implications of culture and diversity, as well as the cyclical migration of skilled workers on the interregional transfer of knowledge and skills as well as on the knowledge-heterogeneity within each region and between regions.

Final version accepted 24 September 2009.

---

27 See also Duranton and Puga (2004).

REFERENCES

Abdel-Rahman, H. and M. Fujita (1990) "Product Variety, Marshallian Externalities, and City Sizes", *Journal of Regional Science*, Vol. 30, pp. 165–183.

Alonso, W. (1964) *Location and Land Use*, Cambridge MA: Harvard University Press.

Anas, A., R. Arnott and K. A. Small (1998) "Urban Spatial Structure", *Journal of Economic Literature*, Vol. 36, pp. 1426–1464.

Anderson, S. P., A. de Palma and J. F. Thisse (1992) *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: MIT Press.

Arrow, K. and G. Debreu (1954) "Existence of An Equilibrium for A Competitive Economy", *Econometrica*, Vol. 22, pp. 265–290.

d'Aspremont, C., J. J. Gabszewicz and J. F. Thisse (1979) "On Hotelling's 'Stability in Competition' ", *Econometrica*, Vol. 47, pp. 1045–1050.

Auerswald, P., S. Kauffman, J. Lobo and K. Shell (2000) "The Production Recipes Approach to Modelling Technological Innovation: An Application to Learning by Doing", *Journal of Economic Dynamics and Control*, Vol. 24, pp. 389–450.

Baldwin, R. E. (1999) "Agglomeration and Endogenous Capital", *European Economic Review*, Vol. 43, pp. 253–280.

——, R. Forslid, P. Martin, G. Ottaviano and F. Robert-Nicoud (2003) *Economic Geography and Public Policy*, Princeton, NJ: Princeton University Press.

——, P. Martin and G. I. P. Ottaviano (2001) "Global Income Divergence, Trade and Industrialization: The Geography of Growth Take-Offs", *Journal of Economic Growth*, Vol. 6, pp. 5–37.

Baumont, C. and J. M. Huriot (2000) "Urban Economics in Retrospect: Continuity Or Change?", in J. M. Huriot and J. F. Thisse, eds, *Economics of Cities: Theoretical Perspectives*, Cambridge: Cambridge University Press, pp. 74–107.

Beckmann, M. J. (1957) *On the Distribution of Rent and Residential Density in Cities*, mimeo., Yale University.

—— (1972) "Spatial Cournot Oligopoly", *Papers of the Regional Science Association*, Vol. 28, pp. 37–47.

—— (1976) "Spatial Equilibrium in the Dispersed City", in Y. Y. Papageorgiou, ed., *Mathematical Land Use Theory*, Lexington, MA: Lexington Books, pp. 117–125.

Berliant, M. and M. Fujita (2008) "Knowledge Creation As A Square Dance on the Hilbert Cube", *International Economic Review*, Vol. 49, pp. 1251–1295.

—— and —— (2009) "Dynamics of Knowledge Creation and Transfer: The Two Person Case", *International Journal of Economic Theory*, Vol. 60, pp. 69–95.

——, R. Reed and P. Wang (2003) "Knowledge Exchange, Matching, and Agglomeration", *Journal of Urban Economics*, Vol. 5, No. 2, pp. 155–179.

Borukhov, E. and O. Hochman (1977) "Optimum and Market Equilibrium in A Model of A City without A Predetermined Center", *Environment and Planning A*, Vol. 9, pp. 849–856.

Chamberlin, E. (1933) *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press.

Christaller, W. (1933) *Die Zentralen Orte in Süddeutschland*, Jena: Gustav Fischer Verlag. Translated from German by C. W. Baskin (1966) *The Central Places of Southern Germany*, Englewood Cliffs, NJ: Prentice-Hall.

Dunn, E. S. (1954) "The Equilibrium of Land-Use Pattern in Agriculture", *Southern Economic Journal*, Vol. 21, pp. 173–187.

Duranton, G. and D. Puga (2004) "Micro-Foundations of Urban Increasing Returns: Theory", in J. V. Henderson and J. F. Thisse, eds, *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: North Holland, pp. 2063–2117.

Ekelund, R. B. Jr, and R. F. Hébert (1999) *Secret Origins of Modern Microeconomics. Dupuit and the Engineers*, Chicago, IL: The University of Chicago Press.

Fujita, M. (1988) "A Monopolistic Competition Model of Spatial Agglomeration: A Differentiated Product Approach", *Regional Science and Urban Economics*, Vol. 18, pp. 87–124.

—— (1989) *Urban Economic Theory: Land Use and City Size*, Cambridge: Cambridge University Press.

—— (1990) "Spatial Interactions and Agglomeration in Urban Economics", in M. Chatterji and R. E. Kunne, eds, *New Frontiers in Regional Science*, London: Macmillan Publishers, Chapter 12, pp. 184–221.

—— and P. Krugman (1995) "When Is the Economy Monocentric? Von Thünen and Chamberlin Unified", *Regional Science and Urban Economics*, Vol. 25, pp. 505–528.

—— and —— (2004) "The New Economic Geography: Past, Present and the Future", *Papers in Regional Science*, Vol. 83, No. 1, pp. 139–164.

——, —— and A. J. Venables (1999a) *The Spatial Economy: Cities, Regions and International Trade*, Cambridge MA: MIT Press.

——, —— and T. Mori (1999b) "On the Evolution of Hierarchical Urban Systems", *European Economic Review*, Vol. 43, pp. 209–251.

—— and T. Mori (1997) "Structural Stability and Evolution of Urban Systems", *Regional Science and Urban Economics*, Vol. 27, pp. 399–442.
—— and H. Ogawa (1982) "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations", *Regional Science and Urban Economics*, Vol. 12, pp. 161–196.
—— and J. F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*, Cambridge MA: Cambridge University Press.
—— and —— (2003) "Does Geographical Agglomeration Foster Economic Growth? And Who Gains and Loses from It?", *The Japanese Economic Review*, Vol. 54, No. 2, pp. 121–145.
Harris, C. (1954) "The Market as A Factor on the Localization of Industry in the United States", *Annals of the Association of American Geographers*, Vol. 64, pp. 315–348.
Heffley, D. R. (1982) "Competitive Equilibria and the Core of A Spatial Economy", *Journal of Regional Science*, Vol. 22, pp. 423–440.
Henderson, J. V. (1974) "The Sizes and Types of Cities", *American Economic Review*, Vol. 64, pp. 640–656.
—— (1977) *Economic Theory and the Cities*, New York: Academic Press.
Hoover, E. M. (1936) *Location Theory and the Shoe and Leather Industries*, Cambridge, MA: Harvard University Press.
—— (1937) "Spatial Price Discrimination", *Review of Economic Studies*, Vol. 4, pp. 182–191.
Hotelling, H. (1929) "Stability in Competition", *Economic Journal*, Vol. 39, pp. 41–57.
Imai, H. (1982) "CBD Hypothesis and Economies of Agglomeration", *Journal of Economic Theory*, Vol. 28, pp. 275–299.
Isard, W. (1949) "The General Theory of Location and Space-Economy", *Quarterly Journal of Economics*, Vol. 63, pp. 476–506.
—— (1956) *Location and Space-Economy*, Cambridge, MA: MIT Press.
Jacobs, J. (1969) *The Economy of Cities*, New York: Random House.
Jovanovic, B. and R. Rob (1989) "The Growth and Diffusion of Knowledge", *The Review of Economic Studies*, Vol. 56, pp. 569–582.
—— and —— (1990) "Long Waves and Short Waves: Growth through Intensive and Extensive Search", *Econometrica*, Vol. 58, pp. 1391–1409.
—— and Y. Nyarko (1996) "Learning by Doing and the Choice of Technology", *Econometrica*, Vol. 64, pp. 1299–1310.
Kaldor, N. (1935) "Market Imperfection and Excess Capacity", *Economica*, Vol. 2, pp. 35–50.
Keely, L. C. (2003) "Exchanging Good Ideas", *Journal of Economic Theory*, Vol. 111, pp. 192–213.
Koopmans, T. C. (1957) *Three Essays on the State of Economic Science*, New York: McGraw-Hill.
—— and M. J. Beckmann (1957) "Assignment Problems and the Location of Economic Activities", *Econometrica*, Vol. 25, pp. 1401–1414.
Krugman, P. (1991a) "Increasing Returns and Economic Geography", *Journal of Political Economy*, Vol. 99, pp. 483–499.
—— (1991b) *Geography and Trade*, Cambridge, MA: MIT Press.
—— (1993) "On the Number and Location of Cities", *European Economic Review*, Vol. 37, pp. 293–298.
—— (1995) *Development, Geography, and Economic Theory*, Cambridge, MA: MIT Press.
—— (1996) *The Self-Organizing Economy*, Oxford: Blackwell Publishers.
—— (2009) "The Increasing Returns Revolution in Trade and Geography", *American Economic Review*, Vol. 99, No. 3, pp. 561–571.
—— and R. Livas Elizondo (1996) "Trade Policy and the Third World Metropolis", *Journal of Development Economics*, Vol. 49, pp. 137–150.
—— and A. J. Venables (1995) "Globalization and the Inequality of Nations", *Quarterly Journal of Economics*, Vol. 110, No. 4, pp. 857–880.
—— and A. J. Venables (1997) *The Seamless World: A Spatial Model of International Specialization, Mimeograph*, Cambridge, MA: Massachusetts Institute of Technology.
Launhardt, W. (1885) *Mathematische Begründung Der Volkswirtschaftslehre*, Leipzig: B.G. Teubner. Translated from German by B.G. Teubner (1993) *Mathematical Principles of Economics*, Aldershot: Edward Elgar.
Lösch, A. (1940) *Die Räumliche Ordnung Der Wirtschaft*, Jena: Gustav Fischer. Translated from German by W. H. Woglom (1954) *The Economics of Location*, New Haven, CN: Yale University Press.
Lucas, R. E. (2001) "Externalities and Cities", *Review of Economic Dynamics*, Vol. 4, pp. 245–274.
Marshall, A. (1890) *Principles of Economics*, London: Macmillan, 8th edition published in 1920.
Mills, E. S. (1972a) *Studies in the Structure of the Urban Economy*, Baltimore: The Johns Hopkins Press.
—— (1972b) *Urban Economics*, Glenview, IL: Scott, Foresman and Company.
Mori, T. (1997) "A Model of Megalopolis Formation: the Maturing of City Systems", *Journal of Urban Economics*, Vol. 42, pp. 133–157.
Nerlove, M. L. and E. Sadka (1991) "Von Thünen's Model of the Dual Economy", *Journal of Economics*, Vol. 54, pp. 97–123.

O'Hara, D. J. (1977) "Location of Firms within A Square Central Business District", *Journal of Political Economy*, Vol. 85, pp. 1189–1207.
Odland, J. (1978) "The Condition for Multi-Center Cities", *Economic Geography*, Vol. 54, pp. 234–244.
Ogawa, H. and M. Fujita (1980) "Equilibrium Land Use Patterns in A Non-Monocentric City", *Journal of Regional Science*, Vol. 20, pp. 455–475.
Ohlin, B. (1933) *Interregional and International Trade*, Cambridge, MA: Harvard University Press, Revised version published in 1968.
Ota, M. and M. Fujita (1993) "Communication Technologies and Spatial Organization of Multi-Unit Firms in Metropolitan Areas", *Regional Science and Urban Economics*, Vol. 23, pp. 695–729.
Ottaviano, G. I. P., T. Tabuchi and J. F. Thisse (2002) "Agglomeration and Trade Revisited", *International Economic Review*, Vol. 43, pp. 409–435.
—— and J. F. Thisse (2005) "New Economic Geography: What about the N?", *Environment and Planning A*, Vol. 37, pp. 1707–1725.
Palander, T. (1935) *Beiträge Zur Standortstheorie*, Uppsala: Almqvist & Wiksells Boktryckeri AB.
Papageorgiou, Y. Y. and J-F. Thisse (1985) "Agglomeration as Spatial Interdependence between Firms and Households", *Journal of Economic Theory*, Vol. 37, pp. 19–31.
Porter, M. E. (1998) *On Competition*, Cambridge, MA: Harvard Business School Press.
Puga, D. and A. J. Venables (1996) "The Spread of Industry; Spatial Agglomeration and Economic Development", *Journal of the Japanese and International Economies*, Vol. 10, No. 4, pp. 440–464.
Robinson, J. (1933) *The Economics of Imperfect Competition*, London: Macmillan.
Samuelson, P. A. (1952) "Spatial Price Equilibrium and Linear Programming", *American Economic Review*, Vol. 42, pp. 283–303.
—— (1983) "Thünen at Two Hundred", *Journal of Economic Literature*, Vol. 21, pp. 1468–1488.
Schumpeter, J. A. (1954) *History of Economic Analysis*, New York: Oxford University Press.
Schweizer, U. (1978) "A Spatial Version of the Nonsubstitution Theorem", *Journal of Economic Theory*, Vol. 19, pp. 307–320.
—— and P. V. Varaiya (1976) "The Spatial Structure of Production with A Leontief Technology", *Regional Science and Urban Economics*, Vol. 6, pp. 231–251.
—— and —— (1977) "The Spatial Structure of Production with A Leontief Technology II: Substitute Techniques", *Regional Science and Urban Economics*, Vol. 7, pp. 293–320.
——, —— and J. Hartwick (1976) "General Equilibrium and Location Theory", *Journal of Urban Economics*, Vol. 3, pp. 285–303.
Solow, R. M. and W. S. Vickrey (1971) "Land Use in A Long Narrow City", *Journal of Economic Theory*, Vol. 3, pp. 430–447.
Starrett, D. (1978) "Market Allocations of Location Choice in A Model with Free Mobility", *Journal of Economic Theory*, Vol. 9, pp. 418–448.
Tabuchi, T. and J. F. Thisse (2002) "Taste Heterogeneity, Labor Mobility and Economic Geography", *Journal of Development Economics*, Vol. 69, pp. 155–177.
Thisse, J. F. (1996) "Introduction", in J. F. Thisse, K. J. Button and P. Nijkomp, eds, *Location Theory*, Glos: Edward Elgar Publishing, pp. xvii–xxxii.
Thünen, J. H. (1826) *Der Isolierte Staat in Beziehung Auf Landwirtschaft Und Nationalökonomie*, Hamburg: Perthes. Translated from German by C. M. Wartenberg (1966) *Von Thünen's Isolated State*, Oxford: Pergamon Press.
Turing, A. (1952) "The Chemical Basis of Morphogenesis", *Philosophical Transactions of the Royal Society of London*, Vol. 237, pp. 37–72.
Venables, A. J. (1996) "Equilibrium Locations of Vertically Linked Industries", *International Economic Review*, Vol. 37, pp. 341–359.
Wartenberg, C.M. (1966) *Von Thünen's Isolated State*, Oxford: Pergamon Press.
Weber, A. (1909) *Über Den Standort Der Industrien*, Tübingen: J.C.B. Mohr. Translated from German by C. J. Friedrich (1929) *The Theory of the Location of Industries*, Chicago: Chicago University Press.
Wingo, L. Jr (1961) *Transportation and Urban Land*, Washington, DC: Resources for the Future.