JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES **10,** 339–378 (1996)
ARTICLE NO. 0021

Economics of Agglomeration*

MASAHISA FUJITA

Kyoto University and University of Pennsylvania

AND

JACQUES-FRANÇOIS THISSE

CORE, Université Catholique de Louvain and CERAS–ENPC (URA 2036, CNRS)

Received January 11, 1996; revised August 20, 1996

Fujita, M., and Thisse, J.-F.—Economics of Agglomeration

We address the fundamental question arising in geographical economics: why do economic activities agglomerate in a small number of places? The main reasons for the formation of economic clusters involving firms and/or households are analyzed: (i) externalities under perfect competition; (ii) increasing returns under monopolistic competition; and (iii) spatial competition under strategic interaction. We review what has been accomplished in these three domains and identify a few general principles governing the organization of economic space. A few alternative, new approaches are also proposed. *J. Japan. Int. Econ.*, December 1996, **10**(4), pp. 339–378. Kyoto University and University of Pennsylvania; and CORE, Université Catholique de Louvain and CERAS–ENPC (URA 2036, CNRS). © 1996 Academic Press, Inc.

Journal of Economic Literature Classification Numbers F12, L13, R12.

1. INTRODUCTION

''Nearly half the world's population and three-quarters of all westerners live in cities'' (*The Economist*, July 29, 1995). This mere, crude fact can no

* Paper prepared for the Trilateral TCER/NBER/CEPR Conference on ''Economic Agglomeration,'' Tokyo, January 11–12, 1996. The authors are grateful to Simon Anderson and Vernon Henderson for helpful discussions during the preparation of this article. They also thank Gilles Duranton, Louis-André Gérard-Varet, Jean-Marie Huriot, Yoshitsugu Kanemoto, Xavier Martinez-Giralt, Dominique Peeters, Diego Puga, Tony Smith, Tetsushi Sonobe Takatoshi Tabuchi, and one referee for useful comments. The extended version of this paper has been published as a CEPR discussion paper.

0889-1583/96 $18.00

longer be put aside. We are therefore led to raise the following, fundamental question: *why do economic activities tend to agglomerate in a small number of places* (*typically cities*)?

More precisely, we want to explain why some particular economic activities choose to establish themselves in some particular places, and what is the resulting geographical organization of the economy. Intuitively the equilibrium spatial configuration of economic activities can be viewed as the outcome of a process involving two opposing types of forces, that is, *agglomeration* (or centripetal) *forces* and *dispersion* (or centrifugal) *forces*. This view agrees with very early work in economic geography. For example, in his *Principes de Géographie humaine* published in 1921, the famous French geographer Vidal de la Blache argues that all societies, rudimentary or developed, face the same dilemma: individuals must get together in order to benefit from the advantages of the division of labor, but various difficulties restrict the gathering of many individuals.[1]

Among the several questions that are investigated in the literature, the following ones are central: (i) *why are there agglomeration or dispersion forces*? (ii) *why do we observe agglomerations formed by different agents*? and (iii) *why do regions and cities specialize in different activities*? In order to answer these questions, we must consider a variety of models focusing on different aspects. Indeed it would be futile to look for *the* model explaining the economic landscape of economies at different stages of development and in different institutional environments. As mentioned in the paragraph above, an interesting model of economic geography must include both centripetal and centrifugal forces. The corresponding spatial equilibrium is then the result of a complicated balance of forces that push and pull consumers and firms until no one can find a better location. As will be seen, the major models which have been developed do reflect such an interplay.

Though convenient at a high level of abstraction, it should be clear that the concept of agglomeration used in this paper does refer to different real world phenomena. For example, one type of agglomeration arises when restaurants, movie theaters, or shops selling similar products are clustered within the same neighborhood of a city. At the other extreme of the spectrum lies the core-periphery structure corresponding to North–South dualism. Other types of agglomeration can be found in the existence of strong regional disparities within the same country, in the formation of cities having different sizes, or in the emergence of industrial districts where firms have strong technological and/or informational linkages. At this stage, it is probably not necessary to distinguish between these different entities, though the intensity of the forces at work is likely to vary according to the cases.

[1] The term agglomeration is less ambiguous than concentration which is used to describe different phenomena. It has been introduced in location theory by Weber (1909, Chap. 1).

In recent years, a growing number of economists have become interested in the study of location problems. This is probably best illustrated by the work of Lucas (1988), Krugman (1991a, 1991b), Becker and Murphy (1992), among several others, which triggered a new flow of interesting contributions in the field. No doubt, this increased interest has been fostered by the integration of national economies within trading blocks such as the European Union or NAFTA, as well as by its impact on the development of their regions and cities. As market integration dissolves economic barriers between nations, national boundaries no longer provide the most natural unit of analysis. Contrary to a widespread opinion, this question is not new: it has been raised by some scholars at the outset of what had to become the European Union (Giersch, 1949). However, the subject matter remains neglected for a long time despite the suggestions made by Ohlin (1933, 1968, Part III) who proposed to unify interregional trade and location theory. Connections with the new theories of growth are also under scrutiny. Indeed cities, and more generally economic agglomerations, are considered as the main institutions where both technological and social innovations are developed through market and nonmarket interactions. Furthermore, city specialization changes over time, thus creating a geographically diversified pattern of economic development. It seems therefore reasonable to say that *growth is localized*, a fact that had been recognized by early development theorists, such as Myrdal (1957) and Hirshman (1958).

Thus it is fair to say that the *new* economic geography, which we will call *geographical economics*, is in many respects more rooted in standard economic theory than the traditional theories of location. As we will see in the course of this paper, geographical economics has strong connections with several branches of modern economics, including industrial organization and urban economics, but also with the new theories of international trade and of economic growth and development.[2] This suggests that this field has high potentials for further developments and that cross-fertilization can be expected. It has also generated a large flow of empirical studies that use the modern tools of econometrics, thus leading to more solid conclusions.

As in any economic field, several lines of research have been explored in geographical economics. The earliest line was initiated by von Thünen (1826) who sought to explain the pattern of agricultural activities surrounding many cities in pre-industrial Germany. More generally, von Thünen's theory has proven to be very useful in studying land use when economic activities are perfectly divisible. Despite his monumental contribution to economic thought (Samuelson, 1983), von Thünen's ideas

[2] References to potential connections with various field of economics can be found in the extended version of the paper published as the CEPR Discussion Paper 1344.

languished for more than a century without attracting widespread attention. Yet, Alonso (1964) succeeded in extending von Thünen's central concept of bid rent curves to an urban context, in which a marketplace is replaced by an employment center (the Central Business District). Since that time urban economics has advanced rapidly.

However the von Thünen model has several limitations. Indeed the following question suggests itself: why is there a unique city in von Thünen's isolated state? Or a unique Central Business District in most urban economic models? This is likely because increasing returns are at work in the design of trading places or in the production of some private goods. Conceding the point, Lösch (1940) argued that scale economies in production are essential for understanding the formation of economic space, and built a spatial model of monopolistic competition involving increasing returns. Similarly Koopmans (1957, p. 157) claims that:

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

The assumption of nonincreasing returns has indeed dramatic implications for geographical economics. Under nonincreasing returns and a uniform distribution of resources, the economy reduces to a Robinson Crusoe type, where each individual produces for his/her own consumption (backyard capitalism). Each location could thus be a base for an autarkic economy, where goods are produced on an arbitrarily small scale, except possibly (as in the neoclassical theory of international trade) that trade might occur if the geographic distribution of resources was nonuniform. While pertinent, the unequal distribution of resources seems insufficient as the only explanation of specialization and trade. Furthermore, when capital or labor can move freely, the neoclassical model of trade does not allow for the prediction of the size of regions when natural resources are uniformly distributed. We can therefore safely conclude that *increasing returns to scale are essential for explaining the geographical distribution of economic activities*.[3] However, when indivisibilities are explicitly introduced, nonexistence of a competitive equilibrium in a spatial economy is common, as shown by Koopmans and Beckmann (1957) and Starrett (1978). Furthermore, as noticed by Drèze and Hagen (1978) in a somewhat different context, scale economies in production have another far-reaching implication for the working of the economy: the number of market places open at a competitive equilibrium is likely to be suboptimal. Or, to use a different terminology, *spatial markets*

[3] This statement has sometimes been referred to as the "Folk Theorem" of geographical economics because it has been rediscovered several times by various scholars (see Scotchmer and Thisse, 1992, for a more detailed discussion).

are typically incomplete so that an equilibrium allocation is in general not Pareto-optimal.

If production involves increasing returns, a finite economy accommodates only a finite number of firms which are imperfect competitors. Treading in Hotelling's footsteps, Kaldor (1935) argued that space gives this competition a particular form. Since consumers buy from the firm with the lowest "full price," defined as the posted price plus the transport cost, *each firm competes directly with only a few neighboring firms*, regardless of the total number of firms in the industry. The very nature of the process of spatial competition is, therefore, oligopolistic and should be studied within a framework of interactive decision making. This was one of the central messages conveyed by Hotelling (1929) but was ignored until economists became fully aware of the power of game theory for studying competition in modern market economies (see Gabszewicz and Thisse, 1986, for a more detailed discussion). Following the outburst of industrial organization since the late 70s, it became natural to study the implications of space for competition. New tools and concepts are now available to revisit and formalize the questions raised by early location theorists.

Despite its factual and policy relevance, the question of *why a hierarchical system of cities emerges* remains open. In particular, it is a well-established fact that cities tend to be distributed according to some specific relationship relating their size and their rank in the urban system (what is called the *rank-size rule*). The first attempt to build a spatial theory of the urban hierarchy goes back at least to the German geographer Christaller (1933) who pioneered "central place theory," based on the clustering of market places for different economic goods and services. Though the theory proposed by Christaller, and developed by Lösch, has served as a cornerstone in classical economic geography, it is fair to say that the microeconomic underpinnings of central place theory are still to be developed.

The topic is difficult because it involves various types of nonconvexities which are even more complex to deal with than increasing returns in production. For example, a consumer organizes his shopping itinerary so as to minimize the total cost of purchases, including transport costs. This problem is extremely complex: determining the optimal geographical pattern of purchases requires solving a particularly difficult combinatorial problem, and finding an equilibrium becomes very problematic (Eaton and Lipsey, 1982). In the same vein, there are often considerable scale economies in carrying the goods bought by a consumer when shopping. These various nonconvexities affect demand functions in complex ways which have not been fully investigated. This is just one example of the many difficulties one encounters in attempting to construct a general spatial model that would consider cities of different sizes trading different commodities. It is therefore no surprise that we still lack such a model since it is well known

that economic theory has serious problems in dealing with nonconvexities. Yet, this turns out to be a real embarrassment because the rank-size rule is one of the most robust statistical relationships known so far in economics.

A major centripetal force can be found in the existence of ''externalities'' since the geographical concentration of economic activities can be viewed as a snowball effect. Specifically, more and more agents want to agglomerate because of the various factors that allow for a larger diversity and a higher specialization in the production processes, and the wider array of products available for consumption. The setting up of new firms in such regions gives rise to new incentives for workers to migrate there because they can expect better job matching and, therefore, higher wages. This in turn makes the place more attractive to firms which may expect to find the types of workers and services they need, as well as new outlets for their products. Hence, both types of agents benefit from being together. This process has been well described by Marshall (1890, 1920, p. 225) in the following quotation:

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

More generally, the ''Marshallian externalities'' arise because of (i) mass-production (the so-called internal economies which are similar to the scale economies mentioned above), (ii) the formation of a highly specialized labor force based on the accumulation of human capital and face-to-face communications, (iii) the availability of specialized input services, and (iv) the existence of modern infrastructures. Not surprisingly Marshallian externalities are the engine of economic development in the new growth theories.

The advantages of proximity for production have their counterpart on the consumption side. For example, cities are typically associated with a wide range of products and a large spectrum of public services so that consumers can reach higher utility levels and, therefore, have stronger incentives to migrate toward cities. Furthermore the propensity to interact with others, the desire of man for man, is a fundamental human attribute, as are the pleasure to discuss and to exchange ideas with others. Distance is an impediment to such interactions, thus making cities the ideal institution for the development of social contacts corresponding to various kinds of externalities (Fisher, 1982, Chaps. 2 and 3).

Before describing the content of the paper, we want to clarify the following issue. For many years, the concept of *externality* has been used to

describe a great variety of situations. Following Scitovsky (1954), it has been customary to consider two categories: ''technological externalities'' (such as spillovers) and ''pecuniary externalities.'' The former deals with the effects of nonmarket interactions which are realized through processes directly affecting the utility of an individual or the production function of a firm. By contrast, the latter refers to the benefits of economic interactions which take place through usual market mechanisms via the mediation of prices. For obvious reasons Marshall was not aware of this distinction, and his externalities turn out to be a mixture of technological and pecuniary externalities. As a consequence, each type of externality may lead to the agglomeration of economic activities.

In order to understand how an agglomeration occurs when Marshallian externalities are present, it is useful to divide human activities into two categories: *production* and *creation*. The former stands for the routine ways of processing or assembling things (such as the preparation of a dinner or the working of an assembly line). For an agglomeration of firms and households to be based on this type of production activity, the presence of pecuniary externalities is crucial. However human beings enjoy more pleasure from, and put much value on, creative activities. Furthermore, in economic life, much of the competitiveness of individuals and firms is due to their creativity. Consequently, as emphasized by Jacobs (1969), economic life is creative in the same way as are arts and sciences and, as pointed out more recently by Lucas (1988, p. 38), personal communication within groups of individuals sharing common interests can be a vital input to creativity:

> New York City's garment district, financial district, diamond district, advertising district and many more are as much intellectual centers as is Columbia or New York University.

In this respect, it is well known that *face-to-face communication* is most effective for rapid product development. For example, Saxenian (1994, p. 33) emphasizes the importance of this factor in the making of the Silicon Valley as an efficient productive system:

> By all accounts, these informal conversations were pervasive and served as an important source of up-to-date information about competitors, customers, markets, and technologies. Entrepreneurs came to see social relationships and even gossips as a crucial aspect of their business. In an industry characterized by rapid technological change and intense competition, such informal communication was often of more value than more conventional but less timely forums such as industry journals.

Given that different people have different skills (by nature as well as by nurture), the size of such groups also gives rise to significant scale effects. Furthermore, information and ideas have characteristics of public goods and, hence, tend to generate spillover effects. In this way, the creative process itself can lead to strong agglomeration tendencies.

Thus *an economic agglomeration is created through both technological and pecuniary externalities*, often working together. Recent advances in geographical economics have mainly concentrated on the Chamberlinian models of monopolistic competition developed in industrial organization by Spence (1976) and Dixit and Stiglitz (1977). As will be seen below, this approach allows one to decipher the working of the pecuniary externalities discussed above (Krugman, 1991a). Accordingly, the section devoted to (technological) externalities will concentrate on production or consumption externalities as they are now defined in modern economic theory, i.e. non-market interactions. These externalities seem to play an increasing role in advanced economies, which are more and more involved in the production and consumption of less tangible goods for which distance matters in a more subtle way than in less advanced economies. This has been observed both in high-tech industries (Saxenian, 1994) and in traditional sectors (Pyke *et al*., 1990).

The remainder of this paper will elaborate on many of the issues discussed above. Because of space constraints, we will concentrate on the main issues only. They will be organized into three themes dealing respectively with *externalities*, *increasing returns* and *spatial competition*. However progress in these three areas has not been the same. In particular, the area of externalities has attracted most attention and, hence, will be discussed first. For the reasons discussed above, we will limit ourselves to a discussion of technological externalities. Formally, such externalities often stand for particular nonconvexities in production or consumption processes. As usual, assuming a continuum of firms and of households permits us to retain the assumption of a competitive behavior while circumventing the many difficulties encountered when nonconvexities are present. In Section 3 we focus on models of monopolistic competition with increasing returns, and show how they can serve to illuminate several aspects of the agglomeration process. One of the most severe limitations of monopolistic competition à la Spence–Dixit–Stiglitz is that price competition is nonstrategic. Yet, as we saw above, spatial competition is inherently strategic because it takes place among the few. Intuitively, one can say that this approach aims at dealing with the strategic externalities generated by the proximity of rival firms or suppliers in economic space. Despite the real progress made during the last decade, spatial competition models are still difficult to manipulate and much work remains to be done in this area. In Section 4 we will review what has been accomplished and will discuss the corresponding implications for geographical economics. In Section 5 we identify a few general principles that seem to emerge from the literature and suggest new lines of research.[4]

[4] The reader is refereed to the excellent book of Ponsard (1983) for a historical survey of spatial economic theory.

Before proceeding, one final remark is in order. Contrary to general beliefs, location problems have attracted a great deal of attention in various disciplines. In economics alone, the topic has been blooming since the early 90s. Thus we have chosen to be selective. As a result, it is fair to say that this survey reflects our idiosyncrasies as much as the state of the art. We owe our apologies to those who have contributed to the field and who feel frustrated by our choice of menu.

2. Externalities

Models involving externalities describe spatial equilibria under the influence of nonmarket interactions among firms and/or households. In modern cities or industrial districts, nonmarket interactions typically take the form of information exchanges between agents. Since most of the corresponding models have been developed in urban economics with the aim to explain the internal structure of cities, we will concentrate on the agglomeration of various economic activities within a city. However, it should be clear that the same principles apply to the spatial organization of broader areas such as regions or nations.[5]

The central idea behind the formation of cities has been very well summarized by Lucas (1988, p. 30):

> What can people be paying Manhattan or downtown Chicago rents *for*, if not for being near other people?

To the best of our knowledge, the first contribution focusing on the role of interaction among individuals as an explanation for cities is due to Beckmann (1976). More precisely, the utility of a household is assumed to depend on the average distance to all households in the city and on the amount of land bought on the market. In equilibrium *the city exhibits a bell-shaped population density distribution*, which is supported by a similarly shaped land rent curve. Focusing on firms instead, Borukhov and Hochman (1977) and O'Hara (1977) studied models of firm location in which interactions between firms generate agglomeration.

The basic contribution, in that the key-variables are independent of the economic system, is due to Papageorgiou and Smith (1983). They consider a trade-off between the need for social contacts, which is negatively affected by distance, and the need for land, which is negatively affected by crowding.

[5] However, they do not necessarily apply to multinational spaces when different national governments are present. Such governments have indeed very specific and powerful instruments, such as money or trade policy, that strongly affect the economic environment in which the agents operate. The study of location problems in the international marketplace is still in infancy and constitutes a very promising line of research.

Initially the preferences are such that the uniform distribution of individuals over a borderless landscape is an equilibrium. When the propensity to interact with others increases enough, this equilibrium becomes unstable: any marginal perturbation is sufficient for the population to evolve toward an irregular distribution. In this model, cities are considered as the outcome of a social process combining basic human needs which are not (necessarily) expressed through the market. It is probably fair to say that this model captures much of the intuition of early geographers interested in the spatial structure of human settlements. However, it is important to consider less general, abstract formulations and to study models based on explicit and economic forms of interactions.

2.1.

To illustrate more concretely the fundamental mechanism of agglomeration involving both firms and households, we give a brief description of a model developed by Fujita, Imai, and Ogawa. The agglomeration force is *the existence of informational spillovers among firms* (see, e.g., Saxenian, 1994, Chap. 2). An important characteristic of information is its public-good nature: the use of a piece of information by a firm does not reduce its content for other firms. Hence the diffusion of information within a set of firms generates externality-like benefits to each of them. Provided that the information owned by firms is different, the benefits of communication generally increase as the number of firms involved rises. Furthermore, since the quality of information involves distance–decay effects, the benefits are greater if firms locate closer to each other. Therefore, all other things being equal, each firm has an incentive to be close to others, thus fostering the agglomeration of firms. On the other hand, the clustering of many firms in a single area increases the average commuting distance for their workers which, in turn, increases the wage rate and land rent in the area surrounding the cluster. Such high wages and land rents tend to discourage the agglomeration of firms in the same area. Consequently the equilibrium distributions of firms and households are determined as the balance between these opposite forces.

Suppose that in a given location space X there is a continuum of firms that are symmetric in the pattern of spillovers. However, they are different in the information they own as well as in the goods they produce. Therefore, each firm gains from the informational spillovers generated by others. Let $a(x, y)$ be the resulting benefit for a firm at x obtained from a firm at y. Then, if $f(y)$ denotes the density of firms at each location $y \in X$,

$$A(x) \equiv \int_X a(x, y) f(y)\, dy \tag{2.1}$$

expresses the *aggregate benefit* that a firm at x can enjoy from the information field within the city. Assume also that each firm needs some given amount of land (S_f) and of labor (L_f). Consequently, if $R(x)$ and $W(x)$ represent the land rent and wage rate prevailing at x, the profit of a firm located at $x \in X$ is equal to

$$\Pi(x) = A(x) - R(x)S_f - W(x)L_f. \tag{2.2}$$

Next there is a continuum of homogeneous households who seek location in the same space. The utility of a household is given by $U(s, z)$, where s represents the land consumption and z the consumption of a composite good. For simplicity, we assume that the land consumption is fixed and equal to S_h. Furthermore, each household supplies one unit of labor and the composite good is imported at a constant price normalized to one. Then, if a household chooses to reside at $x \in X$ and to work at $x_w \in X$, his budget constraint is given by:

$$z + R(x)S_h + t_h|x - x_w| = W(x_w),$$

where t_h is the unit commuting cost. Since the lot size is fixed, the objective of a household is to choose a residential location and a working location which maximize the consumption of the composite good given by

$$z(x, x_w) = W(x_w) - R(x)S_h - t_h|x - x_w|.$$

Finally, in line with mainstream urban economics, it is supposed that land is owned by absentee landlords.

Following the standard approach in land use theory where firms and households are free to choose their locations, the equilibrium configuration is determined through the interplay of the firms' and households' bid rent functions (see Fujita, 1989, Chap. 2, for a detailed discussion of this procedure). An equilibrium is then reached when all the firms achieve the same maximum profit, all the households the same maximum utility, while rents and wage clear the land and labor markets. The unknowns are the firm distribution, the household distribution, the land rent function, the wage function, the commuting pattern, the maximum utility level, and the maximum profit level.

The case of a linear, unbounded space has been studied by Fujita–Ogawa and Imai in different papers. They show that the properties of the equilib-

rium configuration crucially depend on the shape of the local benefit function. Consider the following two examples:[6]

$$a(x, y) = \beta \exp(-\alpha|x - y|) \tag{2.3}$$

and

$$a(x, y) = \beta - \alpha|x - y|, \tag{2.4}$$

where α and β are two positive constants, α measuring the intensity of the distance–decay effect. The former corresponds to a *spatially discounted benefit*, while the latter corresponds to a *linear benefit*.

In the case of a linear benefit, Ogawa and Fujita (1980) and Imai (1982) show that a unique equilibrium configuration exists for each parameter constellation. The equilibrium configuration is *monocentric*, *incompletely mixed* or *completely mixed*. The first (second and third, respectively) configuration occurs, not surprisingly, when α/t_h is large (intermediate and small, respectively).[7] Hence multiple centers cannot arise under linear benefit functions. The case of a spatially discounted benefit leads to more possible cases. Fujita and Ogawa (1982) show that, in addition to the three configurations just mentioned, several other equilibrium configurations may arise. Examples include a *duocentric city*, where each business district is segmented into two labor pools associated with the adjacent residential areas; a city with *one central business district and two subcenters*; and a system of three cities, each having its own CBD, or as *one city with three subcenters*. Furthermore the solution is not necessarily unique: multiple equilibria occur over a wide range of parameter values. Finally *the city may undergo a catastrophic structural transition when parameters take some critical values*. Hence these models are successful in explaining several important features of modern cities such as the endogenous formation of CBDs and subcenters, as well as the transition from a monocentric city to a polycentric one.

2.2.

In the models above, the firm is considered as a single-unit entity. Consequently they are not able to explain a basic trend observed in the spatial organization of large cities, that is, *the location of firm-units in suburban*

[6] These two functions can be derived from explicit benefit functions (Fujita and Smith, 1990).

[7] This type of externality has been further explored by Kanemoto (1990) who considers the case where firms can engage into transactions with others. Combining the exchange of intermediate inputs between firms with indivisibilities in their production creates externalities similar to those considered by Fujita-Imai-Ogawa. If τ is the unit transportation cost of the intermediate goods, Kanemoto then shows that the monocentric configuration is an equilibrium when the ratio τ/t_h is large, a condition similar to that stated above.

areas. For example, many firms (e.g., banks or insurance companies) have recently moved part of their activities (such as book-keeping, planning, and employee training) to the suburbs; similar moves have been observed earlier in the case of industrial activities (cf. Hohenberg and Lees, 1985). In this case, a firm typically conducts some of its activities (such as communications with other firms) at the *front-office* located in the CBD while the rest of its activities are carried out at the *back-office* set up in the suburbs.

This problem has been recently tackled by Ota and Fujita (1993). Keeping the other assumptions of the Fujita–Imai–Ogawa model unchanged, it is now assumed that each firm consists of a front-unit and a back-unit. Each front-unit is assumed to interact 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 belonging to the same firm. Each firm must choose the location of the front-unit and back-unit so as to maximize its profit. If a firm sets up its front-unit at $x \in X$ and back-unit at $y \in X$, the firm incurs an intrafirm communication cost $\Gamma(x, y)$ which depends only upon the locations x and y. As before, each front-unit needs S_f units of land and L_f units of labor; each back-unit requires S_b units of land and L_b units of labor.

In this context, the only change from the previous model is in the profit function (2.2). A firm having a front-unit at x, a back-unit at y and choosing a level of contact activity $q(x, z)$ with the front-unit of any other firm at $z \in X$ has now a profit function defined as follows

$$\Pi(x, y) = A(x) - R(x)S_f - W(x)L_f - R(y)S_b - W(y)L_b - \Gamma(x, y).$$

Assuming that the linear benefit function is linear (see (2.4)) and that the intrafirm communication cost is linear in distance, Ota and Fujita (1993) show that no less than eleven different equilibrium configurations are possible, depending on the values of the various parameters. These configurations are the result of two basic effects: (i) as the commuting cost of workers decreases, the segregation of business and residential areas raises, and (ii) as the intrafirm firm communication cost gets smaller, back-units separate from front-units. The most typical one when intrafirm communication costs are low involves the agglomeration of the front-units at the city center, surrounded by a residential area, while back-units are established at the outskirts of the city together with their employees. Hence *the advancement of intrafirm communication technologies provides a major cause for job suburbanization*. In particular, the recent developments of telecommunication technologies should play a central role on the new spatial organization of production.

3. Increasing Returns

The general principle that lies behind most modern contributions to geographical economics is that *product and/or input differentiation gives rise to agglomeration forces*. This idea is then grafted onto the trade-off between increasing returns and transport costs highlighted in central place theory, in order to generate cumulative processes resulting in the formation of cities and/or industrial districts. In a sense, this corresponds to a revival of ideas advocated by early development theorists, who used related concepts such as the "big push" of Rosenstein–Rodan (1943), the "growth poles" of Perroux (1955), the "circular and cumulative causation" by Myrdal (1957, Chap. 2), and the "backward and forward linkages" by Hirshman (1958, Chap. 1).

In this section, our primary objective is to show how simple models of monopolistic competition may capture agglomeration forces suggested by some of the authors above. In particular, we will see that a major contribution of this approach is to uncover some of the economic mechanisms that underlie the pecuniary externalities evoked in the regional development literature. As mentioned in the introduction, we retain interpretations based on product variety in consumption and/or intermediate goods which are in line with modern theories of growth and international trade.

Consider a population of homogeneous consumers/workers. Each consumes a *homogeneous good* together with *varieties of a differentiated good*. More precisely, when a continuum of varieties of size n is supplied, the utility of a worker is given by a CES-type utility with $0 < \rho < 1$

$$U = z_o^{\alpha} \left\{ \int_0^n [z(\omega)]^{\rho} \, d\omega \right\}^{(1-\alpha)/\rho}, \tag{3.1}$$

where the preferences between the homogeneous good (z_o) and the differentiated goods ($z(\omega)$) is of the Cobb–Douglas-type. When $0 < \rho < 1$, it is well known that ρ measures the degree of substitution between the differentiated varieties and that a low value for ρ means that consumers have a strong preference for variety. More important for our purpose, *the utility of each consumer increases with the number n of varieties*.

Alternatively, as observed by Ethier (1982), the right-hand side of (3.1) can be interpreted as the production function of a competitive firm, which has constant returns in a homogeneous input (z_o) and a composite of differentiated intermediate goods ($z(\omega)$). However this function exhibits *increasing returns in the number n of specialized intermediate goods* used by this firm while ρ now expresses its desire for employing a greater variety of intermediate goods in the production of a final good. In other words,

$$x = z_o^{\alpha} \left\{ \int_0^n [z(\omega)]^{\rho} \, d\omega \right\}^{(1-\alpha)/\rho} \tag{3.2}$$

can be viewed as the "dual" of the utility model (3.1) in the production sector. The importance of specialized intermediate goods (such as legal and communication services, nontraded industrial inputs, maintenance and repair services, finance, etc.) for agglomeration and regional development is a well-documented fact.

In both interpretations, because of specialization in production, each differentiated good $z(\omega)$ is produced by a single firm according to an identical technology, where the only input is labor. The total amount of labor $L(\omega)$ required to produce the quantity $z(\omega)$ is assumed to be given by

$$L(\omega) = f + az(\omega), \tag{3.3}$$

where f is the fixed labor requirement and a the marginal labor requirement. Clearly, this technology exhibits increasing returns to scale. These firms choose their mill (f.o.b.) price and their location in a *nonstrategic* manner in the spirit of Chamberlin (Spence, 1976; Dixit and Stiglitz, 1977). In other words, there is free entry and the number of firms producing the differentiated good/service is very large. Finally, as in von Thünen, an iceberg-type transport cost, in which only a fraction of the good shipped reaches its destination, is assumed (Samuelson, 1983). These assumptions put together have a strong implication. Since the impact of a price change on the total consumption of the differentiated good is negligible (firms are nonstrategic by assumption), a consumer's demand can be shown to be isoelastic. In consequence, because of the multiplicative structure of the transport cost, the elasticity of an individual demand is the same across locations, thus implying that the elasticity of the aggregate demand is independent of the spatial distribution of consumers. For a firm located at x, the equilibrium price for its product is then unique and given by

$$p^*(x) = aW(x)/\rho, \tag{3.4}$$

where $W(x)$ is the equilibrium wage prevailing at x (see below for an example). Thus the equilibrium price is equal to the marginal production cost, $aW(x)$, times a relative mark-up given by $1/\rho > 1$ which rises with the degree of product differentiation.

Two groups of papers, using variants of the model described above, are now discussed. In the first group, we focus on models of *city formation* in the case of a linear space, using a partial equilibrium approach. In the second, a two-region economy is considered and the emphasis is on the

emergence of a *core-periphery* structure. Working with more than two regions (or countries) is known to be complex. We then review some recent work dealing with the formation of an *urban system*. In this second group of papers, the approach is in the spirit of general equilibrium.

3.1.

In the first group, differentiation in consumption and/or intermediate goods is shown to generate endogeneously a city. This idea was developed in a series of contributions published in the late 80s, including Papageorgiou and Thisse (1985), Abdel-Rahman (1988), Fujita (1988; 1989, Ch. 8), Rivera-Batiz (1988), and Abdel-Rahman and Fujita (1990).

Papageorgiou and Thisse (1985) and Fujita (1988) deal with the following system of centripetal/centrifugal forces. Firms are attracted by places where consumers are many because they have a better access to consumers, but are repulsed by places involving many firms because competition is fierce; households are attracted by places where sellers are many in order to have accessibility to a large variety of goods, but are repulsed by places where households are many because of high land rents. While Papageorgiou and Thisse use reduced forms, Fujita assumes explicit market interactions and obtains reduced forms similar to those supposed by the former authors. In Papageorgiou and Thisse the equilibrium configuration is such that both distributions of firms and households are bell-shaped when the purchasing pattern of consumers is dispersed enough, i.e., when the products sold by the firms are sufficiently differentiated. In Fujita two configurations may emerge depending on the relative sizes of consumers and sellers: if there are relatively more (less) consumers than sellers, then most sellers (consumers) agglomerate while most consumers (sellers) surround them. The equilibrium configurations explain here the formation of a *downtown area* where people can find a large number of small stores, restaurants, theaters, and other commercial activities.

On the supply side, it has often been argued that one of the main causes for industrial agglomeration is the availability of specialized local producer services, such as repair and maintenance services, engineering and legal support, transportation and communication services, and financial and advertising services. Based on this observation, Abdel-Rahman and Fujita (1990) consider a city with a final good industry and an intermediate good industry, where the latter supplies a large variety of specialized services to the former. The production function of a firm belonging to the final good industry is given by (3.2), where z_o stands for labor while $z(\omega)$ represents a specialized service. Finally, the production function of the service-firms is as in (3.3). Abdel-Rahman and Fujita then show that the aggregate production function of the city is given by

$$X(N) = AN^{(1-\alpha+\alpha\rho)/\rho},$$

where N is the labor force in the city and A a constant depending on the parameters of the model. Thus, *in the aggregate*, *production in the final sector exhibits increasing returns in the labor force* (the exponent of N is larger than one). The reason for this result lies in the fact that the number of specialized service-firms at the free-entry zero-profit equilibrium rises with N, permitting a finer supply of the intermediate good and the emergence, in turn, of increasing returns at the aggregate level. It is hard here not to think of Marshall (1890, 1920, p. 225):

> The economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind, even though no individual capital employed in the trade be very large. For subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbours, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expense, though its original cost may have been high.

Furthermore, since labor is homogeneous, the equilibrium wage is common to both sectors and also increases with the labor force. Indeed, having more service-firms enhances the productivity of the final sector and, hence, leads to higher wages in both industries. Nevertheless, increasing N leads to an expansion of the residential area which in turn yields higher land rents and transport costs. Thus, in equilibrium, the city achieves a finite size.

Note, finally, that the analysis of Abdel-Rahman and Fujita remains incomplete in that they assume that both types of firms are located at the CBD. When the final sector firms are set up at the CBD, it is reasonable to conjecture that the agglomeration of the service-firms in the CBD is an equilibrium when the intermediate good is differentiated enough, as in the consumption models discussed above.

3.2.

The initial objective of the second family of models is to show *the possibility of divergence between two regions*, while the neoclassical model of interregional trade based on constant returns necessarily leads to the convergence either under free trade or under perfect mobility of labor or capital.[8] The prototype model has been proposed by Krugman (1991a; 1991b) which

[8] Michel *et al.* (1996) show that new conclusions emerge when production externalities (as in modern growth theory) and amenities (as in urban economics) are added to the neoclassical model. In the presence of amenities, the skilled workers may receive different earnings in equilibrium, while a core-periphery structure similar to Krugman (1991b) may emerge as an equilibrium outcome when production externalities are at work. Such an approach extends the neoclassical model following the line of research described in Section 2.

triggered subsequent developments in trade and growth, such as Krugman and Venables (1995a; 1995b), Premer and Walz (1994), Englmann and Walz (1995), Kubo (1995), and Venables (1996), to mention a few.[9]

(*a*). The basic framework can be described as follows. There are two regions, two sectors, and two types of labor. As in the foregoing, agglomeration may arise because of preference for variety on the consumption side or diversity in intermediate goods on the production side. For the sake of brevity, we deal with the first context only. In (3.1), z_o stands for a homogeneous agricultural good (A-good), produced under constant returns using one type of labor (A-workers) and sold on a competitive national market (transport costs are zero). The varieties $z(\omega)$ correspond to differentiated industrial goods (I-goods), produced according to (3.3) where $L(\omega)$ is the other type of labor (I-workers) and sold on monopolistically competitive regional markets (transport costs are positive). The A-workers are immobile, while the I-workers are perfectly mobile. Finally, all workers/consumers have a preference for variety expressed by the utility (3.1).

In this model, the immobility of A-workers is a centrifugal force because they consume both types of goods. The centripetal force is more involved. If a larger number of producers are located in a region, the number of regional products is greater. Then, because firms are mill pricers, the full equilibrium prices are lower there in comparison to the other region, thus generating a real income effect for the corresponding workers (who are also consumers). This, in turn, induces workers to migrate toward this region.[10] The resulting increase in the number of consumers (=workers) creates a larger demand for the I-goods in the region, which therefore leads more firms to locate there. This implies the availability of even more varieties of the differentiated good in the region in question. In this way, a *circular causation* for the agglomeration of firms and workers is generated through *forward linkages* (the supply of more varieties of the I-goods increases the workers' real income) and *backward linkages* (a greater number of consumers attracts more firms). Therefore, through these linkage effects, *scale economies at the individual firm level are transformed into increasing returns at the level of the region as a whole.*

Krugman shows that this mechanism may give rise to *a core-periphery pattern in which the whole production of the I-goods is concentrated into one region*, a regional structure considered by Kaldor (1970) as being more reasonable than the convergence between regions precisely because of the existence of increasing returns. The core-periphery pattern is likely to occur when (i) the transportation rate of the I-goods is low enough, (ii) when

[9] An earlier analysis that anticipated several aspects of Krugman's work was developed by Faini (1984).

[10] This effect of product variety on consumer migration was first emphasized in Stahl (1983).

the I-goods are sufficiently differentiated, or (iii) when the share of the industrial sector in the national economy is large enough. Furthermore, because of the existence of multiple equilibria, minor changes in the values of the critical parameters may generate dramatic changes in the equilibrium spatial configuration. This suggests that history matters (the initial conditions) to explain actual industrial patterns, while circular causation generates a snowball effect that leads manufacturing firms to be *locked-in* within the same region for long periods of time (examples are provided by the ''industrial belt'' in the United States or the ''banane bleue'' in Europe).

Note that a broader set of configurations has been obtained by Helpman (1995) who supposes that the dispersion force is given by a fixed stock of housing, while all individuals are assumed to be perfectly mobile. It is then shown that *both regions accommodate industrial firms*, even though transportation costs are very low, *when the demand for land is high or when products are close substitutes*. However, as in Krugman, industrial concentration arises provided that the demand for land is low or products are differentiated enough, but when transport costs are high instead of low. These results imply that new configurations may emerge when ingredients from urban economics are added into the model. Indeed land is consumed by individuals in Helpman while cities are supposed to be 'punctual' in Krugman and subsequent work. Another interpretation of Helpman's is that the transportation costs of the A-goods are assumed to be prohibitive while they are zero in Krugman's.

(*b*). A line of research in the spirit of Krugman, exploiting the dual model, has been pursued by Englmann and Walz (1995) who study *growth* in a two-region economy. There are two types of labor, the skilled and the unskilled; the former are mobile and the latter are immobile. There are three sectors: the agricultural sector using both types of labor, the industrial sector where the production function is similar to (3.2) but in which the homogeneous input is replaced by the two types of labor, and the R&D sector. Using an endogeneous growth device, these authors suppose that the R&D sector, where only skilled workers are employed and knowledge is accumulated, produces intermediate inputs which are nontraded. Hence, the immobility of the unskilled is a centrifugal force, while the existence of nontraded inputs is a centripetal force. When technological progress is localized, Englmann and Walz show that, at the steady-state, *the production of innovations and of the I-good will take place in the region with the initial advantage in the number of intermediate, nontraded inputs*. The reason for the persistence of leadership lies in the fastest accumulation of knowledge in the region having the initial advantage, while growth is sustained because the marginal productivity of the R&D sector does not decrease to zero. That result provides an explanation for the continuance of a core-periphery

structure; it also sheds light on the role of historical accidents that define the initial conditions of the development process.

However the core-periphery structure is no longer the unavoidable market outcome *when knowledge spills over the other region*. More diversified patterns of regional development involving interior solutions arise because the impact of local intermediate inputs is lessened by the transfer of knowledge, which is itself induced by the existence of interregional spillovers. Furthermore, a developed and rich region might well be less ready to adopt a new technology, so that the lagging region may 'leapfrog' the leading region as a reaction to a major exogeneous change in technology. This would suggest that there is no point of no return.

(*c*). It is well known that results established for two regions are difficult to extend to the case of an arbitrary number of regions. For this reason, Krugman (1993) has extended his initial model to a linear spatial economy. Under the three conditions stated above, he shows that the whole industry tends to concentrate into *a single city* whose location need not be at the center of the segment. Fujita and Krugman (1995) relaxes the assumption that A-workers are immobile and allow for mobility between regions and sectors. Furthermore the transportation costs of the A-goods are now positive. They show that *a single city*, *surrounded by an agricultural area*, *arises when varieties are differentiated enough* (or when transportation costs are low) *and when the population of workers is not too large*. Indeed, if varieties are close substitutes and/or the population is sufficiently large, an individual producer has an incentive to locate far away from the city and to sell a larger output to local consumers. In this case, there is scope for more than one city. Therefore, the work of Fujita and Krugman provides an endogeneous determination of the central city as in von Thünen but within the context of a completely closed model.

The endogeneous determination of several cities has attracted the attention of many scholars but very few results are so far available. In this respect, a recent contribution by Fujita and Mori (1996a) sheds new light on this classical problem of geographical economics. These authors show that, as the population in the national economy increases continuously, new cities are created periodically because of a catastrophic bifurcation in the existing urban system. As the number of cities increases, the urban system approaches a structure where cities are more or less equally distant. Specifically, starting from one city, population growth leads to a larger agricultural area. Beyond some threshold, the agglomeration of industrial firms within a single city is no longer an equilibrium. Some I-workers and some firms leave the existing city to form a new city located deep in the agricultural area, together with some A-workers while new firms are also created. However the size of the existing agglomeration remains large

enough for the other I-workers and firms to stay put. This process keeps going as the population rises. Thus, exactly for the reason suggested by Marshall in the quotation given in the introduction, the locations of the existing cities remain the same though their sizes may vary with the level of population. Finally, there is intercity trade, in addition to trade between cities and rural areas, because the goods produced in the different cities are differentiated and because consumers have a preference for variety.

(*d*). However only one level of city emerges as the outcome of this process. What remains to investigate is the fundamental question of the formation of an urban hierarchy, that is, the construction of an economic theory of central places. A first step into this direction is taken by Fujita *et al.* (1995) who introduce into (3.1) different groups of I-goods, having each different elasticities of substitution and/or transportation rates. As the population rises, they show that *a* (more or less) *regular hierarchical central place system* à la Christaller emerges within the economy, in which 'higher-order cities' provide a larger number of groups of I-goods. There is *two-way trade between cities*, unlike standard central places theory where trade goes from high-order to low-order cities only. However, as expected, *higher-order cities export more varieties than lower-order cities*.

An alternative, original approach to the formation of a system of cities has been pioneered by Henderson (1974). When the production of a good involves increasing returns (see 3.1) and takes place in the Central Business District (see 2.1 and 3.1), Mills (1967) argues that each city has a finite size because of the commuting costs borne by the workers. Then, assuming a ''market for cities,'' Henderson shows that cities will be created until no opportunity exists for a developer or a local government to build a new one. This corresponds to a free entry equilibrium in which all cities are identical. Henderson also shows that cities have an incentive to specialize in the production of traded goods because the production of different goods within the same city rises commuting costs and land rents. Therefore, if the traded goods involve different degrees of scale economies, cities will be specialized in the production of different goods and will export. This approach explains the existence of an urban system formed by *cities having different sizes*, as well as *inter-city trade involving different goods* (see Henderson, 1987, 1988, for further developments). However, this model does not permit to predict the location of cities nor does it explain the urban hierarchical structure. In a sense, Henderson's and Fujita-Krugman's approaches can be viewed as dual: cities have a spatial extension while transportation costs between cities are supposed to be zero in the former, cities have no dimension but intercity trade is costly in the latter.

Finally, though all the models above use very specific functional forms and rest on particular market and transport structures, it seems fair to say

that they point to the right direction. Therefore, they can be viewed as a first step toward the still missing theories of regional development and of central places. More importantly, combining these various approaches, i.e., preference for variety on the product market and diversity/specialization on the input markets, within the same general equilibrium model seems to be an important and challenging task for future research.

Given what we said in the introduction, one of the main limitations of the monopolistic competition models lies in the assumption that firms do not strategically interact (formally this means that we implicitly assume a continuum of firms). Consequently, it is important to deal with oligopolistic rivalry, something which is done in spatial competition. However, as will be seen below, this is not an easy task to accomplish.

4. Spatial Competition

It is now customary to distinguish between two types of models in spatial competition, i.e., the shopping and shipping models. Roughly speaking, we have a *shopping model* when firms charge mill prices while consumers visit firms and bear the whole transportation costs; in a *shipping model*, firms deliver the product and take advantage of the fact that the customers' locations are observable to price discriminate across locations. The former are rooted in the seminal work of Hotelling (1929) while the latter find their origin in Hoover (1937) and Greenhut and Greenhut (1975). Shopping models seem to be appropriate to study competition between sellers of consumption goods while shipping models would describe better competition between sellers of industrial goods.[11] Strategic interaction is at the heart of these models and space is the reason for this behavior: *competition is localized in shopping models* while *shipping models involve oligopolistic competition in spatially separated markets*. Though shopping and shipping models have different aims, the analysis shows that they are governed by the same centrifugal and centripetal forces, thus leading to similar locational patterns under similar conditions. In particular, high transportation costs are always a centrifugal force that results in distinct locations. In consequence, we will limit ourselves to the case of shopping models.

Typically, models of spatial competition assume that the consumer distribution is given. If we introduce a land market and consumer mobility into the Hotelling model then, as observed by Koopmans (1957, Chap. II.9),

[11] In his study of pricing policies followed by business firms in Japan, the United States, and West Germany, Greenhut (1981) finds that about three-quarters of the firms surveyed price discriminate.

the locations of firms and consumers become interdependent. Not much has been done so far and we briefly discussed the few existing contributions.

4.1.

Ever since Hotelling, it has been generally accepted that *competition for market areas is a centripetal force* that would lead vendors to congregate, a result known in the literature as the Principle of Minimum Differentiation. This principle has generated controversies about the inefficiency of free competition since it suggests that ''buyers are confronted everywhere with an excessive sameness'' (Hotelling, 1929, p. 54).

The two ice cream men problem provides a neat illustration of this principle. Two merchants selling the same ice cream at the same fixed price, compete in location for consumers who are uniformly distributed along a linear segment of length L. Each consumer purchases one unit of the good from the nearer firm. The consumers are thus divided into two segments, with each firm's aggregate demand represented by the length of its market segment. The boundary between the two firms' market areas is given by the location of the *marginal consumer* who is indifferent between buying from either firm. This boundary is endogenous, since it depends upon the locations selected by the firms. Since Lerner and Singer (1937), it is well known that the unique Nash equilibrium in pure strategies of this game is given by the location pair

$$x_1^* = x_2^* = L/2$$

regardless of the shape of the transport cost function. Hence, two firms competing for clients choose to locate together at the market center, minimizing their spatial differentiation. Contrary to a wide-spread opinion, this result is not driven by the existence of boundaries. To see it, consider a continuous distribution over the real line. Then, both firms locate back to back at the median of the distribution. It is our belief that several of the results presented below could be extended to this framework.

However things become more complex when (mill) prices are brought into the picture, as in Hotelling's original contribution. Hotelling considers a two-stage game where the firms first simultaneously choose their locations and afterwards their prices. This decoupling of decisions captures the idea that firms select their locations in anticipation of later competing on price. The boundary between the two firms' markets is now given by the location of the consumer for whom the full prices, defined by the posted prices plus the corresponding transport costs, are equal (transport costs are linear in distance). Because of the continuous dispersion of consumers, a marginal variation in price changes the boundary and each firm's demand by the

same order.[12] For each location pair, Hotelling determines what he thinks will be the equilibrium prices of the corresponding price subgame. He includes these prices, which are functions of the locations, into the firms' profit functions, which then depend only upon locations. These new profit functions are used to study the first-stage location game. As in the foregoing, Hotelling finds an equilibrium where the two firms locate at the market center.

Hotelling's analysis was incorrect. When the two firms are sufficiently close, there does not exist an equilibrium in pure strategies for the corresponding price subgame: at least one firm has an incentive to undercut its rival and to capture the whole market. The study of the location game is accordingly incomplete. Nevertheless, as established by d'Aspremont *et al.* (1979), if the transport costs are quadratic rather than linear, a unique price equilibrium exists for any location pair. Reconstructing Hotelling's analysis, these authors then show that the two firms wish to set up at the endpoints of the market.

The extreme spatial dispersion is the result of a trade-off where price competition pushes firms away from each other while competition for market area tends to pull them together. To illustrate how this trade-off works, let Π_1^* be firm 1's profit evaluated at the equilibrium prices $p_i^*(x_1, x_2)$ corresponding to the location pair (x_1, x_2) such that $x_1 < x_2$. Then, since $\partial \Pi_1 / \partial p_1 = 0$, we have

$$d\Pi_1^*/dx_1 = (\partial \Pi_1/\partial p_2)(\partial p_2^*/\partial x_1) + \partial \Pi_1/\partial x_1.$$

In general, the terms in the right-hand side of this expression can be signed as follows. The first one corresponds to the *strategic effect* (the desire to relax price competition) and is expressed by the impact that a change in firm 1's location has on price competition. Since goods are spatially differentiated, they are substitutes so that $\partial \Pi_1^*/\partial p_2$ is positive; because goods become closer substitutes when x_1 increases, $\partial p_2^*/\partial x_1$ is negative. Hence the first term is negative. The second term, which corresponds to the *market area effect* uncovered by Hotelling, is positive. Consequently the impact of reducing the inter-firm distance upon firms' profits is undetermined. However, when firms are close enough, the first term always dominates the second so that *firms always want to be separated in the geographical space*. This implies that the Principle of Minimum Differentiation ceases to hold when firms are allowed to compete in prices (d'Aspremont *et al.*,

[12] d'Aspremont *et al.* (1979) have demonstrated that the hypotheses of Hotelling do not guarantee continuity at the global level. For that it is necessary to replace the assumption of linear transport costs by one in which transport costs are increasing and strictly convex in distance.

1983); one may even observe maximum differentiation. In other words, *price competition is a strong centrifugal force*.

There is no doubt that Hotelling's contribution to economic theory, and in particular to geographical economics, has been fundamental in many respects. Yet, as such, his analysis is unable to explain the currently observed agglomeration of shops selling similar goods. The dispersion of firms turns out to be very sensitive to a particular assumption of the model, namely that consumers patronize the firm with the lowest full price. Somewhat ironically when one knows Hotelling's purpose, the model above corresponds to a very *sharp* consumer behavior that follows from the fact that firms are supposed to sell identical goods. On the contrary, dramatically different results are obtained when consumers behavior is *smooth enough*, for example because firms sell differentiated products.

The idea that consumers distribute their purchases between several sellers is not new in economic geography and goes back at least to Reilly (1931) who formulated the so-called ''gravity law of retailing.'' For a long time, despite their success in empirical studies, gravity models and their extensions, such as the logit, remained somewhat mysterious to the economists because they did not seem to fit the utility maximization assumption. Psychologists have suggested an alternative model of individual choice which imputes a random term to utility and makes the consumer's decision whether to switch firms probabilistic. The use of such models in economics has been pioneered by McFadden (1981) and surveyed by Anderson *et al.* (1992, Chap. 2).

Thus it is now assumed that consumers are influenced by various tangible as well as intangible factors at the moment of their choice, and that the relative importance of these factors may change due to external factors. This implies that consumers' purchase decisions are not based solely on the full prices, but also on firm-specific factors which are typically perceived differently by different consumers. Such a behavior means that consumers at the same location do not react in the same way to a firm's unilateral change in its strategy. The presumably wide array of factors influencing consumers' shopping behavior makes it problematic for a firm to predict exactly a consumer's reactions to a reduction in price. In other words, *the firm assigns a probability between zero and one to whether a particular consumer on a particular date will respond to a price difference by switching firms*. This is modeled by assuming that consumers maximize a *random utility* rather than a deterministic utility.[13] Firms implicitly sell heterogeneous products and the random term in the consumer's utility expresses her matching with firms at the time of purchase. An alternative interpretation is

[13] For our purpose, it is worth noting that Anas (1983) has shown that many descriptive gravity- and logit-type models can be derived from the maximization of a random utility.

that consumers like product variety (see Section 3) so that, even if prices do not vary, they do not always purchase from the same firm over time. In both cases, the indirect utility of a consumer at x and buying from firm i can be modeled as

$$V_i(x) = a - p_i - t\,|x - x_i| + \varepsilon_{ix} \qquad i = 1, \ldots, n, \tag{4.1}$$

where a is a constant measuring the gross utility of the good and ε_{ix} a random variable (with a zero mean) whose realization expresses the matching of product i with a consumer at x. In the special case of the *multinomial logit* (where the random variables ε_{ix} are independently and identically distributed according the double exponential), the probability for a consumer at x to buy from firm i is given by the following expression derived in the econometrics of discrete choices[14]

$$P_i(x) = \frac{\exp(-p_i - t|x - x_i|)/\mu}{\sum_{j=1}^{n} \exp(-p_j - t|x - x_j|)/\mu} \qquad i = 1, \ldots, n, \tag{4.2}$$

where t is the transport rate and μ the standard deviation of the variables ε_{ix} (up to a numerical factor). The values of the choice probabilities $P_i(x)$ reflect those of the full prices: the higher the latter, the lower the former. Consequently, the consumer behavior described by (4.2) encapsulates a tendency to buy from the cheapest shops. Note also that the logit and the CES are closely related in that both models can be derived from the same distribution of consumer tastes; the only difference is that consumers buy one unit of the product in the former and a number of units inversely related to its price in the latter (Anderson *et al.*, 1992, Chaps. 3 and 4).

The expected demand to firm i is equal to the integral of the choice probabilities over the market space; it is *smooth* in prices and locations when μ is strictly positive. However the continuity of profits does not suffice to restore the existence of an equilibrium. Additional restrictions on the parameters are necessary. As will be seen, these restrictions can be given a simple and intuitive interpretation: *the relative importance of the transport costs must be small compared to that of the idiosyncratic components of the individual preferences* (4.1). Formally, this means that μ/tL must be "large enough."

Let c be the common marginal production cost. In the case of a simultaneous choice of prices and locations by firms, the following result holds true:

[14] This formula is well known in classical economic geography but it is usually applied to describe the flows of commodities and of individuals.

if the choice probabilities are given by (4.2) and if the inequality

$$\mu/tL \geq 1/2$$

is satisfied, then the configuration

$$x_i^* = L/2 \text{ and } p_i^* = c + n\mu/(n-1) \qquad i = 1, \ldots, n \tag{4.3}$$

is a Nash equilibrium (de Palma *et al.*, 1985). Therefore, *firms choose to agglomerate at the market center*, as Hotelling thought, *when their products are differentiated enough and when transportation costs* (or market size) *are low enough*. When firms are gathered at the market center, they constitute a very attractive pole for the consumers who may find there the best product, as in Fujita and Krugman (1995) discussed in 3.2. However, products must be differentiated enough for the advantage of being agglomerated to dominate the incentive to move away from the cluster and to charge a higher price.

When transport costs are low, the benefits of geographical separation are reduced and prices are lower. Firms then choose to reconstruct their profit margins by differentiating their products in terms of some nongeographical characteristics, which may be tangible or intangible. Stated differently, *product differentiation is substituted to geographical dispersion* (this is shown in a partial equilibrium model of spatial competition by Irmen and Thisse, 1996). In this case, they no longer fear the effects of price competition (the centrifugal force is weakened by the differentiation of products) and strive to be as close as possible to the consumers with whom the matching is the best. Since these consumers are spread all over the market space, they set up at the market center and, therefore, minimize their geographical differentiation. This is reminiscent of the market potential theory, developed by Harris (1954) in classical economic geography, according to which firms tend to locate where they have the ''best'' access to markets where they can sell their product. The difference is that here the point of highest ''potential'' corresponds to a Nash equilibrium. Furthermore, firms charge a price equal to marginal cost plus an absolute markup.

Consider now the implications of the logit for the sequential, original Hotelling duopoly model. Anderson *et al.* (1992, Chap. 9) have shown the existence and the uniqueness of a price equilibrium for any location pair when μ/tL is large enough. Using this price equilibrium, these authors are then able to study the location game. The following results emerge. As μ/tL rises from 0 to 0.062, there is no location equilibrium. For $0.062 \leq \mu/tL < 1.47$, there is a symmetric dispersed equilibrium which initially entails increasing geographical separation of firms. However, when μ/tL

goes beyond some threshold (around 0.30), the geographical separation starts to decrease. For $0.76 \leq \mu/tL < 1.47$, an agglomerated equilibrium exists along the dispersed one; however the former is unstable while the latter is stable. Finally, for $\mu/tL \geq 1.47$ there is a unique equilibrium that involves central agglomeration.

The intuition behind these results, which is reminiscent of what we saw with the CES, is as follows. An arbitrarily small amount of heterogeneity among products/consumers is not sufficient to restore existence because consumers' shopping behavior remains very sharp. When existence is guaranteed, firms' market areas overlap, thus making price competition so fierce that firms want to move apart. Beyond some threshold, the product differentiation effect tends to dominate the price competition effect and firms set up closer to the market center because price competition is relaxed. Finally, for a sufficiently large degree of differentiation, the market area effect becomes predominant and the agglomeration of sellers is the market outcome as in the nonprice competition context. In both the simultaneous and sequential games, the message is the same: *agglomeration arises when price competition is weakened enough*.[15]

Furthermore, unlike what we observe in the homogeneous product case, the agglomeration may be socially desirable. The social welfare function includes both product differentiation benefits and transportation costs in an entropy-like function where μ plays the role of preference for variety. As μ/tL rises, from 0 to 1/2, the inter-firm distance decreases from $L/2$ (firms are located at the quartiles) to 0 (firms are located at the market center). When μ/tL exceeds 1/2, it is socially optimal for the duopolists to be located back to back; thus, in this model, *the market tends to provide excessive geographical dispersion* (Anderson *et al*., 1992, Chap. 9). In other words, spatial competition does not necessarily lead to excessive sameness, as Hotelling thought.

Another approach, based on the idea of *search*, is explored by Schulz and Stahl (1996). Building on early work by Stahl (1982) and Wolinsky (1983), these authors suppose that the total demand is variable: consumers have the same reservation price but are uniformly distributed along the real line. Thus, if consumers have different tastes and are uncertain about the characteristics of the products on offer, the firms can manipulate the search cost structure by joining an existing market or by establishing a new one. The trade-off faced by a firm is as follows: a firm captures a small market share when setting up in a large market or obtains the whole market when opening a new one. Since total demand is elastic, a *demand externality* arises when more firms are located together because more consumers will

[15] Another example is provided by price collusion in the context of a repeated price game (Friedman and Thisse, 1993).

benefit from economies of scope in searching (that is, the extent of the product market is endogeneous) and, therefore, will visit the cluster. Such an externality is obviously a centripetal force.[16]

Though collectively several firms may want to form a new market, it may not pay an individual firm to open a new market in the absence of a coordinating device. Consequently, a new firm entering the market will choose instead to join the incumbents, thus leading to an increase in the agglomeration size. The entry of a new firm creates a positive externality for the existing firms by making total demand larger. Though price competition becomes fiercer, it appears here that firms take advantage of the extensive margin effect to increase their prices in equilibrium.[17]

A related idea is explored by Gehrig (1996) when two differentiated markets are located at the endpoints of a linear segment. Unlike Schulz and Stahl, Gehrig supposes that the aggregate demand over the two local markets is fixed. The number of products available in a local market increases with the number of consumers visiting this place, thus reducing the average matching costs. The attractiveness of market therefore depends on the size of its clientele. Gehrig then shows that, in such a setting, *an entrant is likely to join one of the existing markets, especially when transportation costs are low*.

4.2.

In the models of spatial competition discussed above, the distribution of consumers is fixed. Ideally, one would like to make it endogeneous. So far there have been few attempts to do so because of the complexity of the problem. Since firms have more market power than consumers, it seems reasonable to assume that firms locate first, anticipating the subsequent consumers' locations and demand functions. When products are homogeneous, such a process may reinforce the tendency toward dispersion. Indeed, when firms are dispersed, consumers pay smaller transport costs on average and may also pay lower average land prices since the supply of attractive lots (those close to firms) is greater. The resulting income effect would increase consumers' demand for private goods and make geographical

[16] Observe that such an externality cannot arise in the standard model of spatial competition because total demand is fixed. On the other hand, it is at the heart of the monopolistic competition models through the forward linkage effect.

[17] In work in progress, Stahl (1995) develops an alternative shopping model in which transport costs are lump-sum. Indeed, there are often considerable scale economies in carrying the goods bought by a consumer. In the limit, consumers' outlays on transportation can be considered as independent of the purchased quantities. Therefore, if the utility functions are homothetic, a more distant consumer has a lower income and demands fall in the same proportion. This leads to much simpler aggregate demand functions and allows Stahl to derive new results and to extend those in Schulz and Stahl (1996).

isolation even more profitable than in the Hotelling model (Fujita and Thisse, 1986).

However the mere existence of a public facility or of a major transportation node might be enough to attract firms within the same urban area. Indeed, other things the same, transportation costs are reduced for consumers who then have higher disposable incomes to buy the composite good sold by the firms (Thisse and Wildasin, 1992). In other words, *the existence of a pre-existing public facility yields an incentive for agglomeration of firms and consumers within an urban area*.[18]

5. Conclusion

Though the economic analysis of agglomeration is still in infancy, a few general principles seem to emerge from the results discussed in the foregoing, they are briefly discussed in 5.1. We will conclude with some suggestions for future research in Section 5.2.

5.1.

First, it should be clear that the existence of scale economies at the firms' level is a critical factor for explaining the emergence of agglomeration. Indeed the mere existence of indivisibilities in production makes it profitable for firms to concentrate production in a relatively small number of plants producing for dispersed consumers, so that increasing returns to scale constitute a strong centripetal force. However, we cannot leave the argument at that. Indeed, the geographical extension of markets, and the corresponding transportation costs, imply that the entire production is generally not concentrated in one place. In other words, the spatial dispersion of demand is a centrifugal force. Therefore, *there is a fundamental trade-off between scale economies and transportation costs in the geographical organization of markets*.

Second, the secular fall in transportation costs often intensifies the tendency toward agglomeration. Although this decrease could have suggested that firms become indifferent about their location, we have seen in various models that *low transport costs, or more generally trade costs, tend to favor the formation of geographical clusters or to deter the creation of new ones*. There are at least two reasons behind this phenomenon. First, as transportation costs decrease, firms have an incentive to concentrate their production in a smaller number of sites in order to reduce fixed costs, as suggested by the trade-off mentioned above. Second, as seen in 4.1, low transportation

[18] See Fujita and Mori (1996b) who study the formation of port-cities using a monopolistic competition model similar to those discussed in 3.2.

costs makes price competition fierce, thus inducing firms to differentiate their products to relax price competition. This in turn leads firms to benefit from the advantages of ''central locations'' where, on average, they are the closest possible to the consumers for whom the matching is best. The counterpart of that result is that *product differentiation is a strong force toward agglomeration*. Behind this result lies the following fundamental cause: when price competition is relaxed (e.g., price collusion or quantity competition with a homogeneous product), firms no longer fear the devastating effects of price competition and the various centrifugal forces discussed in this paper might well be predominant. In other words, even if products are potential substitutes, additional forces make them *complements*. Agglomeration may then emerge as an equilibrium outcome because competition is overcome by other effects (Matsuyama, 1995; Stahl, 1995). Observe also that similar arguments apply to labor: wage competition is a centrifugal force as is price competition, while a better access to a diversified labor pool for both firms and workers is a centripetal force.

However, it should be kept in mind that the models surveyed in this paper are still very simple. In richer models integrating more realistic patterns of migration, new effects might emerge that could more than offset the direct effects identified above. For example, Puga (1996) shows that a drastic fall in communication and transportation costs may lead to geographic dispersion when the mobility costs of workers between regions are arbitrarily large while the mobility costs between sectors are positive but finite (unlike Krugman, 1991a, who assumes prohibitive costs). He also supposes the existence of input/output linkages, as in Venables (1996), otherwise there would be no agglomeration force. In this context, *the agglomeration of firms into a single region intensifies competition on the corresponding local labor market because workers are not spatially mobile*. Though firms can attract workers from the agricultural sector, the latter effect turns out to be a dominant centrifugal force when transportation costs are low enough because they are able to supply the other region at low cost while benefiting of low wages. Mori (1995) obtains comparable results in a continuous model with a land market: firms are willing to locate away from cities because of the lower wages they pay in the agricultural areas and because the supplying costs of the manufactured goods in cities are low; workers are willing to leave cities because the cost of agricultural goods is lower and, therefore, real wages are higher in the rural hinterlands while manufactured goods are available at prices close to those prevailing in cities. Thus new cities may emerge when the population is large enough, leading to a more dispersed pattern of economic activities. Finally, using simulations, Krugman and Venables (1995a) also predict *the collapse of the core-periphery structure and the convergence between regions when trade*

costs are sufficiently low for reasons similar to those discussed above. They summarize their results as follows (p. 476):

> The world economy must achieve a certain critical level of integration before the forces that cause differentiation into core and periphery can take hold; and when differentiation occurs, the rise in core income is partly at peripheral expense. As integration proceeds further, however, the advantages of the core eroded, and the resulting rise in peripheral income may be partly at the core's expense.

These preliminary findings suggest that there would be no monotonic relationship between the degree of geographical concentration and the level of transportation costs. *Very high or very low trade costs would favor the dispersion of economic activities, while agglomeration would emerge for intermediate values of these costs once the spatial mobility of workers is low*. In other words, the relationship between trade costs and the degree of geographical concentration of the economy would be U-shaped. More work is needed to check the robustness of these results, while empirical studies are required to evaluate their real-world implications.[19] They are indeed very important since they would suggest that ''incompletely'' integrated markets, that characterize most trading blocks, would favor the polarization of space while full integration would be associated with a large diffusion of economic activities across regions.

In addition, as discussed in 2.2, low transportation costs may also favor the delocation of activities that need not be close to other producers and/or are labor-intensive. More generally, most models in geographical economics suppose a space homogeneous in terms of ''socio-economic'' factors. Still, it is well known that some firms seek a location in areas where the labor cost is very low because their products can be produced by means of labor-intensive techniques. Under such circumstances, the fall in transportation costs can be viewed as a dispersion force that fosters convergence across countries, as in the neoclassical model. In this perspective, it seems promising to model the firm as a multi-location agent in the context of the new theories of the firm. Indeed the way firms organize their activities may induce particular forms of convergence between various areas since the choice of a location for a particular activity of the firm may obey different logics. Clearly, there is a strong need to integrate the agglomeration models surveyed in this paper with neoclassical models of trade based on comparative advantage in order to study the new international division of labor. Some preliminary attempts dealing with the tension between agglomeration

[19] Note that all these models do not integrate externalities as such discussed in Section 2. They also assume horizontal, and not vertical product differentiation which permits a better description of product innovation. The combination of these two factors seems to be essential for the localized growth of some industries (Saxenian, 1994, Chap. 5), and it is not clear that the results above would remain the same when they are taken into account.

and factor price equalization can be found in Matsuyama and Takahashi (1994) as well as in Puga and Venables (1996).

Third, *the size of the population is also an important determinant of the urban structure of the economy*. We have seen that more cities are likely to emerge when the population rises. Indeed, since production is characterized by increasing returns to scale, larger markets allows for the entry of more firms that can serve as a basis for new clusters and a denser urban pattern. Furthermore, a larger population also permits a better match between consumers/workers and products/job requirements, as well as a wider range of intermediate inputs. In the aggregate, this is reflected by a higher degree of returns to scale on the production side, but also by higher welfare levels on the consumption side. However this process comes to an end when the addition of a consumer/worker leads to an increase in transportation and congestion costs that offset the benefit this individual may derive from variety. Hence, beyond some threshold, firms and consumers/workers have an incentive to form a new city. However, in a hierarchical urban system, a population increase can instead boost the growth of the highest order cities (for example, think of New York, Paris, Tokyo, but also of some urban giants in the Third World).

Finally, in many models of geographical economics, *there is multiplicity of equilibria*. This is because the agglomeration of economic activities has the nature of a cumulative, self-reinforcing process and because the emergence of a particular site as a major agglomeration does not only depend upon the intrinsic features of this site. In other words, *history matters for economic geography* in that initial conditions appear to be essential in the selection of a particular equilibrium. It is then well known that minor changes in the socio-economic environment occurring at some critical periods may result in very different geographical configurations. This might well explain why the location of new agglomerations is difficult to predict.

Furthermore, some equilibria turn out to be socially preferable to others so that *there is scope for regional policy*. More precisely, the role of the government would be here to create the conditions for the ''best'' equilibrium to arise. Yet it is fair to say that our knowledge of the underlying dynamics is by far too rudimentary to permit us making detailed policy recommendations. In addition, even though a spatial structure might well be inefficient, it is likely to be difficult to change it because of the lock-in effects associated with existing agglomerations. Another reason for this inertia, related to Krugman (1991c) and Matsuyama (1995), is the formation of self-fulfilling prophecies about the development of some areas. Indeed, it seems reasonable to consider existing cities as focal points that help agents coordinate their spatial decisions. In such a context, reshaping the urban landscape would then require major changes in agents' expectations.

Accordingly, we seem to have a *putty-clay geography*: there is *a priori* a great deal of flexibility in the choice of locations but a strong rigidity of the urban structure once the process of urbanization has started. Those factors, together or in isolation, could explain why, in many countries and at different times, many ''planned cities'' have failed to develop once governments have stopped supporting most of the urban activities before some critical mass was reached (exceptions involving massive involvement of national governments through huge coordination programs include Brasilia and Saint Petersburg).

Though more work is called for, it is worth mentioning that these preliminary results seem to fit well the waves of urbanization observed in Europe in the 12th century, as well as the process of urban growth that took place during the Industrial Revolution in Europe and the United States, so well analyzed by Bairoch (1985) in his masterful economic history of cities.

5.2.

One of the main limitations of most models of geographical economics is that results seem to heavily depend on strong assumptions made about the economy; in particular very specific functional forms, like the CES or the logit, are used in most models. It is reasonable to think that such simplifying assumptions are needed at an early stage of development of the theory. However, one should now strive for more robust results. For example, it might well be possible to extend the CES model by using the random utility model of monopolistic competition developed in Anderson *et al.* (1992, Chap. 6) that permits to retain the idea of symmetry within a much more general framework. (See also Mirrlees (1995) for another extension dealing with several goods.) Similarly using an iceberg transport cost function implies that any increase in the mill price is accompanied with a proportional increase in transport cost, which seems both unrealistic and undesirable. It is also known that aggregating local demands across locations may lead to demand systems that exhibit undesirable features, such as discontinuities or outward kinks. Still, even if simplifying assumptions are probably unavoidable, more attention should be paid to the aggregation problem over space. In the monopolistic competition models discussed in Section 3, it is not clear what the nonstrategic interaction assumption implies and one should try to relax it. In spatial competition models (see Section 4), there is clearly a need for more work devoted to the endogeneous determination of the consumers' locations and the combination of atomistic and nonatomistic markets.

There are several important questions which remain on the research agenda. First more work is called for about the emergence of urban hierarchies. Central place theory is probably the main topic in geographical economics, though very few major results are so far available. There is no

doubt that the problem is hard, but it is too important to be ignored any longer (insightful suggestions for new developments can be found in Stahl, 1987). In particular, it would be interesting to pursue the comparison of the self-organization approach advocated by Krugman (1996) to that developed by Henderson (Becker and Henderson, 1996; Henderson and Mitra, 1996) for whom modern urban landscapes are mold by large agents. Each approach has its own merits that should be further investigated. In this perspective, there is a new line of research that emerges in modern economic theory in which agents have a certain probability to meet, which depends on socio-economic and geographical factors; when the interaction occurs, a transaction may happen between the corresponding agents (Kirman, 1996). This type of model might prove to be useful to study the emergence of market-towns where people meet in order to trade goods or exchange information.[20] It also seems to be in accord with the self-organization approach.

Second, the question of regional convergence/divergence has at last received the attention it has long deserved, especially in the empirical literature. However models are still too preliminary to draw strong policy recommendations, and more developments are required. In particular, we do not know much about the circumstances that lead a region to recover. In the real world, we observe that some regions are successful in their economic revival while others seem to decline inexorably. It is not always clear why such differences arise. Third, the role of infrastructure, emphasized in the endogeneous growth literature, has not been studied in the new theories of regional economics. So far we have very few insights about what could well be a ''good'' infrastructure policy in the context of a spatial economy. Building transportation infrastructures is often presented as the main remedy to regional imbalance, but this is a policy in search of a theory. See, however, Martin and Rogers (1995) for a first attempt to evaluate the impact of infrastructures on the regional distribution of production in a model similar to those reviewed in 3.2. Fourth, most models of economic agglomeration assume a one-dimensional world. Though acceptable as a first approximation, one should try to go further and to construct more general models allowing for a second dimension. This creates unsuspected difficulties in that the metrics proposed in location theory for measuring distance in a two-dimensional space have different mathematical properties.

Last, all existing models of geographical economics assume full employment (Zenou and Smith, 1995, is a noticeable exception). Even during the Golden Sixties, some regions have experienced persistent unemployment. Nowadays the distribution of unemployment seems to be fairly uneven

[20] For general, competitive equilibrium models of marketplace formation, see Berliant and Wang (1993) and Wang (1993).

across regions, even within the same country. We have a very poor understanding of these questions, and the appeal to low regional mobility of workers, though relevant in some cases, seems weak at the main explanation for regional unemployment disparities. This important economic and social problem should be given more attention in the future. A possible line of research would be to integrate concepts of labor economics and of matching and search models of unemployment into the corpus of geographical economics.

References

Abdel-Rahman, H. (1988). Product differentiation, monopolistic competition and city size, *Region. Sci. Urban Econ.* **18,** 69–86.

Abdel-Rahman, H., and Fujita, M. (1990). Product variety, Marshallian externalities, and city sizes, *J. Region. Sci.* **30,** 165–183.

Alonso, W. (1964). "Location and Land Use," Harvard Univ. Press, Cambridge, MA.

Anas, A. (1983). Discrete choice theory, information theory, and the multinomial logit and gravity models, *Transport. Res. B* **17,** 13–23.

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

Bairoch, P. (1985) "De Jéricho à Mexico. Villes et économie dans l'histoire," Gallimard, Paris.

Becker, G., and Murphy, K. (1992). The division of labor, coordination costs, and knowledge, *Quart. J. Econ.* **107,** 1137–1160.

Becker, R., and Henderson, J. V. (1996). City formation, mimeo, Brown University.

Beckmann, M. J. (1976). Spatial equilibrium in the dispersed city, *in* "Mathematical Land Use Theory (Y. Y. Papageorgiou, Ed.), pp. 117–125. Lexington Books, Lexington, MA.

Berliant, M., and Wang, P. (1993). Endogenous formation of a city without agglomerative externalities or market imperfections: market places in a regional economy, *Region. Sci. Urban Econ.* **23,** 121–144.

Borukhov, E., and Hochman, O. (1977). Optimum and market equilibrium in a model of a city without a predetermined center, *Environ. Planning A* **9,** 849–856.

Christaller, W. (1933). "Die Zentralen Orte in Süddeutschland," Gustav Fischer Verlag, Jena.

d'Aspremont, C., Gabszewicz, J. J., and Thisse, J.-F. (1979). On Hotelling's stability in competition, *Econometrica* **47,** 1045–1050.

d'Aspremont, C., Gabszewicz, J. J., and Thisse, J.-F. (1983). Product differences and prices, *Econ. Lett.* **11,** 19–23.

de Palma, A., Ginsburgh, V., Papageorgiou, Y. Y., and Thisse, J.-F. (1985). The principle of minimum differentiation holds under sufficient heterogeneity, *Econometrica* **53,** 767–781.

Dixit, A. K., and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity, *Amer. Econ. Rev.* **67,** 297–308.

Drèze, J., and Hagen, K. (1978). Choice of product quality: equilibrium and efficiency, *Econometrica* **48.**

EATON, B. C., AND LIPSEY, R. G. (1982). An economic theory of central places, *Econ. J.* **92,** 56–72.

ENGLMANN, F. C., AND WALZ, U. (1995). Industrial centers and regional growth in the presence of local inputs, *J. Region. Sci.* **35,** 3–27.

ETHIER, W. (1982). National and international returns to scale in the modern theory of international trade, *Amer. Econ. Rev.* **72,** 389–405.

FAINI, R. (1984). Increasing returns, non-traded inputs and regional development, *Econ. J.* **94,** 308–323.

FISCHER, CL. (1982). "To Dwell among Friends: Personal Networks in Town and City," Univ. of Chicago Press, Chicago.

FRIEDMAN, J. W., AND THISSE, J.-F. (1993). Partial collusion fosters minimum product differentiation, *Rand J. Econ.* **24,** 631–645.

FUJITA, M. (1988). A monopolistic competition model of spatial agglomeration: a differentiated product approach, *Region. Sci. Urban Econ.* **18,** 87–124.

FUJITA, M. (1989). "Urban Economic Theory. Land Use and City Size," Cambridge Univ. Press, Cambridge, MA.

FUJITA, M., AND KRUGMAN, P. (1995). When is the economy monocentric? von Thünen and Chamberlin unified, *Region. Sci. Urban Econ.* **25,** 505–528.

FUJITA, M., KRUGMAN, P., AND MORI, T. (1995). On the evolution of hierarchical urban systems, Institute of Economic Research, Discussion Paper No 419, Kyoto University.

FUJITA, M., AND MORI, T. (1996a). Structural stability and evolution of urban systems, *Region. Sci. Urban Econ*., forthcoming.

FUJITA, M., AND MORI, T. (1996b). The role of port in the making of major cities: self-organization and hub-effects, *J. Develop. Econ.* **49,** 93–120.

FUJITA, M., AND OGAWA, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations, *Region. Sci. Urban Econ.* **12,** 161–196.

FUJITA, M., AND SMITH, T. E. (1990). Additive-interaction models of spatial agglomeration, *J. Region. Sci.* **30,** 51–74.

FUJITA, M., AND THISSE, J.-F. (1986). Spatial competition with a land market: Hotelling and von Thünen unified, *Rev. Econ. Stud.* **53,** 819–841.

GABSZEWICZ, J. J., AND THISSE, J.-F. (1986). Spatial competition and the location of firms, *in* "Location Theory" (J. J. Gabszewicz, J.-F. Thisse, M. Fujita, and U. Schweizer, Eds.), pp. 1–71 Harwood Academic, Chur.

GEHRIG, T. (1996). Competing exchanges, *Euro. Econ. Rev*., forthcoming.

GIERSCH, H. (1949). Economic union between nations and the location of industries. *Rev. Econ. Stud*. **17,** 87–97.

GREENHUT, J., AND GREENHUT, M. L. (1975). Spatial price discrimination, competition and locational effects, *Economica* **42,** 401–419.

GREENHUT, M. L. (1981). Spatial pricing in the U.S.A., West Germany, and Japan, *Economica* **48,** 79–86.

HARRIS, C. (1954). The market as a factor on the localization of industry in the United States, *Ann. Assoc. Amer. Geograph.* **64,** 315–348.

HELPMAN, H. (1995). The size of regions, The Eitan Berglas School of Economics, Working Paper No. 14-95, University of Tel Aviv.

HENDERSON, J. V. (1974). The sizes and types of cities, *Amer. Econ. Rev.* **64,** 640–656.

HENDERSON, J. V. (1987). Systems of cities and inter-city trade, *in* "Systems of Cities and

Facility Location,'' (P. Hansen, M. Labbé, D. Peeters, J.-F. Thisse, and J. V. Henderson, Eds.), pp. 71–119. Harwood Academic, Chur.

Henderson, J. V. (1988). ''Urban Development. Theory, Fact and Illusion.'' Oxford Univ. Press, Oxford.

Henderson, J. V., and Mitra, A. (1996). The new urban landscape: developers and edge cities, *Region. Sci. Urban Econ*., forthcoming.

Hirschman, A. O. (1958). ''The Strategy of Development,'' Yale Univ. Press, New Haven, CT.

Hoover, E. M. (1937). Spatial price discrimination, *Rev. Econ. Stud.* **4,** 182–191.

Hotelling, H. (1929). Stability in competition, *Econ. J.* **39,** 41–57.

Imai, H. (1982). CBD hypothesis and economies of agglomeration. *J. Econ. Theory* **28,** 275–299.

Irmen, A., and Thisse, J.-F. (1996). ''Competition in Multi-Characteristics Spaces: Hotelling Was Almost Right,'' Discussion Paper No. 1446, CEPR, London.

Jacobs, J. (1969). ''The Economy of Cities,'' Random House, New York.

Kaldor, N. (1935). Market imperfection and excess capacity, *Economica* **2,** 35–50.

Kaldor, N. (1970). The case for regional policies, *Scot. J. Polit. Econ.* **17,** 337–348.

Kanemoto, Y. (1990). Optimal cities with indivisibilities in production and interactions between firms, *J. Urban Econ.* **27,** 46–59.

Kirman, A. P. (1996). Economies with interacting agents, *Games Econ. Behav*., forthcoming.

Koopmans, T. C. (1957). ''Three Essays on the State of Economic Science,'' McGraw–Hill, New York.

Koopmans, T. C., and Beckmann, M. J. (1957). Assignment problems and the location of economic activities, *Econometrica* **25,** 1401–1414.

Krugman, P. (1991a). ''Geography and Trade,'' MIT Press, Cambridge, MA.

Krugman, P. (1991b). Increasing returns and economic geography, *J. Polit. Econ.* **99,** 483–499.

Krugman, P. (1991c). History versus expectations, *Quart. J. Econ.* **106,** 651–667.

Krugman, P. (1993). First nature, second nature, and metropolitan location, *J. Region. Sci.* **33,** 129–144.

Krugman, P. (1996). ''The Self-organizing Economy,'' Basil Blackwell, Oxford.

Krugman, P., and Venables, A. J. (1995a), Globalization and the inequality of nations, *Quart. J. Econ.* **110,** 857–880.

Krugman, P., and Venables, A. J. (1995b), ''The Seamless World: A Spatial Model of International Specialization,'' Discussion Paper No. 1230, CEPR, London.

Kubo, Y. (1995). Scale economies, regional externalities, and the possibility of uneven regional development, *J. Region. Sci.* **35,** 29–42.

Lerner, A., and Singer, H. W. (1937). Some notes on duopoly and spatial competition, *J. Polit. Econ.* **45,** 145–186.

Lösch, A. (1940). ''Die Räumliche Ordnung der Wirtschaft,'' Gustav Fischer Verlag, Jena.

Lucas, R. E. (1988). On the mechanics of economic development, *J. Monetary Econ.* **22,** 3–22.

Marshall, A. (1890). ''Principles of Economics,'' Macmillan, London (8th ed. published in 1920).

Martin, Ph., and Rogers, C. A. (1995). Industrial location and public infrastructure, *J. Int. Econ.* **39,** 335–351.

Matsuyama, K. (1995). Complementarities and cumulative process in models of monopolistic competition, *J. Econ. Lit.* **33,** 701–729.

Matsuyama, K., and Takahashi, T. (1994). ''Self-Defeating Regional Concentration,'' Cen-

ter for Mathematical Studies in Economics and Management Science, Discussion Paper No. 1086, Northwestern University.

McFadden, D. (1981). Econometric models of probabilistic choice, *in* "Structural Analysis of Discrete Data with Econometric Applications," (C. F. Manski and D. McFadden, Eds.), pp. 198–272. MIT Press, Cambridge, MA.

Michel, Ph., Perrot, A., and Thisse, J.-F. (1996). Interregional equilibrium with heterogeneous labor, *J. Pop. Econ.* **9,** 95–114.

Mills, E. S. (1967). An aggregate model of resource allocation in a metropolitan area. *Amer. Econ. Rev.* **57,** 197–210.

Mirrlees, J. (1995). Welfare economics and economies of scale, *Japan. Econ. Rev.* **46,** 38–62.

Mori, T. (1995). A model of megapolis: the maturing of city systems, mimeo, University of Pennsylvania.

Myrdal, G. (1957). "Economic Theory and Underdeveloped Regions," Duckworth, London.

Ogawa, H., and Fujita, M. (1980). Equilibrium land use patterns in a non-monocentric city, *J. Region. Sci.* **20,** 455–475.

O'Hara, D. J. (1977). Location of firms within a square central business district, *J. Polit. Econ.* **85,** 1189–1207.

Ohlin, B. (1932). "Interregional and International Trade," Harvard Univ. Press, Cambridge, MA (revised version published in 1968).

Ota, M., and Fujita, M. (1993). Communication technologies and spatial organization of multi-unit firms in metropolitan areas, *Region. Sci. Urban Econ.* **23,** 695–729.

Papageorgiou, Y. Y., and Smith, T. R. (1983). Agglomeration as local instability of spatially uniform steady-states, *Econometrica* **51,** 1109–1119.

Papageorgiou, Y. Y., and Thisse, J.-F. (1985). Agglomeration as spatial interdependence between firms and households, *J. Econ. Theory* **37,** 19–31.

Perroux, F. (1955). Note sur la notion de pôle de croissance, *Economique appliquée* **7,** 307–320.

Ponsard, C. (1983). "History of Spatial Economic Theory," Springer-Verlag, Heidelberg.

Premer M., and Walz, U. 1994. Divergent regional development, factor mobility, and non-traded goods, *Region. Sci. Urban Econ.* **24,** 707–722.

Puga, D. (1996). The rise and fall of economic agglomerations, mimeo, London School of Economics.

Puga, D., and Venables, A. J. (1996). The spread of industry: spatial agglomeration in economic development, *J. Japan. Int. Econ.* **10,** 440–464.

Pyke, F., Becattini, G., and Sengenberger, W. (1990), "Industrial Districts and Inter-firm Cooperation in Italy," International Institute for Labour Studies, Geneva.

Reilly, W. J. (1931). "The Law of Retail Gravitation," Pilsbury, New York.

Rivera-Batiz, F. (1988). Increasing returns, monopolistic competition, and agglomeration economies in consumption and production, *Region. Sci. Urban Econ.* **18,** 125–153.

Rosenstein–Rodan, P. N. (1943). Problems of industrialization of Eastern and South-Eastern Europe, *Econ. J.* **53,** 202–211.

Samuelson, P. A. (1983). Thünen at two hundred, *J. Econ. Lit.* **21,** 1468–1488.

Saxenian, A. (1994). "Regional Advantage: Culture and Competition in Silicon Valley and Route 128," Harvard Univ. Press, Cambridge, MA.

Schulz, N., and Stahl, K. (1996). Do consumers search for the highest price? Equilibrium and monopolistic optimum in differentiated products markets, *Rand J. Econ.* **27,** 542–562.

SCITOVSKY, T., (1954). Two concepts of external economies, *J. Polit. Econ.* **62,** 143–151.

SCOTCHMER, S., AND THISSE, J.-F. (1992). Space and competition: a puzzle, *Ann. Region. Sci.* **26,** 269–286.

SPENCE, M. (1976). Product selection, fixed costs, and monopolistic competition, *Rev. Econ. Stud.* **43,** 217–235.

STAHL, K. (1982). Differentiated products, consumer search, and locational oligopoly, *J. Indust. Econ.* **31,** 97–114.

STAHL, K. (1983). A note on the microeconomics of migration, *J. Urban Econ.* **14,** 318–326.

STAHL, K. (1987). Theories of urban business location, *in* ''Handbook of Regional and Urban Economics,'' (E. S. Mills, Ed.), Volume 2, pp. 759–820. North-Holland, Amsterdam.

STAHL, K. (1995). Towards a microeconomic theory of the retailing sector, mimeo, Universität Mannheim.

STARRETT, D. (1978). Market allocations of location choice in a model with free mobility, *J. Econ. Theory* **17,** 21–37.

THISSE, J.-F., AND WILDASIN, D. (1992). Public facility location and urban spatial structure, *J. Pub. Econ.* **48,** 83–118.

VENABLES, A. J. (1996). Equilibrium locations of vertically linked industries, *Int. Econ. Rev.* **37,** 341–359.

VON THÜNEN, J. H. (1826). ''Det Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie,'' Perthes, Hamburg.

WANG, P. (1993). Agglomeration in a linear city with heterogeneous households, *Region. Sci. Urban Econ.* **23,** 291–306.

WEBER, A. (1909). ''Ueber den Standort der Industrien,'' J. C. B. Mohr, Tübingen.

WOLINSKY, A. (1983). Retail trade concentration due to consumers' imperfect information, *Bell J. Econ.* **14,** 275–282.

ZENOU, Y., AND SMITH, T. E. (1995). Efficiency wages, involuntary unemployment and urban spatial structure, *Region. Sci. Urban Econ.* **25,** 547–573.