Agglomeration and Market Interaction*

Masahisa Fujita† and Jacques-François Thisse‡

23 February 2002

1 Introduction

The most salient feature of the spatial economy is *the presence of a large variety of economic agglomerations*. Our purpose is to review some of the main explanations of this universal phenomenon, as they are proposed in urban economics and modern economic geography. Because of space constraints, we restrict ourselves to the most recent contributions, referring the reader to our forthcoming book for a more complete description of the state of the art.

Although using agglomeration as a generic term is convenient at a certain level of abstraction, it should be clear that the concept of economic agglomeration refers to very distinct real world situations. At one extreme lies the core-periphery structure corresponding to North-South dualism. For example, Hall and Jones (1999) observe that high income nations are clustered in small industrial cores in the Northern Hemisphere, whereas productivity per capita steadily declines with distance from these cores.

As noted by many historians and development analysts, economic growth tends to be localized. This is especially well illustrated by the rapid growth of East Asia during the last few decades. We view East Asia as comprising Japan and nine other countries, that is, Republic of Korea, Taiwan, Hong

*We should like to thank P. Belleflamme, D. Laussel, Th. Mayer, T. Tabuchi and D.-Z. Zeng for comments and suggestions. Both authors are grateful to FNRS (Belgium) and Grants in Aid for COE Research 09CE2002 and Science Research S 13851002 (Japan) for financial support.

†Institute of Economic Research, Kyoto University

‡CORE, Université catholique de Louvain (Belgium), CERAS, Ecole nationale des ponts et chaussées (France) and CEPR.

Kong, Singapore, Philippines, Thailand, Malaysia, Indonesia, and China. In 1990, the total population of East Asia was approximately 1.6 billion. With only 3.5% of the total area and 7.9% of the total population, Japan accounted for 72% of the GDP and 67% of the manufacturing GDP of East Asia. In Japan itself, the economy is very much dominated by its core regions formed by the five prefectures containing the three major metropolitan areas of Japan: Tokyo and Kanagawa prefectures, Aichi prefecture (containing Nagoya MA), and Osaka and Hyogo prefectures. These regions account for only 5.2% of the area of Japan, but for 33% of its population, 40% of its GDP, and 31% of its manufacturing employment. Hence, for the whole of East Asia, the Japanese core regions with a mere 0.18% of the total area accounted for 29% of East Asia's GDP.

Strong regional disparities within the same country imply the existence of agglomerations at another spatial scale. For example, in Korea, the capital region (Seoul and Kyungki Province), which has an area corresponding to 11.8% of the country and 45.3% of the population, produces 46.2% of the GDP. In France, the contrast is even greater: the Ile de France (the metropolitan area of Paris), which accounts for 2.2% of the area of the country and 18.9% of its population, produces 30% of its GDP. Inside the Ile-de-France, only 12 % of the available land is used for housing, plants and roads, the remaining land being devoted to agricultural, forestry or natural activities.

Regional agglomeration is also reflected in large varieties of cities, as shown by the stability of the urban hierarchy within most countries. Cities themselves may be specialized in a very small number of industries, as are many medium-size American cities. However, large metropolises like Paris, New York or Tokyo are highly diversified in that they nest a large variety of industries, which are not related through direct linkages. Industrial districts involving firms with strong technological and/or informational linkages (e.g., the Silicon Valley or Italian districts engaged in more traditional activities) as well as of factory towns (e.g., Toyota City) manifest various types of local specialization. Therefore, it appears that highly diverse size/activity arrangements exist at the regional and urban level.

Although the sources are dispersed, not always trustworthy and hardly comparable, data clearly converge to show the existence of an urban revolution. In Europe, the proportion of the population living in cities increased very slowly from 10% in 1300 to 12% in 1800. It was approximately 20% in 1850, 38% in 1900, 52% in 1950, and is close to 75% nowadays, thus showing an explosive growth in the urban population. In the United States, the rate of urbanization increased from 5% in 1800 to more than 60% in 1950 and is now near 77%. In Japan, the rate of urbanization was about 15% in 1800, 50% in 1950, and is now about 78%. The proportion of the urban population in the world increased from 30% in 1950 to 45% in 1995 and should exceed 50% in 2005. Furthermore, concentration in very big cities keeps rising. In 1950, only two cities had populations above 10 millions: New York and Greater London. In 1995, 15 cities belong to this category. The largest one, Tokyo, with more than 26 millions, exceeds the second one, New York, by 10 millions. In 2025, 26 mega-cities will exceed 10 millions.

Economists must explain *why firms and households concentrate in large metropolitan areas,* whereas empirical evidence suggests that the cost of living in such areas is typically higher than in smaller urban areas (Richardson, 1987). Or, as Lucas (1988, p.39) put it in a neat way: "What can people be paying Manhattan or downtown Chicago rents for, if not for being near other people?" But Lucas did not explain why people want, or need to be near other people.

The increasing availability of high-speed transportation infrastructure and the fast-growing development of new informational technologies might suggest that our economies enter an age that would culminate in the "death of distance". If so, locational difference would gradually fade because agglomeration forces would be vanishing. In other words, cities would become a thing of the past. Matters are not that simple, however, because the opposite trend may as well happen.[1] Indeed, one of the general principles that will come out from our analysis is that the relationship between the decrease in transport costs and the degree of agglomeration of economic activities is not that expected by many analysts: *agglomeration happens provided that transport costs are below some critical threshold*, although further decreases may yield dispersion of some activities due to factor price differentials.[2] In addition, technological progress brings about new types of innovative activities that benefit most from being agglomerated and, therefore, tend to arise in

[1]For example, recent studies show that, in the US, 86% of Net delivery capacity is concentrated in the 20 largest cities. This suggests that the U.S. is quickly becoming a country of digital haves and have-nots, with many small businesses unable to compete, and minority neighborhoods and rural areas getting left out.

[2]Transportation (or transfer) costs are broadly defined to include all the factors that drive a wedge between prices at different locations, such as shipping costs per se, tariff and non-tariff barriers to trade, different product standards, difficulty of communication, and cultural differences.

developed areas (Audretsch and Feldman, 1996). Consequently, the wealth or poverty of people seems to be more and more related to the existence of prosperous and competitive clusters of specific industries, as well as to the presence of large and diversified metropolitan areas.

The recent attitude taken in several institutional bodies and medias seems to support this view. For example, in its *Entering the 21st Century. World Development Report 1999/2000*, the World Bank stresses the importance of economic agglomerations and cities for boosting growth and escaping from the poverty trap. Another example of this increasing awareness of the relevance of cities in modern economies can be found in *The Economist* (1995, p.18):

> "The liberalization of world trade and the influence of regional trading groups such as NAFTA and the EU will not only reduce the powers of national governments, but also increase those of cities. This is because an open trading system will have the effect of making national economies converge, thus evening out the competitive advantage of countries, while leaving those of cities largely untouched. So in the future, the arenas in which companies will compete may be cities rather than countries."

The remainder of the paper is organized as follows. In Section 2, we show why the competitive framework can hardly be the foundation for the economics of agglomeration. We then briefly review the alternative modeling strategies. In the hope to make our paper accessible to a broad audience, Sections 3 presents in detail the two (specific) models that have been used so far to study the spatial distribution of economic activities. Several extensions of these models are discussed in Section 4. Section 5 concludes with some suggestions for further research and policy implications.

2 Modeling Strategies of Economic Agglomerations

As a start, it is natural to ask the following question: to what extent is the competitive paradigm useful in understanding the main features of the economic landscape? The general competitive equilibrium model is indeed the benchmark used by economists when they want to study the market properties of an economic issue. Before proceeding, we should remind the reader that the essence of this model is that all trades are impersonal: when making their production or consumption decisions, economic agents need to know the price system only, which they take as given. At a competitive equilibrium, prices provide firms and consumers with all the information they must know to maximize their profit and their utility.

The most elegant and general model of a competitive economy is undoubtedly that developed by Arrow and Debreu. In this model, a commodity is defined not only by its physical characteristics, but also by the place it is made available. This implies that the same good traded at different places is treated as different economic commodities. Within this framework, choosing a location is part of choosing commodities. This approach integrates spatial interdependence of markets into general equilibrium in the same way as other forms of interdependence. Thus, the Arrow-Debreu model seems to obviate the need for a theory specific to the spatial context.

Unfortunately, as will be seen below, *the competitive model cannot generate economic agglomerations without assuming strong spatial inhomogeneities*. More precisely, we follow Starrett (1978) and show that introducing a homogenous space (in a sense that will be made precise below) in the Arrow-Debreu model implies that *total transport costs in the economy must be zero at any spatial competitive equilibrium*, and thus trade and cities cannot arise in equilibrium. In other words, the competitive model per se cannot be used as the foundation for the study of a spatial economy because we are interested in identifying purely economic mechanisms leading agents to agglomerate in a featureless plain.[3] This is because we concur with Hoover (1948, p.3) for whom:

> "Even in the absence of any initial differentiation at all, *i.e.*, if natural resources were distributed uniformly over the globe, patterns of specialization and concentration of activities would inevitably appear in response to economic, social, and political principles."

[3]Ellickson and Zame (1994) disagree with this claim and argue that the introduction of moving costs in a dynamic setting may be sufficient to save the competitive paradigm. To the best of our knowledge, however, the implications of their approach have not yet been fully worked out.

2.1 The breakdown of the competitive price mechanism in a homogenous spatial economy

The economy is formed by agents (firms and households) and by commodities (goods and services). A firm is characterized by a set of production plans, each production plan describing a possible input-output relation. A household is identified by a relation of preference, by a bundle of initial resources and by shares in firms' profits. A competitive equilibrium is then described by a price system (one price per commodity), a production plan for each firm and a consumption bundle for each household that satisfy the following conditions: at the prevailing prices (i) supply equals demand for each commodity; (ii) each firm maximizes its profit subject to its production set; and (iii) each household maximizes her utility under her budget constraint defined by the value of her initial endowment and her shares in firms' profits. In other words, all markets clear while each agent chooses her most preferred action at the equilibrium prices.

Space involves a finite number of locations. Transportation within each location is costless but shipping goods from one location to another requires the use of resources. Without loss of generality, transportation between any two locations is performed by a profit-maximizing carrier who purchases goods in a location at the market prices prevailing in this location and sell them in the other location at the corresponding market prices, while using goods and land in each location as inputs.

A typical firm produces in a small number of places. Likewise, a household has a very small number of residences. For simplicity, we therefore assume that each firm (each household) chooses a single location and engages in production (consumption) activities there. However, firms and households are free to choose any location they want (the industry is footloose). For expositional convenience, we distinguish explicitly prices and goods by their location. Given this convention, space is said to be *homogenous* when (i) the utility function and the consumption set are the same regardless of the location in which the household resides, and (ii) the production set of a firm is independent of the location elected by this firm. In other words, consumers and producers have no intrinsic preferences for one location over the others. In this context, the following unsuspected result, which we call the *Spatial Impossibility Theorem*, has been proven by Starrett (1978).

Theorem 1 *Consider an economy with a finite number of agents and loca-*

tions. If space is homogenous, transport is costly and preferences are locally nonsatiated, then there is no competitive equilibrium involving transportation.

What does it mean? If economic activities are perfectly divisible, a competitive equilibrium exists and is such that each location operates as an autarky. For example, when households are identical, locations have the same relative prices and the same production structure (backyard capitalism). This is hardly a surprising outcome since, by assumption, there is no reason for economic agents to distinguish among locations and each activity can operate at an arbitrarily small level. Firms and households thus succeed in reducing transport costs at their absolute minimum, namely zero.

However, as observed by Starrett (1978, p.27), when economic activities are *not* perfectly divisible the transport of some goods between some places becomes unavoidable:

> "as long as there are some indivisibilities in the system (so that individual operations must take up space) then a sufficiently complicated set of interrelated activities will generate transport costs" (Starrett, 1978, p.27)

In this case, the Spatial Impossibility Theorem tells us that no competitive equilibrium exists.

This is clearly a surprising result that requires more explanations. For simplicity, we restrict ourselves to the case of two locations, A and B. When both locations are not in autarky, one should keep in mind that the price system must do two different jobs simultaneously: (i) *to support trade between locations* (while clearing the markets in each location), and (ii) *to prevent firms and households from relocating.* The Spatial Impossibility Theorem says that, in the case of a homogenous space, it is impossible to hit two birds with one stone: the price gradients supporting trade bear wrong signals from the viewpoint of locational stability. Indeed, if a set of goods is exported from A to B, then the associated positive price gradients induce producers located in A (who seek a higher revenue) to relocate in B, whereas location B's buyers (who seek lower prices) want to relocate in A. Likewise, the export of another set of goods from B to A encourages such "cross-relocation". The land rent differential between the two locations can discourage the relocation in one direction only. Hence, as long as trade occurs at positive costs, some agents always want to relocate.

To ascertain the fundamental cause for this nonexistence, it is helpful to illustrate the difficulty encountered by using a standard diagram approach. Depicting the whole trade pattern between two locations would require a diagram with six dimensions (two tradable goods and land at each location), which is a task beyond our capability. We thus focus on a two-dimensional subspace of the whole pattern by considering the production of good i only, which is traded between A and B, while keeping the other elements fixed. Because the same physical good available at two distinct locations corresponds to two different commodities, this is equivalent to studying the production possibility frontier between two different economic goods.

Suppose that at most one unit of good i is produced by one firm at either location by using a fixed bundle of inputs. For simplicity, the cost of these inputs is assumed to be the same in both locations. The good is shipped according to an iceberg technology: when x_i units of the good are moved between A and B, only a fraction x_i/Υ arrives at destination, with $\Upsilon > 1$, while the rest melts away en route (Samuelson, 1983). In this context, if the firm is located in A, then the output is represented by point E on the vertical axis in Figure 1; if the entire output is shipped to B, then the fraction $1/\Upsilon$ arrives at B, which is denoted by point F on the horizontal axis. Hence, when the firm is at A, the set of feasible allocations of the output between the two locations is given by the triangle OEF. Space being homogenous, if the firm locates at B, the set of feasible allocations between the two places is now given by the triangle $OE'F'$. Hence, when the firm is not located, the set of feasible allocations is given by the union of the two triangles.

Figure 1: The set of feasible allocations in a homogenous space

Let the firm be set up at A and assume that the demand conditions are such that good i is consumed in both locations so that trade occurs. Then, to support any feasible trade pattern, represented by an interior point of the segment EF, the price vector (p_{iA}, p_{iB}) must be such that $p_{iA}/p_{iB} = 1/\Upsilon$, as shown in Figure 1. However, under these prices, it is clear that the firm can obtain a strictly higher profit by locating in B and choosing the production plan E' in Figure 1. This implies that there is no competitive price system that can support both the existence of trade and a profit-maximizing location for the firm.

This difficulty arises from the nonconvexity of the set of feasible allocations. If transportation were costless, the set of feasible allocations would be given by the triangle OEE' in Figure 1, which is convex. In this case, the firm would face no incentive to relocate. Similarly, if the firm's production activity were perfectly divisible, this set would again be equal to the triangle OEE', and no difficulty would arise.

Therefore, even though the individual land consumption is endogenous, we may conclude that *the fundamental reason for the Spatial Impossibility Theorem is the nonconvexity of the set of feasible allocations caused by the existence of positive transport costs and the fact that agents have an address in space.*

Some remarks are still in order. First, we have assumed that each firm locates into a single region. The theorem could be generalized to permit firms to run distinct plants, one plant per location because each plant amounts to a separate firm in the competitive setting (Koopmans, 1957). Second, we have considered a closed economy. The theorem can readily be extended to allow for trade with the rest of the world provided that each location has the same access to the world markets to satisfy the assumption of a homogenous space. Third, the size of the economy is immaterial for the Spatial Impossibility Theorem to hold in that assuming a "large economy", in which competitive equilibria often emerge as the outcome generated by several institutional mechanisms, does not affect the result because the value of total transport costs within the economy rises when agents are replicated. Last, the following result sheds extra light on the meaning of the Spatial Impossibility Theorem (Fujita and Thisse, 2002).

Corollary 2 *If there exists a competitive equilibrium in a spatial economy with a homogenous space, then the land rent must be the same in all locations.*

This result has the following fundamental implication for us: *in a homogenous space, the competitive price mechanism is unable to explain why the land rent is higher in an economic agglomeration* (such as a city, a central business district, or an industrial cluster) *than in the surrounding area.* This clearly shows the limits of the competitive paradigm for studying the agglomeration of firms and households.

2.2 What are the alternative modeling strategies?

Thus, if we want to understand something about the spatial distribution of economic activities and, in particular, the formation of major economic agglomerations as well as regional specialization and trade, the Spatial Impossibility Theorem tells us that we must make at least one of the following three assumptions:

(i) space is heterogenous (as in the neoclassical theory of international trade),

(ii) externalities in production and consumption exist (as in urban economics),

(iii) markets are imperfectly competitive (as in the so-called "new" economic geography).

Of course, in reality, economic spaces are the outcome of different combinations of these three agglomeration forces. However, it is convenient here to distinguish them to figure out what are the effects of each one of them.

A. Comparative advantage models. The heterogeneity of space introduces the uneven distribution of immobile resources (such as mineral deposits or some production factors) and amenities (climate), as well as the existence of transport nodes (ports, transhipment points) or trading places. This approach, while retaining the assumption of constant returns and perfect competition, yields comparative advantage among locations and gives rise to interregional and intercity trade.

B. Externality models. Unlike models of comparative advantage, the basic forces for spatial agglomeration and trade are generated endogeneously through nonmarket interactions among firms and/or households (knowledge spillovers, business communications and social interactions). Again, this approach allows us to appeal to the constant return/perfect competition paradigm.[4]

C. Imperfect competition models. Firms are no longer price-takers, thus making their price policy dependent on the spatial distribution of consumers and firms. This generates some form of direct interdependence between firms and households that may produce agglomerations. However, it is useful to distinguish two types of approach.

C1. Monopolistic competition. This leads to some depart from the competitive model and allows for firms to be price-makers and to produce

[4]See, e.g. the now classical papers by Henderson (1974) and by Fujita and Ogawa (1982).

differentiated goods under increasing returns; however, strategic interactions are weak because one assumes a continuum of firms.

C2. Oligopolistic competition. Here we face the integer aspect of location explicitly. That is, we assume a finite number of large agents (firms, local governments, land developers) who interact strategically by accounting for their market power.

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

Models of comparative advantage have been extensively studied by international and urban economists (Fujita, 1989), whereas models of spatial competition have attracted a lot of attention in industrial organization (Anderson, de Palma and Thisse, 1992). Because Ed Glaeser and Jose Scheinkman deal with non-market interactions, we choose to focus on market interactions, that is, models belonging to class **C1**. Although this class of models has been initially developed in the context of intra-urban agglomeration with a land market (e.g., Fujita, 1988), we restrict ourselves to *multi-regional models of industrial agglomeration.*

3 Core and Periphery: a monopolistic competition approach

The spatial economy is replete with *pecuniary externalities*. For example, when some workers choose to migrate, they are likely to affect both the labor and product markets in their region of origin, thus affecting the well-being of those who stay put. Moreover, the moving workers do not account either for the impact of their decision on the workers and firms located in the region of destination. Still, their moves will increase the level of demand inside this region, thus making the place more attractive to firms. Everything else being equal, they will also depress the local labor market so that the local wage is likely to be affected negatively. In sum, these various changes may increase or decrease the attractiveness of the destination region for outside workers and firms. Such pecuniary externalities are especially relevant in the context of imperfectly competitive markets because prices do not perfectly reflect the social values of individual decisions. They are also better studied within a general equilibrium context to account for the interactions between the product and labor markets. In particular, such a framework allows us to study the dual role of individuals as workers and consumers. At first sight, this seems to be a formidable task. Yet, as shown by Krugman (1991), several of these various effects can be combined and studied within a simple enough general equilibrium model of monopolistic competition, which has come to be known as the *core-periphery model*.

Recall that monopolistic competition à la Chamberlin involves consumers with a preference for variety (*varietas delectat*) while firms producing these varieties compete for a limited amount of resources because they face increasing returns. The prototype that has emerged from the industrial organization literature is the model developed by Spence (1976) and Dixit and Stiglitz (1977), sometimes called the S-D-S model. These authors assume that each firm is negligible in the sense that it may ignore its impact on, and hence reactions from, other firms, but retains enough market power for pricing above marginal cost regardless of the total number of firms (like a monopolist). Moreover the position of a firm's demand depends on the actions taken by all firms in the market (like in perfect competition).

In many applications, the S-D-S model is proven to be a very powerful instrument for studying the aggregate implications of monopoly power and increasing returns, and so especially when these are the basic ingredients of self-sustaining processes such as those encountered in modern theories of growth and geography (Matsuyama, 1995). This is because of the following reasons. First, although each firm is a price-maker, strategic interactions are very weak in this model, thus making the existence of an equilibrium much less problematic than in general equilibrium under imperfect competition (see, e.g. Bonanno, 1990). Second, the assumption of free entry and exit leads to zero profit so that a worker's income is just equal to her wage, another major simplification. Last, the difference between price competition and quantity competition that plagues oligopoly models is immaterial in a monopolistic competitive setting. Indeed, being negligible to the market, each firm behaves as a monopolist on her residual demand, which makes it indifferent between using price or quantity as a strategy.

3.1 The framework

We consider a $2\times2\times2$ setting. The economic space is made of two regions (A and B). The economy has two sectors, the modern sector ($\mathbb{M}$) and the traditional sector ($\mathbb{T}$). There are two production factors, the high-skilled workers (H) and the low-skilled workers (L). The $\mathbb{M}$-sector produces a continuum of varieties of a horizontally differentiated product under increasing returns, using H as the only input. The $\mathbb{T}$-sector produces a homogenous good under constant returns, using unskilled labor L as the only input.

The economy is endowed with L unskilled workers and with H skilled workers (labor dualism). The skilled workers are perfectly mobile between regions, whereas the unskilled workers are immobile. This extreme assumption is justified because the skilled are more mobile than the unskilled over long distances (SOPEMI, 1998). Finally, the unskilled workers are equally distributed between the two regions, and thus regions are a priori symmetric.

The technology in the $\mathbb{T}$-sector is such that one unit of output requires one unit of L. The output of the $\mathbb{T}$-sector is costlessly traded between any two regions and is chosen as the numéraire so that $p^{\mathbb{T}} = 1$. Hence, the wage of the unskilled workers is also equal to 1 in both regions. Each variety of the $\mathbb{M}$-sector is produced according to the same technology such that the production of the quantity $q(i)$ requires $l(i)$ units of skilled labor given by

$$l(i) = f + cq(i) \tag{1}$$

in which f and c are, respectively, the fixed and marginal labor requirements. Because there are increasing returns but no scope economies, each variety is produced by a single firm. This is because, due to the consumers' preference for variety, any firm obtains a higher share of the market by producing a differentiated variety than by replicating an existing one.

The market equilibrium is the outcome of the interplay between a dispersion force and an agglomeration force. The centrifugal force is very simple. It lies in two sources: (i) the spatial immobility of the unskilled whose demands for the manufactured good are to be met and (ii) the fiercer competition that arises when firms locate back to back (d'Aspremont, Gabszewicz and Thisse, 1979). The centripetal force is more involved. If a larger number of firms is located in one region, the number of varieties locally produced is also larger. This in turn induces some skilled living in the smaller region to move toward the larger region in which they may enjoy a higher standard of living. The resulting increase in the numbers of consumers creates a larger demand for the differentiated good which, therefore, leads additional firms to locate in this region. This implies the availability of more varieties in the region in question but less in the others because there are scale economies at the firm's level. Consequently, as noticed by Krugman (1991a, p.486), there is *circular causation* à la Myrdal because these two effects reinforce each other: "manufactures production will tend to concentrate where there is a large market, but the market will be large where manufactures production is concentrated."

Let λ be the fraction of skilled residing in region A and denote by $v_r(\lambda)$ the indirect utility a skilled worker enjoys in region $r = A, B$ when the spatial distribution of skilled is $(\lambda, 1-\lambda)$. A *spatial equilibrium* arises at $\lambda \in (0,1)$ when

$$\Delta v(\lambda) \equiv v_A(\lambda) - v_B(\lambda) = 0$$

or at $\lambda = 0$ when $\Delta v(0) \leq 0$, or at $\lambda = 1$ when $\Delta v(1) \geq 0$. Such an equilibrium always exists when $v_r(\lambda)$ is a continuous function of λ. However, this equilibrium is not necessarily unique. Stability is then used to eliminate some of them. The stability of such an equilibrium is studied with respect to the following equation of motion:[5]

$$\dot{\lambda} = \lambda \Delta v(\lambda)(1-\lambda) \tag{2}$$

If $\Delta v(\lambda)$ is positive and $\lambda \in (0,1)$, workers move from B to A; if it is negative, they go in the opposite direction. Clearly, any spatial equilibrium is such that $\dot{\lambda} = 0$. A spatial equilibrium is *stable* if, for any marginal deviation of the population distribution from the equilibrium, the equation of motion above brings the distribution of skilled workers back to the original one.[6] We assume that local labor markets adjust instantaneously when some skilled workers move from one region to the other. More precisely, the number of firms in each region must be such that the labor market clearing conditions (12) and (22) stated below remain valid for the new distribution of workers.

[5]This dynamics implies that the equilibrium is reached for $t \to \infty$. One could alternately use the dynamic system proposed by Tabuchi (1986) in which the corner solutions $\lambda = 0$ and $\lambda = 1$ are reached within finite times. The difference becomes critical when the economy exhibits different equilibrium patterns over time.

[6]Note that (2) provides one more justification for working with a continuum of agents: this modeling strategy allows one to respect the integer nature of a agent's location (her address) while describing the evolution of the regional share of production by means of a differential equation.

Wages are then adjusted in each region for each firm to earn zero profits in any region having skilled workers because the skilled move according to the utility differential.

3.2 A model with CES utility and iceberg transport costs

Although consumption takes place in a specific region, it is notationally convenient to describe preferences without explicitly referring to any particular region. Preferences are identical across all workers and described by a Cobb-Douglas utility:

$$u = Q^{\mu}T^{1-\mu}/\mu^{\mu}(1-\mu)^{1-\mu} \qquad 0 < \mu < 1 \tag{3}$$

where Q stands for an index of the consumption of the modern sector varieties, and T is the consumption of the output of the traditional sector. Because the modern sector provides a continuum of varieties of size M, the index Q is given by

$$Q = \left[\int_0^M q(i)^{\rho} di\right]^{1/\rho} \qquad 0 < \rho < 1 \tag{4}$$

where $q(i)$ represents the consumption of variety $i \in [0, M]$. Hence, each consumer displays a *preference for variety.* In (4), the parameter ρ stands for the inverse of the intensity of love for variety over the differentiated product. When ρ is close to 1, varieties are close to perfect substitutes; when ρ decreases, the desire to spread consumption over all varieties increases. If $\sigma \equiv 1/(1-\rho)$, then σ is the elasticity of substitution between any two varieties. Because there is a continuum of firms, each firm is negligible and the interactions between any two firms are zero, but aggregate market conditions of some kind (e.g., the average price across firms) affects any single firm. This provides a setting in which firms are not competitive (in the classic economic sense of having infinite demand elasticity), but at the same time they have no strategic interactions with one another (see (5) below).

If y denotes the consumer income and $p(i)$ the price of variety i, then the demand functions are

$$q(i) = \mu y p(i)^{-\sigma} P^{\sigma-1} \qquad i \in [0, M] \tag{5}$$

where P is the price index of the differentiated product given by

$$P \equiv \left[\int_0^M p(i)^{-(\sigma-1)} di \right]^{-1/(\sigma-1)} \tag{6}$$

The corresponding indirect utility function is

$$v = yP^{-\mu} \tag{7}$$

Without loss of generality, we choose the unit of skilled labor such that $c = 1$ in (1). The output of the $\mathbb{M}$-sector is shipped at a positive cost according to the "iceberg" technology: when one unit of the differentiated product is moved from region r to region s, only a fraction $1/\Upsilon$ arrives at destination with $\Upsilon > 1$. Because mill and discriminatory pricing can be shown to be equivalent in the present setting, we may use the mill pricing interpretation in what follows. When variety i is sold in region r at the mill price $p_r(i)$, the price $p_{rs}(i)$ paid by a consumer located in region s ($\neq r$) is

$$p_{rs}(i) = p_r(i)\Upsilon$$

If the distribution of firms is (M_r, M_s), using (6) the price index P_r in region r is then given by

$$P_r = \left\{ \int_0^{M_r} p_r(i)^{-(\sigma-1)} di + \Upsilon^{-(\sigma-1)} \int_0^{M_s} p_s(i)^{-(\sigma-1)} di \right\}^{-1/(\sigma-1)} \tag{8}$$

which clearly depends on the spatial distribution of firms as well as the level of transport costs.

Let w_r denote the wage rate of a skilled worker living in region r. Because there is free entry and exit and, therefore, zero profit in equilibrium, the income of region r is

$$Y_r = \lambda_r H w_r + L/2 \qquad r = A, B \tag{9}$$

where λ_r is the share of skilled workers residing in region r.

Using (5), the total demand of the firm producing variety i and located in region r is

$$q_r(i) = \mu p_r(i)^{-\sigma} Y_r (P_r)^{\sigma-1} + \mu p_r(i)^{-\sigma} Y_s \Upsilon^{-(\sigma-1)} (P_s)^{\sigma-1} \tag{10}$$

Because each firm has a negligible impact on the market, it may accurately neglect the impact of a price change over consumers' income (Y_r) and other firms' prices, hence on the regional price indices (P_r). Consequently, (10) implies that, regardless of the spatial distribution of consumers, each firm faces an iso-elastic demand. This very convenient property depends crucially on the assumption of an iceberg transport cost, which affects here the level of demand but not its elasticity.

The profit function of a firm in r is:

$$\pi_r(i) = [p_r(i) - w_r]q_r(i) - w_r f$$

Because varieties are equally weighted in the utility function, the equilibrium price is the same across all firms located in region r. Solving the first order condition yields the common equilibrium price

$$p_r^* = \frac{w_r}{\rho} \tag{11}$$

Substituting p_r^* into $\pi_r(i)$ leads to

$$\pi_r = \frac{w_r}{\sigma - 1}[q_r - (\sigma - 1)f]$$

Under free entry, profits are zero so that the equilibrium output of a firm is given by $q_r^* = (\sigma - 1)f$, which is independent of the spatial distribution of demand. As a result, in equilibrium a firm's labor requirement is a constant given by $l^* = \sigma f$, and thus the total number of firms in the $\mathbb{M}$-sector is equal to $H/\sigma f$. The corresponding distribution of firms

$$M_r = \lambda_r H/\sigma f \qquad r = A, B \tag{12}$$

depends only upon the distribution of the skilled workers. Hence, the model allows for studying the spatial distribution of the modern sector but not for its size.

Introducing the equilibrium prices (11) and substituting (12) for M_r in the regional price index (8) gives:

$$P_r = \kappa_1 \left[\lambda_r w_r^{-(\sigma-1)} + \lambda_s \left(w_s \Upsilon\right)^{-(\sigma-1)}\right]^{-1/(\sigma-1)} \tag{13}$$

where κ_1 is a positive constant.

Finally, we consider the labor market clearing conditions for a given distribution of workers. The wage prevailing in region r is the highest wage that firms located there can pay under the nonnegative profit constraint. For that, we evaluate the demand (10) as a function of the wage through the equilibrium price (11):

$$q_r(w_r) = \mu(w_r/\rho)^{-\sigma}(Y_r P_r^{\sigma-1} + Y_s \Upsilon^{-(\sigma-1)} P_s^{\sigma-1})$$

Because this expression is equal to $(\sigma-1)f$ when profits are zero, we obtain the following implicit expression for the zero-profit wages:

$$w_r^* = \kappa_2(Y_r P_r^{\sigma-1} + Y_s \Upsilon^{-(\sigma-1)} P_s^{\sigma-1})^{1/\sigma} \tag{14}$$

where κ_2 is a positive constant. Clearly, w_r^* is the *equilibrium wage* in region r when $\lambda_r > 0$. Substituting (9) for Y_r in the indirect utility (7), we obtain the *real wage* as follows:

$$v_r = \omega_r = \frac{w_r^*}{P_r^\mu} \qquad r = A, B \tag{15}$$

Finally, the Walras law implies that the traditional sector market is in equilibrium provided that the equilibrium conditions above are satisfied.

Summarizing the foregoing developments, the basic equations for our economy are given by (9), (13), (14), and (15). From now on, set $\lambda_A = \lambda$ and $\lambda_B = (1-\lambda)$.

3.2.1 The core-periphery structure

Suppose that the modern sector is concentrated in one region, say region A so that $\lambda = 1$. We wish to determine conditions under which the real wage a skilled worker may obtain in region B does not exceed the real wage she gets in region A.

Setting $\lambda = 1$ in (9), (13), (14), and (15), we get:

$$\frac{\omega_B}{\omega_A} = \left(\frac{1+\mu}{2}\Upsilon^{-\sigma(\mu+\rho)} + \frac{1-\mu}{2}\Upsilon^{-\sigma(\mu-\rho)}\right)^{1/\sigma} \tag{16}$$

The first term in the right hand side of (16) is always decreasing in Υ. Therefore, if $\mu \geq \rho$ the second term is also decreasing so that the ratio ω_B/ω_A

always decreases with Υ, thus implying that $\omega_B < \omega_A$ for all $\Upsilon > 1$. This means that the core-periphery structure is a stable equilibrium for all $\Upsilon > 1$. When

$$\mu \geq \rho \tag{17}$$

varieties are so differentiated that firms' demands are not very sensitive to differences in transportation costs, thus making the agglomeration force very strong.

More interesting is the case in which

$$\mu < \rho \tag{18}$$

that is, varieties are not very differentiated so that firms' demand are sufficiently elastic for the agglomeration force to be weak. If (18) holds, $\Upsilon^{-\mu\sigma+\sigma-1}$ goes to infinity when $\Upsilon \rightarrow \infty$ and the ratio ω_B/ω_A is as depicted in Figure 2.

Figure 2. The determination of the sustain point

In this case, there exists a single value $\Upsilon_{sustain} > 1$ such that $\omega_B/\omega_A = 1$. Hence, the agglomeration is a stable equilibrium for any $\Upsilon \leq \Upsilon_{sustain}$. This occurs because firms can enjoy all the benefits of agglomeration without losing much of their business in the other region. Such a point is called the *sustain point* because, once firms are fully agglomerated, they stay so for all smaller values of Υ. On the other hand, when transportation costs are sufficiently high ($\Upsilon > \Upsilon_{sustain}$), firms lose much on their exports, and thus the core-periphery structure is no longer an equilibrium.

Summarizing this discussion, we obtain:

Proposition 3 *Consider a two-region economy.*

(i) If $\mu \geq \rho$, then the core-periphery structure is always a stable equilibrium.

(ii) If $\mu < \rho$, then there exists a unique solution $\Upsilon_{sustain} > 1$ to the equation

$$\frac{1+\mu}{2}\Upsilon^{-\sigma(\mu+\rho)} + \frac{1-\mu}{2}\Upsilon^{-\sigma(\mu-\rho)} = 1$$

such that the core-periphery structure is a stable equilibrium for any $\Upsilon \leq \Upsilon_{sustain}$.

Interestingly, this proposition provides formal support to the claim made by Kaldor (1970, p.241) more than 30 years ago:

> "When trade is opened up between them, the region with the more developed industry will be able to supply the need of the agricultural area of the other region on more favourable terms: with the result that the industrial centre of the second region will lose its market and will tend to be eliminated"

3.2.2 The symmetric structure

Proposition 3 suggests that the modern sector is geographically dispersed when transportation costs are high, at least when (18) holds. To check this, we consider the symmetric configuration ($\lambda = 1/2$). In this case, for a given Υ, the symmetric equilibrium is stable (unstable) if the slope of $\Delta\omega(\lambda)$ is negative (positive) at $\lambda = 1/2$. Checking this condition requires fairly long calculations using all the equilibrium conditions. However, Fujita, Krugman and Venables (1999) have shown the following results. First, when (18) does not hold, the symmetric equilibrium is always unstable. Second, when (18) holds, this equilibrium is stable (unstable) if Υ is larger (smaller) than some threshold value Υ_{break} given by

$$\Upsilon_{break} = \left[\frac{(\rho+\mu)(1+\mu)}{(\rho-\mu)(1-\mu)}\right]^{1/(\sigma-1)} \tag{19}$$

which is clearly larger than one. This is called the *break point* because symmetry between the two regions is no longer a stable equilibrium for lower values of Υ. It is interesting to note that Υ_{break} depends on the same parameters as $\Upsilon_{sustain}$. It is immediate from (19) that Υ_{break} is increasing with the share of the modern sector (μ) and with the degree of product differentiation ($1/\rho$).

Because $\Upsilon_{break} < \Upsilon_{sustain}$ can be shown to hold,[7] there exists a domain of parameters over which there is multiplicity of equilibria, namely agglomeration and dispersion, as depicted in Figure 3.

[7] See Neary (2001) for a simple proof.

Figure 3. Bifurcation diagram for the core-periphery model

More precisely, when $\Upsilon > \Upsilon_{sustain}$, the economy necessarily involves dispersion. When $\Upsilon < \Upsilon_{break}$, agglomeration always arises, the winning region depending on the initial conditions. Finally, when $\Upsilon_{break} \leq \Upsilon \leq \Upsilon_{sustain}$, both agglomeration and dispersion are stable equilibria. In this domain, the economy displays some hysteresis because dispersion (agglomeration) still prevails when transport costs rise above the sustain point (fall below the break point) while staying below the break point (above the sustain point).

Summarizing these results, when transportation costs are sufficiently low, all manufactures are concentrated in a single region which becomes the core of the economy, while the other region, called the periphery, supplies only the traditional good. Firms in the modern sector are able to exploit increasing returns by selling more in the large market without losing much business in the small market. For exactly the opposite reason, the economy displays a symmetric regional pattern of production when transportation costs are large. Hence, this model allows for *the possibility of divergence between regions*, whereas the neoclassical model, based on constant returns and perfect competition in the two sectors, would predict symmetry only.

3.3 A linear model of core-periphery

The conclusions derived in Section 3.2 are very important for the space-economy. This is why it is crucial to know how they depend on the specificities of the framework employed. The use of both the CES utility and iceberg cost leads to a convenient setting in which demands have a constant elasticity. However, such a result conflicts with research in spatial pricing theory in which demand elasticity is shown to vary with distance. Moreover, if using the iceberg cost is able to capture the fact that shipping is resource-consuming, such a modeling option implies that any increase in the mill price is accompanied with a proportional increase in transport cost, which seems unrealistic. Last, although models of the type considered in the foregoing are based on very specific assumptions, they are often beyond the reach of analytical resolution.

The setting considered here, which has been developed by Ottaviano, Tabuchi and Thisse (2002), is very similar to that used in Section 3.2. However, there are two major differences. First, the output of the $\mathbb{M}$-sector is traded at a cost of τ units of the numéraire per unit shipped between regions. This characteristic agrees more with reality as well as with location theory than the iceberg technology does. Second, preferences are given by a quasi-linear utility encapsulating a quadratic subutility instead of a Cobb-Douglas preference on the homogenous and differentiated goods with CES subutility. These two specifications correspond to rather extreme cases: the former assumes an infinite elasticity of substitution between the differentiated product and the numéraire, the latter a unit elasticity. Moreover, firms' demands are linear and not iso-elastic. Despite such major differences in settings, we will see that conclusions are qualitatively the same in the two models, thus suggesting that they hold for a whole class of models.

3.3.1 A model with quadratic utility and linear transport costs

Preferences are identical across individuals and described by a quasi-linear utility with a quadratic subutility which is supposed to be symmetric in all varieties:

$$\begin{aligned} u(q_0; q(i), i \in [0, M]) &= \alpha \int_0^M q(i)di - (\beta - \delta) \int_0^M [q(i)]^2 di \\ &\quad -\delta \left[\int_0^M q(i)di \right]^2 + q_0 \end{aligned} \tag{20}$$

where $q(i)$ is the quantity of variety $i \in [0, M]$ and q_0 the quantity of a homogenous good chosen as the numéraire. The parameters in (20) are such that $\alpha > 0$ and $\beta > \delta > 0$. In this expression, α expresses the intensity of preferences for the differentiated product, whereas $\beta > \delta$ means that consumers' preferences exhibit love of variety. Finally, for a given value of β, the parameter δ expresses the substitutability between varieties: the higher δ, the closer substitutes the varieties.

Admittedly, a quasi-linear utility abstracts from general equilibrium income effects and gives the corresponding framework a fairly strong partial equilibrium flavor. However, it does not remove the interaction between product and labor markets, thus allowing us to develop a full-fledged model of agglomeration formation, independently of the relative size of the manufacturing sector.

Any individual is endowed with one unit of labor (of type H or L) and $\overline{q}_0 > 0$ units of the numéraire. Her budget constraint can then be written as follows:

$$\int_0^M p(i)q(i)di + q_0 = y + \overline{q}_0$$

where y is the individual's labor income and $p(i)$ the price of variety i. The initial endowment $\overline{q}_0$ is supposed to be large enough for the residual consumption of the numéraire to be strictly positive for each individual. Hence, individual demand $q(i)$ for variety i is given by:

$$q(i) = a - (b + dM)\, p(i) + dP \tag{21}$$

where

$$P \equiv \int_0^M p(i)di$$

which can be interpreted as the price index in the modern sector, while $a \equiv 2\alpha/[(\beta+(M-1)\delta]$, $b \equiv 1/[\beta+(M-1)\delta]$ and $d \equiv \delta/(\beta-\delta)[\beta+(M-1)\delta]$.

Finally, each variety can be traded at a positive cost of τ units of the numéraire for each unit transported from one region to the other, regardless of the variety. The technologies are the same as in Section 3.1 but, for simplicity, c is set equal to zero in (1).

Labor market clearing implies that the numbers of firms belonging to the $\mathbb{M}$-sector in region r is:

$$M_r = \lambda_r H/f \tag{22}$$

Consequently, the total number of firms in the economy is constant and equal to $M = H/f$.

Discriminatory and mill pricing are no longer equivalent in this model. In the sequel, we focus on discriminatory pricing because this policy endows firms with flexibility in their price choice, something that could affect the process of agglomeration. This means that each firm sets a delivered price specific to each region. Hence the profit function of a firm located in region r is as follows:

$$\pi_r = p_{rr}q_{rr}(p_{rr})(L/2 + \lambda_r H) + (p_{rs} - \tau)q_{rs}(p_{rs})(L/2 + \lambda_s H) - fw_r$$

To illustrate the type of interaction that characterizes this model of monopolistic competition, we describe how the equilibrium prices are determined. Each firm i in region r maximizes its profit π_r, assuming accurately that its price choice has no impact on the regional price indices

$$P_r \equiv \int_0^{M_r} p_{rr}(i)di + \int_0^{M_s} p_{sr}(i)di \qquad s \neq r$$

Since, by symmetry, the prices selected by the firms located within the same region are identical, the result is denoted by $p^*_{rr}(P_r)$ and $p^*_{rs}(P_s)$. Clearly, it must be that

$$M_r p^*_{rr}(P_r) + M_s p^*_{sr}(P_r) = P_r$$

Given (22), it is then readily verified that the equilibrium prices are as follows:

$$p^*_{rr} = \frac{1}{2}\frac{2a + \tau d\lambda_r M}{2b + dM} \tag{23}$$

$$p^*_{rs} = p_{ss} + \frac{\tau}{2} \tag{24}$$

Clearly, these prices depend directly on the firms' distribution. In particular, p^*_{rr} decreases with the number of firms in region r and increases with the degree of product differentiation when τ is sufficiently small for the demands of the imported varieties to be positive. These results agree with what we know from standard models of product differentiation.

It is easy to check that the equilibrium operating profits earned in each market by a firm established in r are as follows:

$$\begin{aligned} \pi^*_{rr} &= (b + dM)(p^*_{rr})^2(L/2 + \lambda_r H) \\ \pi^*_{rs} &= (b + dM)(p^*_{rs} - \tau)^2(L/2 + \lambda_s H) \end{aligned}$$

Increasing λ_r has two opposite effects on π^*_{rr}. First, as λ_r rises, the equilibrium price (23) falls as well as the quantity of each variety bought by each consumer living in region r. However, the total population of consumers residing in this region is now larger so that the profits made by a firm located in r on local sales may increase. What is at work here is *a global demand effect due to the increase in the local population that may compensate firms for the adverse price effect as well as for the decrease in each worker's individual demand.*

Entry and exit are free so that profits are zero in equilibrium. Hence, (22) implies that any change in the population of workers located in one region must be accompanied by a corresponding change in the number of firms. The equilibrium wage rates w^*_r of the skilled are obtained from the zero profit condition evaluated at the equilibrium prices: $w^*_r(\lambda_r) = (\pi^*_{rr} + \pi^*_{rs})/f$.

3.3.2 The debate agglomeration vs. dispersion revisited

The indirect utility differential $\Delta v(\lambda)$ is obtained by plugging the equilibrium prices (23)-(24) and the equilibrium wages $w_r^*(\lambda)$ into the indirect utility associated with (20):

$$\Delta v(\lambda) \equiv v_A(\lambda) - v_B(\lambda) = C^* \tau(\tau^* - \tau)(\lambda - 1/2) \tag{25}$$

where C^* is a positive constant and

$$\tau^* \equiv \frac{4af(3bf + 2dH)}{2bf(3bf + 3dH + dL) + d^2H(L + H)} > 0 \tag{26}$$

It follows immediately from (25) that $\lambda = 1/2$ is always an equilibrium. Moreover, because $\Delta v(\lambda)$ is linear in λ and $C^* > 0$, for $\lambda \neq 1/2$ the indirect utility differential has always the same sign as $\lambda - 1/2$ if and only if $\tau < \tau^*$; if $\tau > \tau^*$, it has the opposite sign. In particular, *when there are no increasing returns in the manufacturing sector* ($f = 0$), the coefficient of $(\lambda - 1/2)$ is always negative because $\tau^* = 0$, and thus *dispersion is the only (stable) equilibrium*. This shows once more the importance of increasing returns for the possible emergence of an agglomeration.[8] The same holds for product differentiation because τ^* becomes arbitrarily small when varieties become less and less differentiated ($d \to \infty$).

It remains to determine when τ^* is sufficiently low for all demands to be positive at the equilibrium prices. This is so if and only if

$$L/H > \frac{6b^2f^2 + 8bdfH + 3d^2H^2}{dH(2bf + dH)} \tag{27}$$

The inequality (27) means that the population of unskilled is large relative to the population of skilled. When (27) does not hold, the coefficient of $(\lambda - 1/2)$ in (25) is always positive for all transport costs that allow for interregional trade. In this case, the advantages of having a large home market always dominate the disadvantages incurred while supplying a distant periphery. The condition (18) plays a role similar to (17).

[8]Sonnenchein (1982) shows, a contrario, a related result: if the initial distribution of firms is uneven along a given circle, then the spatial adjustment of firms in the direction of higher profit leads the economy toward a uniform long-run equilibrium, each local economy being perfectly competitive.

More interesting is the case when (27) holds. Although the size of the industrial sector is captured here through the relative population size L/H and not through its share in consumption, the intuition is similar: the ratio L/H must be sufficiently large for the economy to display different types of equilibria according to the value of τ. This result does not depend on the expenditure share on the manufacturing sector because of the absence of general equilibrium income effects: small or large sectors in terms of expenditure share are agglomerated when τ is small enough.

Finally, stability is studied using (2). When $\tau > \tau^*$, it is straightforward to see that the symmetric configuration is the only stable equilibrium. In contrast, when $\tau < \tau^*$ the symmetric equilibrium becomes unstable and workers agglomerate in region r provided that the initial fraction of workers residing in this region exceeds $1/2$. In other words, agglomeration arises when the transport cost is low enough.

Proposition 4 *Consider a two-region economy with segmented markets.*

(i) When (27) does not hold, the core-periphery structure is the only stable equilibrium under trade.

(ii) When (27) is satisfied, we have: for any $\tau > \tau^*$ the symmetric configuration is the only stable equilibrium with trade; for any $\tau < \tau^*$ the core-periphery pattern is the unique stable equilibrium; for $\tau = \tau^*$ any configuration is an equilibrium.

Because (25) is linear in λ, the break point and the sustain point are the same, and thus history alone matters for the selection of the agglomerated outcome.

Looking at the threshold value τ^* as given by (26), we first observe that τ^* increases with the degree of product differentiation (d falls) when (27) holds. This is intuitively plausible because the agglomeration process is driven by the mobility of the skilled workers, whence their population must be sufficiently large for product differentiation to act as an agglomeration force. Second, higher fixed costs leads to a smaller number of firms/varieties. Still, it is readily verified that τ^* also increases when increasing returns become stronger (f rises) when (27) holds. In other words, the agglomeration of the modern sector is more likely, the stronger are the increasing returns at the firm's level. Last, τ^* increases when the number of unskilled (L) decreases because the dispersion force is weaker.

Both models studied in this sections yield similar results, suggesting that the core-periphery structure is robust against alternative specifications. Each model has its own merit. The former allows for income effects and the latter for a finer description of the role played by the key-parameters of the economy. As will be seen below, both have been used in various extensions of the core-periphery model.

4 Further Topics in Economic Geography

In this section, we present an abbreviated version of a few recent contributions. The interested reader will find the models at greater length in the corresponding references.

4.1 On a $\cap$-shaped relationship between agglomeration and transport costs

The assumption of zero transport costs for the homogenous good is not innocuous. Indeed, introducing positive transport costs for this good leads to some fundamental changes in the results presented above. In order to permit trade of the traditional good even at the symmetric configuration, we assume that this good is differentiated too (e.g. orange in A and apples in B). Thus, T as it appears in (3) is now given by

$$T = (T_A^\eta + T_B^\eta)^{1/\eta}$$

where $0 < \eta < 1$. The numéraire is given by the traditional good in one of the two regions. As shown by Fujita *et al.* (1999), the bifurcation diagram given in Figure 3 changes and is now as in Figure 4. To make things simple, we consider a fixed value for the transport costs of the traditional good and, as before, we concentrate on a decrease in the transport costs in the modern sector. When these costs are high, the symmetric configuration is the only equilibrium. Below some critical value, the core-periphery arises as before.

However, *further reductions in transport costs eventually lead to re-dispersion of the modern sector*. Indeed, the agglomeration of the modern sector within, say, region A generates large imports of the traditional good from region B. When transport costs in the modern sector becomes sufficiently low, the price indices of this good is about the same in the two regions. Then, the relative price of the traditional good in A rises because its transport cost remains unchanged. This in turn lowers region B's nominal wage that guarantees the same utility level in both regions to the skilled. When the transport costs within the modern sector decrease sufficiently, the factor price differential becomes strong enough to induce firms to move away from A to B.

Consequently, as transport costs in the modern sector keep decreasing from high to very low values while transport costs in the traditional sector remain constant, the modern sector is first dispersed, then agglomerated, and re-dispersed, as seen in Figure 4. It is worth stressing that the reasons that lead to dispersion in the first and third phases are different: in the former, the modern sector is dispersed because the cost of shipping its output is high; in the latter, dispersion arises because the periphery develops some comparative advantage in terms of labor cost.

Figure 4. Bifurcation with positive agricultural transport costs

Although transport costs of both types of goods have declined since the beginning of the Industrial Revolution, what matters for the regional distribution of economic activities is not only the absolute levels of transport costs but also their relative values across sectors (Kilkenny, 1998). For example, if both costs decrease proportionally, it can be shown that re-dispersion never occurs. This is not surprising because there is no force creating wage differential any more. However, if agricultural transport costs decrease at a lower pace than those of manufacturing goods, cheaper rural labor should eventually attract industrial firms, whereas the reversal in the relationship between transport costs has the opposite impact (see Fujita *et al.*, 1999, Section 7.4 for more details).

The pattern dispersion/agglomeration/re-dispersion also arises as long as we consider any ingredient giving rise to factor price differentials in favor of the periphery. For example, if we assume that the agglomeration of the modern sector in one region generates higher urban costs, such as land rent and commuting costs, a sufficiently strong decrease in transport costs between regions will foster re-dispersion when firms located in the core region have to pay high wages to their workers. This occurs because workers must be compensated for the high urban costs associated with a large concentration of people within the same urban area (Helpman, 1998; Tabuchi, 1998; Ottaviano *et al.*, 2002). Another example is when all workers are immobile, whereas agglomeration of the industrial sector may arise because of technological linkages with the intermediate sector (more on this below). In this case, wage in the core region may become so high that re-dispersion is profitable for firms (Krugman and Venables, 1995; Puga, 1999).

4.2 Welfare implications of the core-periphery structure

We now wish to determine whether or not agglomeration is efficient. To this end, we assume that the planner is able (i) to assign any number of workers (or, equivalently, of firms) to a specific region and (ii) to use lump sum transfers from all workers to pay for the loss firms may incur while pricing at marginal cost. Because utilities are quasi-linear in the model of Section 3.3, a utilitarian approach may be used to evaluate the global level of welfare (Ottaviano and Thisse, 2002). Observe that no distortion arises in the total number of varieties Because N is determined by the factor endowment (H) and technology (f) in the modern sector and is, therefore, the same at both the equilibrium and optimum outcomes.

Because the setting assumes transferable utility, the planner chooses λ to maximize the sum of individual indirect utilities $W(\lambda)$ (for both types of workers) in which all prices have been set equal to marginal cost. It can be shown that

$$W(\lambda) = C^o \tau(\tau^o - \tau)\lambda(\lambda - 1) + \text{constant} \tag{28}$$

where C^o is a positive constant and

$$\tau^o \equiv \frac{4af}{2bf + d(H + L)}$$

The welfare function (28) is strictly concave in λ if $\tau > \tau^o$ and strictly convex if $\tau < \tau^o$. Furthermore, because the coefficients of λ^2 and of λ are the same (up to their sign), this expression has always an interior extremum at $\lambda = 1/2$. As a result, the optimal choice of the planner is determined by the sign of the coefficient of λ^2, that is, by the value of τ with respect to of τ^o: if $\tau > \tau^o$, the symmetric configuration is the optimum; if $\tau < \tau^o$ any agglomerated configuration is the optimum; if $\tau = \tau^o$ the welfare level is independent of the spatial configuration.

In accordance with intuition, it is efficient to agglomerate the modern sector into a single region once transport costs are low, increasing returns are strong enough and/or the output of this sector is sufficiently differentiated.

On the other hand, the optimum is always dispersed when increasing returns vanish ($f = 0$) and/or when varieties are close substitutes (d is large).

A simple calculation shows that $\tau^o < \tau^*$. This means that *the market yields an agglomerated configuration for a range* ($\tau^o < \tau < \tau^*$) *of transport cost values for which it is efficient to have a dispersed pattern of activities.* In contrast, when transport costs are low ($\tau < \tau^o$) or high ($\tau > \tau^*$), no regional policy is required from the efficiency point of view, although equity considerations might justify such a policy when agglomeration arises. On the contrary, for intermediate values of transport costs ($\tau^o < \tau < \tau^*$), the market provides excessive agglomeration, thus justifying the need for an active regional policy to foster the dispersion of the modern sector on both the efficiency and equity grounds.[9]

This discrepancy may be explained as follows. First, workers do not internalize the negative external effects they impose on the unskilled who stay put, nor do they account for the impact of their migration decisions on the residents in their region of destination. Hence, even though the skilled have individual incentives to move, these incentives do not reflect the social value of their move. This explains why equilibrium and optimum do not necessarily coincide. Second, the individual demand elasticity is much lower at the optimum (marginal cost pricing) than at the equilibrium (Nash equilibrium pricing), and thus regional price indices are less sensitive to a decrease in τ. As a result, the fall in trade costs must be sufficiently large to make the agglomeration of workers socially desirable; this tells us why $\tau^o < \tau^*$.

4.3 On the impact of forward-looking behavior

In the dynamics used in Section 3, workers care only about their current utility level. This is a fairly restrictive assumption to the extent that migration decisions are typically made on the grounds of current and future utility flows and costs (such as search, mismatch and homesickness). In addition, this approach has been criticized because it is not consistent with fully rational forward-looking behavior. It is, therefore, important to determine if and how workers' expectations about the evolution of the economy may influence the process of agglomeration. In particular, we are interested in identifying the conditions under which, when initially the two regions host

[9]Observe that the same qualitative results hold for a second best analysis in which firms price at the Nash equilibrium while the planner controls their locations (Ottaviano and Thisse, 2002).

different numbers of skilled workers, the common belief that these workers will eventually agglomerate in the currently smaller region can reverse the historically inherited advantage of the larger region.

Formally, we want to determine the parameter conditions for which there exists an equilibrium path consistent with this belief, assuming that workers have perfect foresight (*self-fulfilling prophecy*). Somewhat different approaches have been proposed to tackle this problem, but they yield similar conclusions (Baldwin, 2001; Ottaviano, 1999; Ottaviano *et al.*, 2002). In what follows, we use the model of Section 3.3 because it leads to a linear dynamic system that allows for a detailed analysis of the main issues (Krugman, 1991b; Fukao and Bénabou, 1993).

Workers live indefinitely with a rate of time preference equal to $\gamma > 0$. Because we wish to focus on the sole dynamics of migration, we assume that the consumption of the numéraire is positive for each point in time so that there is no intertemporal trade in the differentiated good. For concreteness, consider the case in which workers expect agglomeration to occur in region A while region B is initially larger than A. Formally, we assume that there exists $T \geq 0$ such that, given $\lambda_0 < 1/2$,

$$\begin{array}{rclr} \dot{\lambda}(t) & > & 0 & t \in [0, T) \\ \lambda(t) & = & 1 & t \geq T \end{array} \tag{29}$$

Because workers have perfect foresights, the easiest way to generate a non bang-bang migration behavior is to assume that, when moving from one region to the other, workers incur a utility loss that depends on the rate of migration, perhaps because a migrant imposes a negative externality on the others. Specifically, we assume that the cost $CM(t)$ borne by a migrant at time t is proportional to the corresponding migration flow:

$$CM(t) \equiv \left| \frac{d\lambda(t)}{dt} \right| / \delta \tag{30}$$

where δ is a positive constant whose meaning is given below.

For each region $r = A, B$, let us define

$$V_r(t) \equiv \int_t^T e^{-\gamma(s-t)} v_r(s) ds + e^{-\gamma(T-t)} v_A(T)/\gamma \qquad t \in [0, T) \tag{31}$$

where $v_r(s)$ is the instantaneous indirect utility at time s in region r. By definition, for $r = A$, $V_A(t)$ is the discounted sum of utility flows of a worker who moves from B to A at time t (i.e., today), while for $r = B$, $V_B(t)$ is that of a worker who currently resides in B and plans to move to A at time T. Because workers are free to choose when to immigrate, in equilibrium they must be indifferent about the time t at wich they move. Hence, at any $t < T$ the following equality must hold:

$$V_A(t) - CM(t) = V_B(t) - e^{-r(T-t)}CM(T).$$

Furthermore, because no worker residing currently in B wishes to postpone his migration time beyond T, it must be that $CM(T) = 0$ (Fukao and Bénabou, 1993), and thus

$$V_A(t) - CM(t) = V_B(t) \qquad t \in [0, T).$$

Using (**??**) and (30), we then obtain

$$\frac{d\lambda}{dt} = \delta \Delta V \qquad t \in [0, T) \tag{32}$$

where $\Delta V \equiv (V_A - V_B)$ and δ can be interpreted as the speed of adjustment. This means that the private marginal cost of moving equals its private marginal benefit at any time $t < T$; of course, $\lambda(T) = 1$.

Using (31), we obtain the second law of motion by differentiating $V_A(t) - V_B(t)$, thus yielding

$$\frac{d\Delta V}{dt} = \gamma \Delta V - \Delta v \qquad t \in [0, T) \tag{33}$$

where $\Delta v \equiv v_A - v_B$ stands for the instantaneous indirect utility differential flow given by (25). The expression (33) states that the "annuity value" of being in A rather than in B (i.e. $\gamma \Delta V$) equals the "dividend" (Δv) plus the "capital gain" ($d\Delta V/dt$). As a result, because (25) is linear in λ, we obtain a system of two differential equations instead of one.

The system (32)-(33) has always a steady state at $(\lambda, \Delta V) = (1/2, 0)$ that corresponds to the symmetric configuration. When $\tau > \tau^*$ this steady state is globally stable. So, for the assumed belief (29) to be consistent with equilibrium, it must be $\tau < \tau^*$. Then, the study of the eigenvalues of the system (32)-(33) shows that two cases may arise. In the first one, when workers' migration costs are sufficiently large (δ is such that $\gamma > 2\sqrt{C\delta\tau(\tau^* - \tau)}$), the outcome of the migration dynamics is the same as the one described in Section 3.3. In other words, the equilibrium path is not consistent with (29), thus implying that expectations do not matter.

By contrast, when migration costs are small enough ($\gamma < 2\sqrt{C\delta\tau(\tau^* - \tau)}$), expectations may matter. More precisely, there exist two threshold values for the transport costs $\tau_1 < \tau^*/2 < \tau_2 < \tau^*$, as well as two boundary values $\lambda_1 < 1/2 < \lambda_2 < 1$ such that the equilibrium path is consistent with (29) if and only if $\tau \in (\tau_1, \tau_2)$ and $\lambda_0 \in [\lambda_1, \lambda_2]$. Namely, *as long as obstacles to trade take intermediate values and regions are not initially too different, the region that becomes the core is determined by workers' expectations.* This is more so either the lower the migration costs or the lower the discount rate.

4.4 The impact of a heterogenous labor force

So far, workers have been assumed to be identical in terms of preferences. Although this assumption is fairly standard in economic modeling, it seems highly implausible that potentially mobile individuals will react in the same way to some "gap" between regions. First of all, it is well known that some people show a high degree of attachment to the region in which they are born. They will stay put even though they may guarantee to themselves higher living standards in other places. In the same spirit, life-time considerations such as marriage, divorce and the like play an important role in the decision to migrate. Second, regions are not similar and exhibit different natural and cultural features. Clearly, people value differently local amenities and such differences in attitudes are known to affect the migration process.

These considerations are fundamental ingredients of the migration process and should be accounted for explicitly in workers' preferences. Even though the personal motivations may be quite diverse and, therefore, difficult to model at the individual level, it is possible to identify their aggregate impact on the spatial distribution of economic activities by using discrete choice theory, in much the same way that consumer preferences for differentiated products are modeled (Anderson *et al.*, 1992). Specifically, we assume that the "matching" of workers' with regions is expressed through the *logit* (McFadden, 1974). This assumption turns out to be empirically relevant in migration modeling (see, e.g. Anderson and Papageorgiou, 1994), whereas it is analytically convenient without affecting the qualitative nature of the main results. Then, the probability that a worker will choose to reside in region r is given by

$$p_r(\lambda) = \frac{\exp[v_r(\lambda)/\upsilon]}{\exp[v_A(\lambda)/\upsilon] + \exp[v_B(\lambda)/\upsilon]}$$

where υ expresses the dispersion of individual tastes: the larger υ, the more heterogenous the responsiveness of workers to living standards differences $\Delta v(\lambda)$ given by (25).[10]. When $\upsilon = 0$, the living standard response is overwhelming and workers relocates until standards of living are equal in the two regions; when $\upsilon \to \infty$ mobility responds only to amenity differentials and the probability of moving is exogenous with respect to living standards.

In the present setting, it should be clear that the population of workers changes according to the following equation of motion:

$$\begin{aligned} \frac{d\lambda}{dt} &= (1-\lambda)p_B(\lambda) - \lambda p_A(\lambda) \\ &= \frac{1-\lambda}{1+\exp[-\Delta V(\lambda)/\upsilon]} - \frac{\lambda}{1+\exp[\Delta V(\lambda)/\upsilon]} \end{aligned} \tag{34}$$

in which the first term in the right hand side of (34) stands for the fraction of people migrating into region A, whereas the second term represents those leaving this region for region B.

Using Theorem 5 by Tabuchi (1986), it is then readily verified that, for sufficiently large values of υ, there exists a unique stable equilibrium in which the manufacturing sector is equally distributed between regions. Otherwise, there exist two stable equilibria involving each partial agglomeration of the manufacturing sector in one region, whereas dispersion arises for very low values of these costs. As expected, *taste heterogeneity prevents the emergence of a fully agglomerated equilibrium and favors the dispersion of activities.*[11]

4.5 Intermediate sector and industrial agglomeration

In the models above, agglomeration is the outcome of a circular causation process in which more workers concentrate within the same region because

[10]Alternately, it could be evaluated at $\Delta\omega(\lambda)$ which is defined in Section 3.

[11]See Tabuchi and Thisse (2002) for more details.

they love variety. However, if workers are immobile, no agglomeration can arise. Instead, each region specializes in the production of differentiated varieties on the basis of their initial endowments and intra-industry trade occurs for all values of the transport costs.

However, the agglomeration of industries is a pervasive phenomenon even when labor is sticky (e.g. between countries). Venables (1996) suggests that an alternative explanation is to account for the fact that the modern sector uses an array of differentiated intermediate goods. In this case, *the agglomeration of the final sector in a particular region may occur because of the concentration of the intermediate industry in that region makes the final sector more productive; and vice versa.* Evidence reveals, indeed, the importance of the proximity of high-quality business services for the economic success of an urban area (Kolko, 1999).

Workers being immobile, we may consider a single type of labor. Because its output is taken as homogenous, the $\mathbb{M}$-sector is assumed to operate under constant returns to scale and perfect competition. The $\mathbb{M}$-good is produced according to the production function:

$$X^{\mathbb{M}} = l^{1-\alpha} I^{\alpha} \qquad 0 < \alpha < 1$$

where

$$I = \left\{ \int_0^M [q(i)]^{\rho} di \right\}^{1/\rho} \qquad 0 < \rho < 1$$

is the composite input made of the differentiated intermediate goods and l the quantity of labor. Then, the agglomeration of the intermediate and final sectors into the same region is an equilibrium if and only if the following two conditions are satisfied:

$$\Upsilon^{\mathbb{I}} \geq \left(\frac{\mu}{1-\mu} \right)^{(1-\alpha)/\alpha} (\Upsilon^{\mathbb{M}})^{1/\alpha} \tag{35}$$

$$\Upsilon^{\mathbb{I}} \geq \left(\frac{\mu}{1-\mu} \right)^{1/\rho} \tag{36}$$

where $\Upsilon^{\mathbb{M}}$ ($\Upsilon^{\mathbb{I}}$) stands for the transport costs of the final sector (intermediate sector) good and α the cost share of the intermediate good in the final sector (Fujita and Thisse, 2002).

Hence, *when the transport cost of the intermediate goods is high relative to the transport cost of the final good, there is complete regional specialization* in that the final and intermediate sectors are entirely concentrated in region r, whereas the traditional sector operates only in region s. Condition (35) becomes less stringent as the transport cost of the final good declines. In addition, the transport cost of the intermediate goods must also exceed some threshold value (36) because $\mu/(1-\mu) \geq 1$. Clearly, this threshold rises when the intermediate goods are more differentiated.

Condition (35) means that the $\mathbb{M}$-sector does not find it profitable to start operating in region s because importing the intermediate goods from r turns out to be costly due to high transport costs; by contrast, exporting its output from r to s is not costly because there is no restriction on $\Upsilon^{\mathbb{M}}$. Condition (36) means that no firm of the intermediate sector wants to set up in region s because it has to export all its production to region r at a high transport cost. It should be stressed that both sectors are trapped within the same region even when shipping the final good becomes cheaper and cheaper ($\Upsilon^{\mathbb{M}}$ approaches 1).

In order to break the core region, the transport costs of the intermediate goods must fall below some critical value. This is not necessarily easy to achieve when the provision of specific intermediate goods requires face-to-face contacts as for highly differentiated services (in which case $\Upsilon^{\mathbb{I}}$ is high). This provides some clue why the industrial sector is so much concentrated in many developing countries: in such economies, the transport costs of intermediate goods are often quite high due to poor transport and communication infrastructure.

Furthermore, as long as (35) and (36) hold, μ can rise, thus generating a widening wage gap between the core region and the periphery. It is readily verified that the real wage gap, in turn, becomes even larger. This agrees with the observation that (35) becomes less and less stringent as the role of the intermediate goods plays a growing role in the economy (α rises). By contrast, we see that the modern sector is likely to decentralize some of its activities in the periphery as its share in consumption increases.[12]

[12] See Puga (1999) for a framework unifying Krugman's and Venables' settings.

4.6 On the formation of an urban hierarchy

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 Henderson (1974) who proposes a very original approach to the formation of systems of cities. His work is based on Mills (1967) who supposes that the production of a good involves increasing returns and takes place in the Central Business District. Each city then has a finite size because of 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 argues that each city has an incentive to specialize in the production of one final good to export because combining production of different goods within the same city rises commuting costs and land rents. Because the production of different tradables involve different degrees of scale economies, cities end up with different types and sizes. This approach explains the existence of an urban system with different types of cities and of inter-city trade involving different goods.[13] However, this model does not permit to predict the location of cities nor does it explain the urban hierarchical structure.

To this effect, Fujita, Krugman and Mori (1999) introduce into the monopolistic competition model of Section 3.2 different groups of final goods, having each different elasticities of substitution and/or transportation rates. Specifically, the utility (3) becomes:

$$u = T^{\mu_T} \prod_{k=1}^{K} (Q_k)^{\mu_k} \qquad 0 < \mu_T,\ \mu_k < 1$$

where

$$Q_k = \left[\int_0^{M_k} m_k(i)^{\rho_k} di \right]^{1/\rho_k} \qquad 0 < \rho_k < 1$$

stands for an index of the consumption of the k-product's varieties. Furthermore, the location space is now described by the real line. Labor is

[13] See Henderson (1987) for further developments.

homogenous and workers are free to work either the modern or the traditional sector. Finally, the use of land by the traditional sector (agriculture) creates the dispersion force.

Assuming that trade costs are equal across manufactured goods and that

$$\sum_{k=1}^{K} \frac{\mu_k}{\rho_k} \geq 1$$

a condition that boils down to (17) when $K = 1$.As the population rises, Fujita *et al.* show that a (more or less) regular hierarchical central place system reminiscent of Christaller emerges within the economy, in which *"higher-order cities" provide a larger number of groups of final goods, whereas "lower-order cities" produce a smaller number of goods.* Put simply, the equilibrium is consistent with an urban system involving large and diversified cities along with small and specialized cities, a well-documented empirical fact (Duranton and Puga, 2000). In this setting, there is two-way trade between cities because cities supply differentiated goods. This leads to a more intricate pattern of trade in which horizontal relationships are superimposed on the pyramidal structure of central places theory. As expected, higher-order cities export more varieties than lower-order cities. However, horizontal relationships between cities of the same order may be more important than trade with lower-order cities, so that the resulting urban hierarchy is more fuzzy than in the Christaller model of central places.

The pattern of specialization and trade obtained by Fujita *et al.* seems to fit well the description provided by geographers (Pred, 1966) as well as by historians (Hohenberg and Lees, 1985) about the economic space emerging in the US in the 20th century and in Europe during the 19th century. It combines both *the hierarchy of various centers* with *the existence of networks of cities* exchanging specialized goods and services.

5 Suggestions for Future Research and Policy Implications

At first glance, the economics of agglomeration looks like a collection of examples. And, indeed, this new strand of literature has so far sacrificed generality for tractability. However, it is fair to say that this topic is fraught with most of the difficulties encountered in economic theory: nonconvexities (the locational indivisibility of economic agents and increasing returns), imperfect competition (which form?), externalities, and general interdependence. To a large extent, advances in economic geography depends on the ability of the economics profession to come up with more general models involving imperfectly competitive markets. One possible direction for future research would be to investigate a broader class of models of monopolistic competition in which interactions between firms are weak, although there are several good reasons to believe that competition in space is strategic (Gabszewicz and Thisse, 1986). However, it is interesting to notice that the results presented in Section 3 have the same flavor as those obtained in spatial competition when firms sell differentiated products (Anderson *et al.*, 1992; de Palma *et al.*, 1985). Another major line of research to be explored is the nature and working of local interactions. If assuming the existence of local externalities is an acceptable proxy to study their spatial consequences, this is clearly inadequate for a detailed study of the phenomenon of agglomeration. Instead, one should study the various and contrasted social processes arising among agents set up within the same locale. The main engines of economic growth and local development are likely there (Presscott, 1998; Lucas, 2001).

The economic geography approach reviewed in this paper is by no means the only tool useful for understanding the shaping of the space-economy. It is our contention, however, that the qualitative results presented in Section 3 are fairly robust and representative of general tendencies at work in contemporary economies (Fujita and Thisse, 2002). Furthermore, economic geography has strong connections with several branches of modern economic theory, including industrial organization and urban economics, but also with the new theories of international trade and of economic growth and development. Cross-fertilization can therefore be expected. As seen in Section 4, the basic models show a great versatility that has already allowed them to be extended in various directions.

Finally, models of economic geography offer testable predictions so that, eventually, their validity will be an empirical issue. So far very few attempts have been made. To the best of our knowledge, one of the most elaborated studies has been conducted by Combes and Lafourcade (2001) who perform a structural estimation of a multi-regional, multi-sectorial model with vertical linkages. More precisely, they consider 71 industrial sectors and 341 employment areas in France; transport costs are evaluated by means of a distance and time cost function in 1978 and 1993, built from the road net-work, gas price, highways tolls and carriers' contracts. Their work shows that a decrease of about 40% in road transport costs was associated with a strengthening in regional specialization and inequalities. Clearly, more work remains to be done to confirm (or invalidate) such conclusions.

The results presented here, together with others discussed in Fujita and Thisse (2002), suggest some policy implications that are summarized below.

Modern economies encapsulate a strong system of forces pushing toward more agglomeration in economic activities. What makes these forces so powerful is the combination of a drastic fall in transport and trade costs, which combines with the cumulative nature of the agglomeration process. This gives rise to a new type of economic geography in which space is "slippery", whereas locations are "sticky".

There is a risk of excessive agglomeration if the skilled labor force keeps getting more mobile. Yet, one would go too far in predicting that the space-economy will be much more polarized than what it is today. Urban systems are characterized by a strong inertia that favors dispersion. In addition, the growing concentration of activities in a few large regions is likely to be accompanied with higher urban costs that will make these regions eventually less attractive. Finally, even though innovative activities often benefit from being agglomerated, firms are likely to be attracted by cheaper areas when technologies are well monitored, thus offering a *niche* to less diversified areas that can specialize in the production of specific goods. In this perspective, many cities would do well by improving their provision of public goods and services used directly by firms and by cooperating more with their hinterland.

Local clusters may emerge in very different places, thus opening the door to possible local development within depressed regions. However, one should resist to the temptation of planning and organizing such clusters from above. Indeed, they often rest on informal processes such as discussions among workers within firms, inter-firm mobility of skilled workers, exchange of ideas within families or clubs, and bandwagon effects. The proliferation of externalities within cities leads Anas, Arnott and Small (1998, p.1458) to conclude as follows: "only very comprehensive and detailed planning can overcome the resulting inefficiencies. Because the externalities are so poorly understood, however, attempted cures may well do more harm than the disease". The situation is very similar when we come to the case of regional clusters, although the nature of externalities to take into account is likely to be different.

Still, there is a lot to be learned from the many successful experiences undertaken. Among other things, they concur in saying that the efficiency and quality of local institutions that facilitate communication and social coordination are critical in successful local development stories. This is a far too much neglected factor in development plans designed for lagging regions. For example, the European Commission should be more active in detecting such inefficiencies and in making its regional aids conditional upon significant improvements in local (nonmarket) institutions.

References

[1] Anas A., R. Arnott and K.A. Small (1998) Urban spatial structure, *Journal of Economic Literature* 36, 1426-64.

[2] Anderson S.P., de Palma A. and J.-F. Thisse (1992) *Discrete Choice Theory of Product Differentiation,* Cambridge (Mass.), MIT Press.

[3] Anderson W.P. and Y.Y. Papageorgiou (1994) An analysis of migration streams for the Canadian regional system, 1952-1983. 1. Migration probabilities, *Geographical Analysis* 26, 15-36.

[4] Audretsch D.B. and M.P. Feldman (1996) R&D spillovers and the geography of innovation and production, *American Economic Review* 86, 630-40.

[5] Baldwin R.E. (2001) Core-periphery model with forward-looking expectations, *Regional Science and Urban Economics* 31, 21-49.

[6] Bonanno G. (1990) General equilibrium theory with imperfect competition, *Journal of Economic Surveys* 4, 297-328.

[7] Combes P.-Ph. and M. Lafourcade (2001), Transportation cost decline and regional inequalities: evidence from France, CEPR Discussion Paper 2894.

[8] d'Aspremont C., J.J. Gabszewicz and J.-F. Thisse (1979) On Hotelling's "Stability in Competition", *Econometrica* 47, 1045-50.

[9] de Palma A., V. Ginsburgh, Y.Y. Papageorgiou and J.-F. Thisse (1985), The principle of minimum differentiation holds under sufficient heterogeneity, *Econometrica* 53, 767-81.

[10] Dixit A.K. and J.E. Stiglitz (1977) Monopolistic competition and optimum product diversity, *American Economic Review* 67, 297-308.

[11] Duranton G. and D. Puga (2000) Diversity and specialization in cities: why, where and when does it matter? *Urban Studies* 37, 533-55.

[12] *The Economist* (1995) A survey of cities, July 29.

[13] Ellickson B. and W. Zame (1994) Foundations for a competitive theory of economic geography, University of California at Los Angeles, Department of Economics, memo.

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

[15] Fujita M. (1989) *Urban Economic Theory. Land Use and City Size*, Cambridge, Cambridge University Press.

[16] Fujita M., P. Krugman and T. Mori (1999) On the evolution of hierarchical urban systems, *European Economic Review* 43, 209-51.

[17] Fujita, M., P. Krugman and A.J. Venables (1999) *The Spatial Economy. Cities, Regions and International Trade*, Cambridge (Mass.), MIT Press.

[18] Fujita M. and H. Ogawa (1982) Multiple equilibria and structural transition of non-monocentric urban configurations, *Regional Science and Urban Economics* 12, 161-96.

[19] Fujita M. and J.-F. Thisse (2002) *Economics of Agglomeration*, Cambridge, Cambridge University Press.

[20] Fukao K. and R. Bénabou (1993) History versus expectations: a comment, *Quarterly Journal of Economics* 108, 535-42.

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

[22] Hall R.E and C.I Jones (1999) Why do some countries produce so much more output per worker than others?, *Quarterly Journal of Economics* 114, 83-116.

[23] Helpman, E. (1998) The size of regions, in: D. Pines, E. Sadka and I. Zilcha, eds., *Topics in Public Economics. Theoretical and Applied Analysis,* Cambridge, Cambridge University Press, 33-54.

[24] Henderson J.V. (1974) The sizes and types of cities, *American Economic Review* 64, 640-56.

[25] Henderson J.V. (1987) Systems of cities and inter-city trade, in: P. Hansen, M. Labbé, D. Peeters, J.-F. Thisse, and J.V. Henderson, *Systems of Cities and Facility Location*, Chur, Harwood Academic Publishers, 71-119.

[26] Hohenberg P. and L.H. Lees (1985) *The Making of Urban Europe (1000-1950)*, Cambridge (Mass.), Harvard University Press.

[27] Hoover E.M. (1948) *The Location of Economic Activity*, New York, McGraw-Hill.

[28] Kaldor N. (1970) The case for regional policies, *Scottish Journal of Political Economy* 17, 337-48.

[29] Kilkenny M. (1998) Transport costs and rural development, *Journal of Regional Science* 38, 293-312.

[30] Kolko J. (1999) Can I get a service here. Information technology, service industries, and the future of cities, Harvard University, memo.

[31] Krugman P.R. (1991a) Increasing returns and economic geography, *Journal of Political Economy* 99, 483-99.

[32] Krugman (1991b) History versus expectations, *Quarterly Journal of Economics* 106, 651-67.

[33] Krugman P. and A.J. Venables (1995) Globalization and the inequality of nations, *Quarterly Journal of Economics* 110, 857-80.

[34] Lucas R.E. (1988) On the mechanics of economic development, *Journal of Monetary Economics* 22, 3–22.

[35] Lucas R.E. (2001) Externalities and cities, *Review of Economic Dynamics* 4, 245-74.

[36] Matsuyama K. (1995) Complementarities and cumulative process in models of monopolistic competition, *Journal of Economic Literature* 33, 701-29.

[37] McFadden D. (1974) Conditional logit analysis of qualitative choice behavior, in: P. Zarembka, ed., *Frontiers in Econometrics*, New York, Academic Press, 105-42.

[38] Mills E.S. (1967) An aggregative model of resource allocation in a metropolitan area, *American Economic Review* 57, 197-210.

[39] Neary J.P. (2001) Of hype and hyperbolas: introducing the new economic geography, *Journal of Economic Literature* 39, 536-61.

[40] Ottaviano G.I.P. (1999) Integration, geography, and the burden of history, *Regional Science and Urban Economics* 29, 245-56.

[41] Ottaviano G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 101-127.

[42] Ottaviano G.I.P. and J.-F. Thisse (2002) Integration, agglomeration and the political economics of factor mobility, *Journal of Public Economics*, forthcoming.

[43] Pred A. (1966) *The Spatial Dynamics of US Urban-industrial Growth*, Cambridge (Mass.), MIT Press.

[44] Presscott E.C. (1998) Needed: a theory of total factor productivity, *International Economic Review* 39, 525-51.

[45] Puga D. (1999) The rise and fall of regional inequalities, *European Economic Review* 43, 303-34.

[46] Richardson H.W. (1987) The costs of urbanization: a four-country comparison, *Economic Development and Cultural Change* 35, 561-80.

[47] Samuelson P. A. (1983) Thünen at two hundred, *Journal of Economic Literature* 21, 1468-88.

[48] Sonnenschein H. (1982) Price dynamics based on the adjustment of firms, *American Economic Review* 72, 1088-96.

[49] SOPEMI (1998) *Trends in International Migration*, Paris, OECD.

[50] Spence M. (1976) Product selection, fixed costs, and monopolistic competition, *Review of Economic Studies* 43, 217-35.

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

[52] Tabuchi T. (1986) Existence and stability of city-size distribution in the gravity and logit models, *Environment and Planning A* 18, 1375–89.

[53] Tabuchi T. (1998) Agglomeration and dispersion: a synthesis of Alonso and Krugman, *Journal of Urban Economics* 44, 333-51.

[54] Tabuchi T. and J.-F. Thisse (2002) Taste heterogeneity, labor mobility and economic geography, *Journal of Economic Development*, forthcoming.

[55] Venables A.J. (1996) Equilibrium locations of vertically linked industries, *International Economic Review* 37, 341-59.

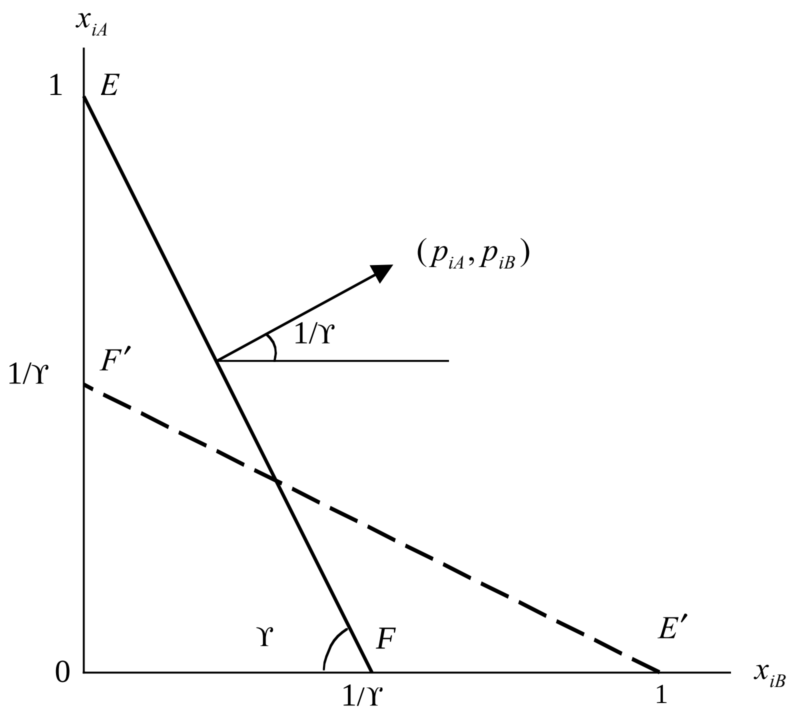

Figure 1. The set of feasible allocations in a homogeneous space

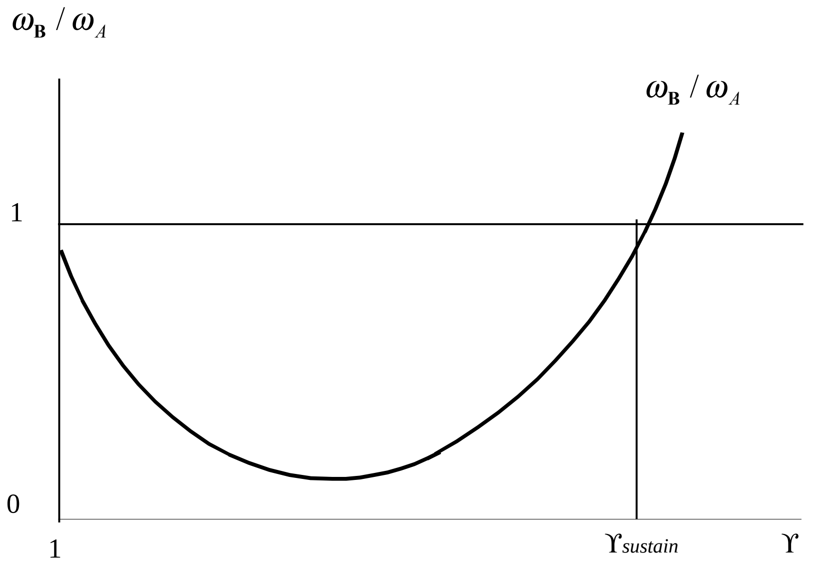

Figure 2 The determination of the sustain point

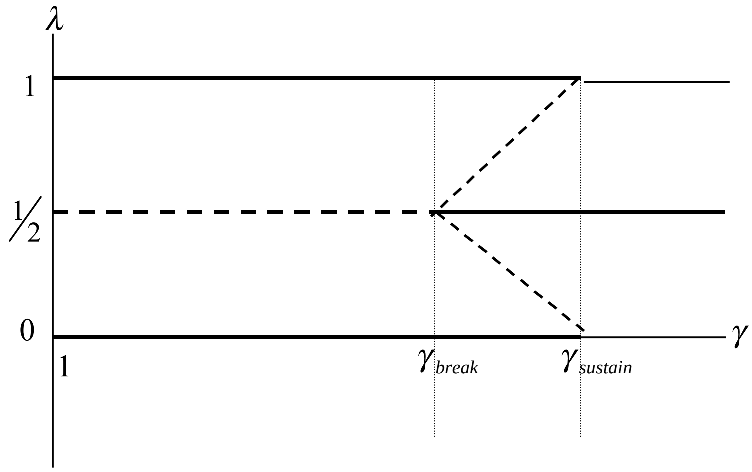

Figure 3. The bifurcation diagram for the core-periphery model

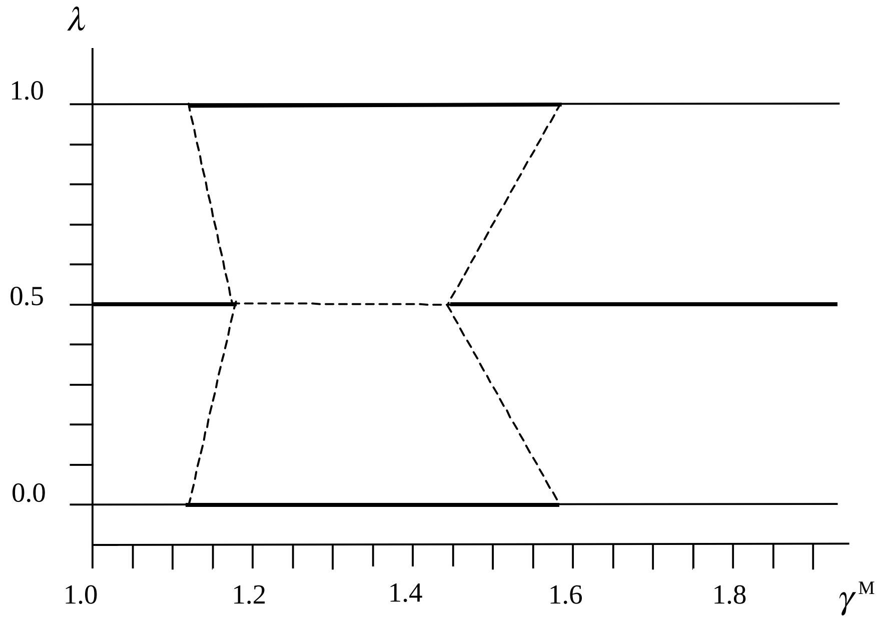

Figure 4. Bifurcation with positive agricultural transport costs