



**Economics Department of the University of Pennsylvania
Institute of Social and Economic Research -- Osaka University**

Agglomeration and Trade Revisited

Author(s): Gianmarco Ottaviano, Takatoshi Tabuchi and Jacques-François Thisse

Source: *International Economic Review*, Vol. 43, No. 2 (May, 2002), pp. 409-435

Published by: [Wiley](#) for the [Economics Department of the University of Pennsylvania](#) and [Institute of Social and Economic Research -- Osaka University](#)

Stable URL: <http://www.jstor.org/stable/826994>

Accessed: 10/09/2013 14:50

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley, Economics Department of the University of Pennsylvania, Institute of Social and Economic Research -- Osaka University are collaborating with JSTOR to digitize, preserve and extend access to *International Economic Review*.

<http://www.jstor.org>

AGGLOMERATION AND TRADE REVISITED*

BY GIANMARCO OTTAVIANO, TAKATOSHI TABUCHI,
AND JACQUES-FRANÇOIS THISSE¹

*Università "L. Bocconi", Italy; University of Tokyo, Japan; CORE, Université
Catholique de Louvain, Belgium, CERAS, Ecole Nationale des Ponts et
Chaussées, France*

The purpose of this article is twofold. First, we present an alternative model of agglomeration and trade that displays the main features of the recent economic geography literature while allowing for the derivation of analytical results by means of simple algebra. Second, we show how this framework can be used to permit (i) a welfare analysis of the agglomeration process, (ii) a full-fledged forward-looking analysis of the role of history and expectations in the emergence of economic clusters, and (iii) a simple analysis of the impact of urban costs on the spatial distribution of economic activities.

1. INTRODUCTION

The agglomeration of activities in a few locations is probably the most distinctive feature of the economic space. Despite some valuable early contributions made by Hirschman, Perroux, or Myrdal, this fact remained unexplained by mainstream economic theory for a long time. It is only recently that economists have become able to provide an analytical framework explaining the emergence of economic agglomerations in an otherwise homogenous space. As argued by Krugman (1995), this is probably because economists lacked a model embracing both increasing returns and imperfect competition, the two basic ingredients of the formation of the economic space, as shown by the pioneering work of Hotelling (1929), Lösch (1940), and Koopmans (1957).

However, even though several modeling strategies are available to study the emergence of economic agglomerations (Fujita and Thisse, 1996), their potential has not been really explored, as recognized by Krugman (1998) himself:

* Manuscript received December 1999; revised February 2000.

¹ We should like to thank four referees for their comments as well as R. Baldwin, G. Duranton, M. Fujita, P. Jehiel, A. Linbeck, K. Middelfart-Knarvik, T. Mori, D. Peeters, P. Picard, D. Pines, K. Stahl, D.-Z. Zeng, participants of seminars at University of Colorado, Kyoto University, IUI in Stockholm, CEPR ERWIT in Rotterdam, SITE in Stanford, ASSET in Bologna, and Kagawa University. The first author is grateful to the EU TMR programme, the second author to the Japan Society for the Promotion of Science through the Future Program, and the first and third authors to FNRS (Belgium) for funding. The revision was made when the first author was Jean Monnet Fellow at EUI Florence (Italy). E-mail: thisse@core.ucl.ac.be.

To date, the new economic geography has depended heavily on the tricks summarized in Fujita et al. (1999) with the slogan “Dixit–Stiglitz, icebergs, evolution, and the computer” (p. 164).

The slogan of the new economic geography is explained by the following methodology (see Fujita et al., 1999): First, the main tool used in the new economic geography is a particular version of the Chamberlinian model of monopolistic competition developed by Dixit and Stiglitz (1977) in which consumers love variety and firms have fixed requirements for limited productive resources (hence, “Dixit–Stiglitz”). Love of variety is captured by a CES utility function that is symmetric in a bundle of differentiated products. Each firm is assumed to be a negligible actor in that it has no impact on overall market conditions. Second, transportation is modeled as a costly activity that uses the transported good itself: In other words, a certain fraction of the good melts on the way (hence, “icebergs”). Taken together, these assumptions yield a demand system in which the own-price elasticities of demands are constant, identical to the elasticities of substitutions, and equal to each other across all differentiated products. This entails equilibrium prices that are independent of the spatial distribution of firms and consumers. Though convenient from an analytical point of view, such a result conflicts with research in spatial pricing theory that shows that demand elasticity varies with distance while prices change with the level of demand and the intensity of competition. Moreover, the iceberg assumption also implies that any increase in the price of the transported good is accompanied by a proportional increase in its trade cost, which is unrealistic. Third, the stability analysis used to select spatial equilibria rests on myopic adjustment processes in which the location of mobile factors is driven by differences in current returns. Despite some analogy with evolutionary game theory (hence, “evolution”), this approach neglects the role of expectations (Krugman, 1991a; Matsuyama, 1991), which may be crucial for locational decisions since they are often made once and for all.² Last, notwithstanding their simplifying assumptions, the models of the new economic geography are often beyond the reach of analytical resolution, so that authors have to appeal to numerical investigations (hence, “the computer”).³

The purpose of this article is to propose a complementary modeling strategy that allows us to go beyond some of the current limits of the new economic geography. In particular, we do that by presenting a model of agglomeration and trade that, while displaying the main features of the core-periphery model by Krugman (1991b), differs under several major respects. First, preferences are not CES in that we adopt an alternative specification of the preference for variety, namely, the *quadratic utility model*, which is also popular in industrial organization (Dixit, 1979; Vives, 1990), in international trade (Krugman and Venables, 1990; Anderson et al., 1995), as well as in demand analysis (Phlips, 1983). Moreover, while firms are still considered as negligible actors, we adopt a broader concept of equilibrium than the one in Dixit

² See, however, Ottaviano (1999) for the analysis of a special case.

³ The most that can be obtained within this framework without resorting to numerical solutions has probably been achieved by Puga (1999). See also some chapters of Fujita et al. (1999).

and Stiglitz (1977). Second, trade costs are assumed to absorb resources that are different from the transported good itself. Taken together, our specifications allow us to disentangle the economic meanings of the various parameters, thus leading to clear-cut comparative static results that are likely to be easier to test than those based on Dixit and Stiglitz (1977).⁴ They also entail elasticities of demand and substitution that vary with prices, while equilibrium prices now depend on all the fundamentals of the market.

Using this framework, we are able to derive analytically the results obtained by Krugman (1991b). Going beyond them, our setting allows us to provide a neat welfare analysis of agglomeration. While natural due to the many market imperfections that are present in new economic geography models, such an analysis is seldom touched due to the limits of the standard approach.⁵ What we show is that the market yields agglomeration for values of the trade costs for which it is socially desirable to keep activities dispersed. Hence, while they coincide for high and low values of the trade costs, the equilibrium and the optimum differ for a domain of intermediate values. In this case, there is room for regional policy interventions grounded on both efficiency and equity considerations.

In addition, our framework allows us to study forward-looking location decisions and to determine the exact domain in which expectations matter for agglomeration to arise. Specifically, we show that expectations influence the agglomeration process in a totally unsuspected way in that they have an influence on the emergence of a particular agglomeration for intermediate values of the trade cost only. For such values, and only for them, *if (for whatever reason) workers expect the lagging region to become the leading one, their expectations will reverse the dynamics of the economy provided that the difference in initial endowments between the two regions is not too large.*

Finally, our model is sufficiently flexible to establish a bridge between the new economic geography and urban economics. We show that it can be easily extended to accommodate urban costs (Fujita, 1989). This trade-off leads to a set of results richer than the core-periphery model. When the manufactured goods' trade costs decrease, *the economy now displays a scheme given by dispersion, agglomeration, and redispersion* (Alonso, 1980).⁶ Such a result confirms and extends preliminary explorations undertaken by Helpman (1998), Tabuchi (1998), and Puga (1999). It also agrees with the observations according to which some developed economies

⁴ As an additional example, in Krugman (1991b) the same parameter turns out to measure not only the elasticities of demand and substitution but also (inversely) the returns to scale that remain unexploited in equilibrium.

⁵ The utilitarian approach is difficult to justify in the CES case because farmers and workers have different incomes, hence a different marginal utility for the numéraire. See, however, Krugman and Venables (1995) and Helpman (1998) for some numerical developments about the welfare implications of agglomeration in related models. See also Trionfetti (2001) for a particular example of a more general insight we will develop in the present article, namely, that in some cases market forces may lead to inefficient agglomeration.

⁶ Other reasons leading to a similar scheme are discussed in Ottaviano and Puga (1998) and Fujita et al. (1999).

(especially the United Kingdom) would experience redispersion (Geyer and Kontuly, 1996).

The organization of the article reflects what we have said earlier. The model is presented in the next section, while the equilibrium prices and wages are determined in Section 3 for any given distribution of firms and workers. The process of agglomeration is analyzed in Section 4 by using the standard myopic approach in selecting the stable equilibria. In Section 5, we compare the optimum and market outcomes. In Section 6, we introduce forward-looking behavior and show how our model can be used to compare history (in the sense of initial endowments) and expectations in the emergence of an agglomeration. The impact of urban costs associated with the formation of an agglomeration is investigated in Section 7. Section 8 concludes.

2. THE MODEL

The economic space is made of two regions, called H and F . There are two factors, called A and L . Factor A is evenly distributed across regions and is spatially immobile. Factor L is mobile between the two regions, and $\lambda \in [0, 1]$ denotes the share of this factor located in region H . For expositional purposes, we refer to sector A as “agriculture” and sector L as “manufacturing.” Accordingly, we call “farmers” the immobile factor A and “workers” the mobile factor L . We want to stress the fact, however, that *the role of factor A is to capture the idea that some inputs (such as land or some services) are nontradeable while some others have a very low spatial mobility (such as low-skilled workers).* Hence our model, as Krugman’s one, should not necessarily be interpreted as an agriculture-oriented model.

There are two goods in the economy. The first good is homogenous. Consumers have a positive initial endowment of this good that is also produced using factor A as the only input under constant returns to scale and perfect competition. This good can be traded freely between regions and is chosen as the numéraire. The other good is a horizontally differentiated product; it is supplied by using L as the only input under increasing returns to scale and imperfect competition.

Each firm in the manufacturing sector has a negligible impact on the market outcome in the sense that it can ignore its influence on, and hence reactions from, other firms. To this end, we assume that there is a continuum N of potential firms, so that all the unknowns are described by density functions. There are no scope economies so that, due to increasing returns to scale, there is a one-to-one relationship between firms and varieties. Since each firm sells a differentiated variety, it faces a downward-sloping demand.

Since there is a continuum of firms, each firm is negligible and the interaction between any two firms is zero. However, aggregate market conditions of some kind (here average price across firms) affect any single firm. This provides a setting in which individual 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.

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, where τ accounts for all the impediments to trade.

Preferences are identical across individuals and described by a *quasi-linear utility with a quadratic subutility* that is supposed to be symmetric in all varieties (see the Appendix for more details):

$$(1) \quad U(q_0; q(i), i \in [0, N]) = \alpha \int_0^N q(i) di - \frac{\beta - \gamma}{2} \int_0^N [q(i)]^2 di - \frac{\gamma}{2} \left[\int_0^N q(i) di \right]^2 + q_0$$

where $q(i)$ is the quantity of variety $i \in [0, N]$, and q_0 the quantity of the numéraire. The parameters in (1) are such that $\alpha > 0$ and $\beta > \gamma > 0$. In this expression, α expresses the intensity of preferences for the differentiated product, whereas $\beta > \gamma$ means that consumers are biased toward a dispersed consumption of varieties. Suppose, indeed, that an individual consumes a total mass of Nq of the differentiated product. If consumption is uniform on $[0, x]$ and zero on $(x, N]$, then the density on $[0, x]$ is Nq/x . Equation (1) evaluated for this consumption pattern is

$$(2) \quad \begin{aligned} U &= \alpha \int_0^x \frac{Nq}{x} di - \frac{\beta - \gamma}{2} \int_0^x \left(\frac{Nq}{x} \right)^2 di - \frac{\gamma}{2} \left[\int_0^x \left(\frac{Nq}{x} \right) di \right]^2 + q_0 \\ &= \alpha Nq - \frac{\beta - \gamma}{2x} N^2 q^2 - \frac{\gamma}{2} N^2 q^2 + q_0 \end{aligned}$$

which is strictly increasing in x since $\beta > \gamma$. Hence, regardless of the values of q and N , (2) is maximized at $x = N$ where variety consumption is maximal. We may then conclude that *the quadratic utility function exhibits love of variety as long as $\beta > \gamma$* . Finally, for a given value of β , the parameter γ expresses the substitutability between varieties: The higher γ , the closer substitutes the varieties.⁷

We use a quasi-linear utility that abstracts from general equilibrium income effects for analytical convenience. Although this modeling strategy gives our framework a fairly strong partial equilibrium flavor, 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 A or L) and $\bar{q}_0 > 0$ units of the numéraire. His budget constraint can then be written as follows:

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

⁷ When $\beta = \gamma$, substitutability is perfect. Indeed, (1) degenerates into a utility function that is quadratic in total consumption $\int_0^N q(i) di$, which is exactly what one would expect with a homogeneous product.

where y is the individual's labor income, $p(i)$ is the price of variety i , and the price of the agricultural good is normalized to one. The initial endowment \bar{q}_0 is supposed to be sufficiently large for the equilibrium consumption of the numéraire to be positive for each individual. By this assumption we want to focus on interior solutions only. This has some costs in terms of generality but, as we will see, larger benefits in terms of simpler analysis. It is also consistent with the idea that each individual is interested in consuming both types of goods. Note that this assumption does not imply that the share of the manufacturing sector must be small. It only requires that the equilibrium expenditure share on the differentiated product is smaller than one.

Solving the budget constraint for the numéraire consumption, plugging the corresponding expression into (1), and solving the first-order conditions with respect to $q(i)$ yields

$$\alpha - (\beta - \gamma)q(i) - \gamma \int_0^N q(j) dj = p(i) \quad i \in [0, N]$$

Therefore, the demand for variety $i \in [0, N]$ is

$$(3) \quad q(i) = a - bp(i) + c \int_0^N [p(j) - p(i)] dj$$

where $a \equiv \alpha/[(\beta + (N - 1)\gamma)]$, $b \equiv 1/[\beta + (N - 1)\gamma]$, and $c \equiv \gamma/(\beta - \gamma)[\beta + (N - 1)\gamma]$.⁸ Increasing the degree of product differentiation among a given set of varieties amounts to decreasing c . However, assuming that all prices are identical and equal to p , we see that the aggregate demand for the differentiated product equals $aN - bpN$, which is independent of c . Hence, (3) has the desirable property that the market size in the industry does not change when the substitutability parameter c varies. More generally, it is possible to decrease (increase) c through a decrease (increase) in the parameter γ in the utility U while keeping the other structural parameters a and b of the demand system unchanged. The own-price effect is stronger (as measured by $b + cN$) than each cross-price effect (as measured by c) as well as the sum of all cross-price effects (cN), thus allowing for different elasticities of substitution between pairs of varieties as well as for different own elasticities at different prices.

The indirect utility corresponding to the demand system (3) is as follows:

$$(4) \quad V(y; p(i), i \in [0, N]) = \frac{a^2 N}{2b} - a \int_0^N p(i) di + \frac{b + cN}{2} \int_0^N [p(i)]^2 di \\ - \frac{c}{2} \left[\int_0^N p(i) di \right]^2 + y + \bar{q}_0$$

⁸ Notice that when $\gamma = \beta$, the indirect demand cannot be inverted in terms of each variety's quantity $q(i)$ but only in terms of total quantity $\int_0^N q(i) di$. Once more, this is what one would expect with homogeneous products and explains why parameter c degenerates to infinity as γ tends to β .

Turning to the supply side, technology in agriculture requires one unit of A in order to produce one unit of output. With free trade in agriculture, the choice of this good as the numéraire implies that in equilibrium the wage of the farmers is equal to one in both regions, that is, $w_H^A = w_F^A = 1$. Technology in manufacturing requires ϕ units of L in order to produce any amount of a variety; that is, the marginal cost of production of a variety is set equal to zero. This simplifying assumption, which is standard in many models of industrial organization, entails no loss of generality when firms' marginal costs are incurred in the numéraire. Clearly, ϕ is a measure of the degree of increasing returns in the manufacturing sector.

Let n_H and n_F be the mass of firms in regions H and F , respectively. Labor market clearing implies that

$$(5) \quad n_H = \lambda L / \phi$$

and

$$(6) \quad n_F = (1 - \lambda) L / \phi$$

Consequently, the total mass of firms (varieties) in the economy is fixed and equal to $N = L / \phi$. This means that, in equilibrium, ϕ can also be interpreted as an inverse measure of the mass of firms. As $\phi \rightarrow 0$ (or $L \rightarrow \infty$), the mass of varieties becomes arbitrarily large. In addition, (5) and (6) show that the region with the larger labor market is also the region accommodating the larger proportion of firms. In addition, (5) and (6) imply that any change in the population of workers located in one region must be accompanied by a corresponding change in the mass of firms.

As to equilibrium wages, they are determined as follows: Due to free entry and exit, profits are zero in equilibrium. As in Krugman (1991b), the equilibrium wages corresponding to (5) and (6) are determined by a bidding process between firms for workers, which ends when no firm can earn a strictly positive profit at the equilibrium market prices. In other words, all operating profits are absorbed by the wage bills.

Since trade costs are positive, firms have the ability to segment markets; that is, each firm is able to set a price specific to the market in which its product is sold. Indeed, even for very low trade costs, empirical work shows that firms succeed in price discriminating among spatially separated markets (McCallum, 1995; Wei, 1996; Head and Mayer, 2000).

In the sequel, we focus on region H . Things pertaining to region F can be derived by symmetry. Using the assumption of symmetry between varieties and (3), demands faced by a representative firm located in H in region H (q_{HH}) and region F (q_{HF}) are given respectively by

$$(7) \quad q_{HH} = a - (b + cN)p_{HH} + cP_H$$

and

$$(8) \quad q_{HF} = a - (b + cN)p_{HF} + cP_F$$

where

$$P_H \equiv n_H p_{HH} + n_F p_{FH}$$

$$P_F \equiv n_H p_{HF} + n_F p_{FF}$$

Clearly, P_H/N and P_F/N are the average prices prevailing in regions H and F , so that P_H and P_F can be interpreted as the corresponding price indices since N is fixed. Finally, the profits made by a firm in H are defined as follows:

$$(9) \quad \Pi_H = p_{HH} q_{HH}(p_{HH})(A/2 + \lambda L) + (p_{HF} - \tau) q_{HF}(p_{HF})[A/2 + (1 - \lambda)L] - \phi w_H$$

where $A/2$ stands for the number of farmers in each region, and w_H for the wage prevailing in region H .

3. SHORT-RUN PRICE EQUILIBRIA

In this section, we study the process of competition between firms for a given spatial distribution of workers. Prices are obtained by maximizing profits, while wages are determined as described above by equating the resulting profits to zero. Since we have a continuum of firms, each one is negligible in the sense that its action has no impact on the market. Hence, when choosing its prices, a firm in H accurately neglects the impact of its decision over the two price indices P_H and P_F . In addition, because firms sell differentiated varieties, each one has some monopoly power in that it faces a demand function with finite elasticity.

When Dixit and Stiglitz use the CES, the same assumption implies that each firm is able to determine its price independently of the others because the price index enters the demand function as a multiplicative term. This no longer holds in our model because the price index now enters the demand function as an additive term (see (7) and (8)). Stated differently, a firm must account for the distribution of the firms' prices through some aggregate statistics, given here by the price index, in order to find its equilibrium price. As a consequence, our market solution is given by a Nash equilibrium with a continuum of players in which prices are interdependent: *Each firm neglects its impact on the market but is aware that the market as a whole has a nonnegligible impact on its behavior.* As a result, the equilibrium prices will depend on key aspects of the market instead of being given by a simple relative markup rule.

Since profit functions are concave in own price, solving the first-order conditions for profit maximization with respect to prices yields the equilibrium prices (denoted by *). In order to illustrate the type of interaction that characterizes our model of monopolistic competition, we describe how the equilibrium prices are determined. First, each firm i in region H maximizes its profit Π_H , assuming accurately that its price choice has no impact on the regional price indices P_H and P_F . By symmetry, the prices selected by the firms located within the same region are identical; hence, they are, respectively, given by two linear expressions $p_{HH}^*(P_H)$ and $p_{HF}^*(P_F)$. Second, these prices must be consistent; that is, they must satisfy the following relations:

$$\begin{aligned} n_H p_{HH}^*(P_H) + n_F p_{FH}^*(P_H) &= P_H \\ n_H p_{HF}^*(P_F) + n_F p_{FF}^*(P_F) &= P_F \end{aligned}$$

Given (5) and (6), it is then readily verified that

$$(10) \quad p_{HH}^* = \frac{1}{2} \frac{2a + \tau c(1 - \lambda)N}{2b + cN}$$

$$(11) \quad p_{FF}^* = \frac{1}{2} \frac{2a + \tau c \lambda N}{2b + cN}$$

$$(12) \quad p_{HF}^* = p_{FF}^* + \frac{\tau}{2}$$

$$(13) \quad p_{FH}^* = p_{HH}^* + \frac{\tau}{2}$$

Consequently, the equilibrium prices under monopolistic competition depend on the demand and firm distributions between regions. In particular, the prices charged by both local and foreign firms fall when the mass of local firms increases (because price competition is fiercer), but the impact is weaker when τ is smaller. In the limit, when τ is negligible, the relocation of firms in say, H , has almost no impact on market prices. In this case, prices are “independent” of the way firms are distributed between the two regions.

Equilibrium prices also rise when the relative desirability of the differentiated product with respect to the numéraire, evaluated by a , gets larger or when the degree of product differentiation, inversely measured by c , increases provided that trade occurs (see (14)). All these results are in accordance with what is known in industrial organization and spatial pricing theory.

Furthermore, there is freight absorption since only a fraction of the trade cost is passed on to the consumers. Indeed, we have

$$p_{HF}^* - p_{HH}^* = \tau \frac{b + c\lambda N}{2b + cN} < \tau \quad \text{which is equal to } \frac{\tau}{2} \text{ when } \lambda = 1/2$$

$$p_{FH}^* - p_{FF}^* = \tau \frac{b + c(1 - \lambda)N}{2b + cN} < \tau \quad \text{which is equal to } \frac{\tau}{2} \text{ when } \lambda = 1/2$$

It is well known that a monopolist facing a linear demand absorbs exactly one-half of the trade cost. By contrast, we see that monopolistic competition leads to less (more) freight absorption than monopoly when the foreign market is the small (large) one: In an attempt to penetrate the distant market, competition leads firms to a price gap that varies with the relative size of the home and foreign markets.

By inspection, it is readily verified that p_{HH}^* (p_{FF}^*) is increasing in τ because the local firms in H (F) are more protected against foreign competition, while $p_{HF}^* - \tau$ ($p_{FH}^* - \tau$) is decreasing because it is now more difficult for these firms to sell on the foreign market. Observe also that arbitrage is never profitable since price differentials are always lower than trade costs. Finally, our demand side happens to be consistent with identical demand functions at different locations but different price levels, as in standard spatial pricing theory.

Deducting the unit trade cost τ from the prices set on the distant markets, that is, (12) and (13), we see that firms’ prices net of trade costs are positive regardless of the workers’ distribution if and only if

$$(14) \quad \tau < \tau_{\text{trade}} \equiv \frac{2a\phi}{2b\phi + cL}$$

which depends only upon the primitives $(A, L, \alpha, \beta, \gamma, \phi)$ once a , b , c , and N are replaced by their values. The same condition must hold for consumers in F (H) to

buy from firms in H (F), that is, for the demand (8) evaluated at the prices (10) and (11) to be positive for all λ . From now on, condition (14) is assumed to hold. Consequently, there is intra-industry trade and reciprocal dumping, as in Anderson et al. (1995). However, *there must be increasing returns for trade to occur*. Indeed, when $\phi = 0$, all potential varieties are produced in each region that becomes autarkic. More generally, it is readily verified that

$$\frac{d\tau_{\text{trade}}}{d\phi} > 0 \quad \frac{d\tau_{\text{trade}}}{d\gamma} < 0$$

so that trade is more likely the higher are the intensity of increasing returns and the degree of product differentiation.

It is easy to check that the equilibrium gross profits earned by a firm established in H on each separated market are as follows:

$$(15) \quad \Pi_{HH}^* = (b + cN)(p_{HH}^*)^2(A/2 + \lambda L)$$

where Π_{HH}^* denotes the profits earned in H , while the profits made from selling in F are

$$(16) \quad \Pi_{HF}^* = (b + cN)(p_{HF}^* - \tau)^2[A/2 + (1 - \lambda)L]$$

Increasing λ has two opposite effects on Π_{HH}^* . First, due to tougher competition, the equilibrium price (10) falls as well as the quantity of each variety bought by each consumer living in region H . At the same time, the total population of consumers residing in this region increases, so that the profits made by a firm located in H on local sales might rise. What is at work here is *an aggregate local demand effect due to the increase in the local population that may compensate firms for the adverse price effect as well as for the individual demand effect generated by a wider array of local varieties*.

The individual consumer surplus S_H in region H associated with the equilibrium prices (10) and (13) is then as follows (a symmetric expression holds in region F):

$$\begin{aligned} S_H(\lambda) = & \frac{a^2 L}{2b\phi} - \frac{aL}{\phi} [\lambda p_{HH}^* + (1 - \lambda)p_{FH}^*] \\ & + \frac{(b\phi + cL)L}{2\phi^2} [\lambda (p_{HH}^*)^2 + (1 - \lambda)(p_{FH}^*)^2] \\ & - \frac{cL^2}{2\phi^2} [\lambda p_{HH}^* + (1 - \lambda)p_{FH}^*]^2 \end{aligned}$$

Differentiating twice this expression with respect to λ shows that $S_H(\lambda)$ is concave. Furthermore, (14) implies that $S_H(\lambda)$ is always increasing in λ over the interval $[0, 1]$.

The equilibrium wage prevailing in region H may be obtained by evaluating $w_H^*(\lambda)(\Pi_{HH}^* + \Pi_{HF}^*)/\phi$, thus yielding the following expression:

$$w_H^*(\lambda) = \frac{b\phi + cL}{4(2b\phi + cL)^2\phi^2} \left\{ [2a\phi + \tau cL(1 - \lambda)]^2 \left(\frac{A}{2} + \lambda L \right) + [2a\phi - 2\tau b\phi - \tau cL(1 - \lambda)]^2 \left[\frac{A}{2} + (1 - \lambda)L \right] \right\}$$

which, after simplifying, turns out to be quadratic in λ . Standard, but cumbersome, investigations reveal that $w_H^*(\lambda)$ is concave and increasing (convex and decreasing) in λ when ϕ is large (small) as well as when τ , c , A , and L are small (large). This implies that both $S_H(\lambda)$ and $w_H^*(\lambda)$ increase with λ when τ is small, while they go in opposite directions when τ is large.

4. WHEN DO WE OBSERVE AGGLOMERATION?

The distribution $\lambda \in [0, 1]$ is a *spatial equilibrium* when no worker may get a higher utility level by changing location. Given that the indirect utility in region H is as follows:

$$V_H(\lambda) = S_H(\lambda) + w_H^*(\lambda) + \bar{q}_0$$

a spatial equilibrium arises at $\lambda \in (0, 1)$ when

$$\Delta V(\lambda) \equiv V_H(\lambda) - V_F(\lambda) = 0$$

or at $\lambda = 0$ when $\Delta V(0) \leq 0$, or at $\lambda = 1$ when $\Delta V(1) \geq 0$.

In order to study the stability of a spatial equilibrium, we assume that local labor markets adjust instantaneously when some 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 (5) and (6) remain valid for the new distribution of workers. Wages are then adjusted in each region for each firm to earn zero profits everywhere. For now, we assume a myopic adjustment process; that is, the driving force in the migration process is workers' current utility differential between H and F :

$$(17) \quad \dot{\lambda} \equiv d\lambda/dt = \begin{cases} \Delta V(\lambda) & \text{if } 0 < \lambda < 1 \\ \min\{0, \Delta V(\lambda)\} & \text{if } \lambda = 1 \\ \max\{0, \Delta V(\lambda)\} & \text{if } \lambda = 0 \end{cases}$$

when t is time. Clearly, a spatial equilibrium implies $\dot{\lambda} = 0$. If $\Delta V(\lambda)$ is positive, some workers will move from F to H ; if it is negative, some will go in the opposite direction. In the sequel, we assume that individual consumption of the numéraire is positive during the adjustment process. This assumption is made to capture the idea that individuals need both types of goods.⁹

A spatial equilibrium is *stable* for (17) if, for any marginal deviation from the equilibrium, this equation of motion brings the distribution of workers back to the original one. Therefore, the agglomerated configuration is always stable when it is an equilibrium, while the dispersed configuration is stable if and only if the slope of $\Delta V(\lambda)$ is nonpositive in a neighborhood of this point.

⁹ An analytically convenient way to achieve that is to assume that each consumer receives the same endowment \bar{q}_0 at each point in time.

The forces at work are similar to those found in the core-periphery model. First, the immobility of the farmers is a centrifugal force, at least as long as there is trade between the two regions. The centripetal force finds its origin in a demand effect generated by the preference for variety. If a larger number of firms are located in region H , there are two effects at work. First, less varieties are imported. Second, (10) and (13) imply that the equilibrium prices of all varieties sold in H are lower. (Observe that the latter effect does not appear in Krugman's model.) This, in turn, induces some consumers to migrate toward this region. The resulting increase in the number of consumers creates a larger demand for the industrial good in the corresponding region, which therefore increases operating profits (hence, wages) and leads to more firms to move there. In other words, both backward and forward linkages are present in our model.

It is readily verified that the indirect utility differential can be written as follows:

$$(18) \quad \Delta V(\lambda) \equiv V_H(\lambda) - V_F(\lambda) \equiv S_H(\lambda) - S_F(\lambda) + w_H^*(\lambda) - w_F^*(\lambda) \\ = C\tau(\tau^* - \tau)(\lambda - 1/2)$$

where

$$C \equiv [2b\phi(3b\phi + 3cL + cA) + c^2L(A + L)] \frac{L(b\phi + cL)}{2\phi^2(2b\phi + cL)^2} > 0$$

and

$$\tau^* \equiv \frac{4a\phi(3b\phi + 2cL)}{2b\phi(3b\phi + 3cL + cA) + c^2L(A + L)} > 0$$

which can also be restated in terms of the primitives of the economy.

It follows immediately from (18) that $\lambda = 1/2$ is always an equilibrium. Since $C > 0$, for $\lambda \neq 1/2$ the indirect utility differential has always the same sign as $\lambda - 1/2$ if and only if $\tau < \tau^*$; otherwise it has the opposite sign. In particular, *when there are no increasing returns in the manufacturing sector* ($\phi = 0$), the coefficient of $(\lambda - 1/2)$ is always negative since $\tau^* = 0$, so that *dispersion is the only (stable) equilibrium*. This shows once more the importance of increasing returns for the possible emergence of an agglomeration.

It remains to determine when τ^* is lower than τ_{trade} . This is so if and only if

$$(19) \quad A/L > \frac{6b^2\phi^2 + 8bc\phi L + 3c^2L^2}{cL(2b\phi + cL)} > 3$$

where the second inequality holds because $b/c = \beta/\gamma - 1 \in (0, +\infty)$. This inequality means that the population of farmers is large relative to the population of workers. When (19) does not hold, the coefficient of $(\lambda - 1/2)$ in (18) is always positive for all $\tau < \tau_{\text{trade}}$.

When $\tau < \tau^*$, the symmetric equilibrium is unstable and workers agglomerate in region $H(F)$ provided that the initial fraction of workers residing in this region exceeds $1/2$. In other words, *agglomeration arises when the trade cost is low enough*, as in Krugman (1991b) and for similar reasons. In contrast, for large trade costs, that

is, when $\tau > \tau^*$, it is straightforward to see that the symmetric configuration is the only stable equilibrium. Hence, the threshold τ^* corresponds to both the critical value of τ at which symmetry ceases to be stable (the “break point”) and the value below which agglomeration is stable (the “sustain point”); this follows from the fact that (18) is linear in λ .

PROPOSITION 1. *Assume that $\tau < \tau_{\text{trade}}$. Two cases may arise:*

- (i) *When (19) holds, we have the following: If $\tau > \tau^*$, then the symmetric configuration is the only stable spatial equilibrium with trade; if $\tau < \tau^*$, there are two stable spatial equilibria corresponding to the agglomerated configurations with trade; if $\tau = \tau^*$, then any configuration is a spatial equilibrium.*
- (ii) *When (19) does not hold, any stable spatial equilibrium involves agglomeration.*

The reverse of (19) plays a role similar to the “black hole” condition in Krugman and Venables (1995) and Fujita et al. (1999): Regardless of the value of the trade costs, the region with the larger initial share of the manufacturing sector always attracts the whole sector. As in their case, more product differentiation (lower c) and stronger increasing returns (higher ϕ) make the black hole condition more likely. Although the size of the industrial sector is captured here through the relative population size A/L and not through its share in consumption, the intuition is similar: The ratio A/L must be sufficiently large for the economy to display different types of equilibria according to the value of τ . Our result does not depend on the expenditure share on the manufacturing sector because of the absence of general equilibrium income effects: Either small or large sectors in terms of expenditure share may be agglomerated when τ is small enough. This does not strike us as being implausible.

Furthermore, when $\tau_{\text{trade}} > \tau^*$, trade occurs regardless of the type of equilibrium that is stable. However, the nature of trade varies with the type of configuration emerging in equilibrium. In the dispersed configuration, there is only intra-industry trade in the differentiated product; in the agglomerated equilibrium, the region accommodating the manufacturing sector only imports the homogenous good from the other region.

When increasing returns are stronger, as expressed by higher values of ϕ , τ^* rises since $d\tau^*/d\phi > 0$. This means that *the agglomeration of the manufacturing sector is more likely, the stronger are the increasing returns at the firm’s level*. In addition, τ^* increases with product differentiation since $d\tau^*/d\gamma < 0$. In other words, *more product differentiation fosters agglomeration*. In particular, γ very small implies that $\tau_{\text{trade}} < \tau^*$, so that agglomeration always arises under trade.

The best way to convey the economic intuition behind Proposition 1 is probably to make use of a graphical analysis.¹⁰ Figure 1 depicts the aggregate inverse demand in region H for a typical local firm after choosing, for simplicity, the units of L so that $b + cN = 1$:

¹⁰ For illustrative purposes, we neglect the impact of relocation on the firm’s profit in F since this one is typically smaller than the impact on its profit in H .

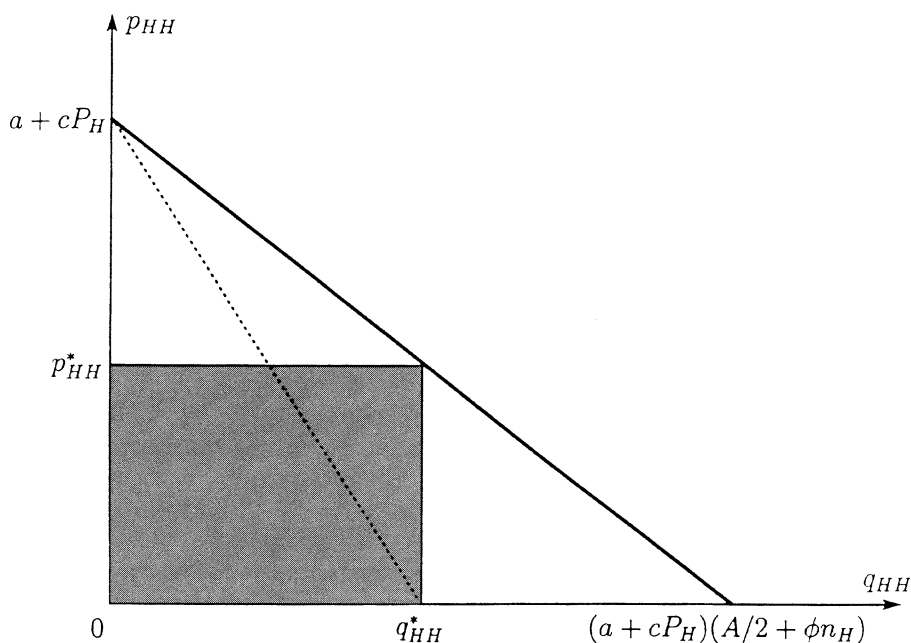


FIGURE 1

DETERMINATION OF THE EQUILIBRIUM OUTCOME

$$(20) \quad p_{HH} = a + cP_H(n_H, \tau) - \frac{q_{HH}}{A/2 + \phi n_H}$$

Since $p_{FH} > p_{HH}$ and the total number of firms is fixed by labor market clearing, the price index P_H is a decreasing function of n_H at a rate that increases with τ :

$$(21) \quad \frac{\partial P_H(n_H, \tau)}{\partial n_H} < 0 \quad \text{and} \quad \left| \frac{\partial^2 P_H(n_H, \tau)}{\partial n_H \partial \tau} \right| > 0$$

The horizontal and vertical intercepts of (20) are, respectively, $[a + cP_H(n_H, \tau)](A/2 + \phi n_H)$ and $[a + cP_H(n_H, \tau)]$. The equilibrium values of q_{HH} and p_{HH} are shown as q_{HH}^* and p_{HH}^* . They are found by setting marginal revenue equal to marginal cost. The operating profits are shown by the shaded rectangle and accrue to the workers while, as usual, the triangle in Figure 1 represents the consumer surplus enjoyed by both types of workers.

Although it depicts a partial equilibrium argument, Figure 1 is a powerful learning device to understand the forces at work in our model. To see why, start from an initial situation where regions are identical ($n_H = n_F$). Suppose that some firms move from the foreign to the home region, so that n_H rises and n_F falls. For these firms to want to stay in the home region, operating profits, thus wages, have to increase. Indeed, were this not the case, the firms would rather go back to the foreign region.

As revealed by Figure 1, an increase in n_H has two opposite effects on operating profits, hence on wages. First, as new firms enter the home region, the price index

$P_H(n_H, \tau)$ decreases. Ceteris paribus, this would shift the inverse demand (20) toward the origin of the axes, and operating profits, hence wages, would shrink. This effect is due to increased competition in the home market and stems from the fact that fewer firms now face trade costs when supplying the home market. But this negative competition effect is not the only effect. For some firms to move to the home region, some workers have to follow, since $n_H = \lambda L / \phi$. This means that, as n_H increases, λ also goes up, so that the market of the home region expands. Ceteris paribus, the horizontal intercept of the inverse demand would move away from the origin, and operating profits, thus wages, would expand. This is a positive demand effect that is induced by the linkage between the locations of firms and workers' expenditures.

Since the two effects oppose each other, the net result is a priori ambiguous. But we can say more than that. In particular, we can assess which effect prevails depending on parameter values. Start with the competition effect that goes through $[a + cP_H(n_H, \tau)]$. This effect is strong if c is large, that is, if varieties are good substitutes. It is also strong if $|\partial P_H(n_H, \tau) / \partial n_H|$ is large. As shown in (21), this happens if τ is large, because, when obstacles to trade are high, competition from the other region is weak and home firms care a lot about their competitors' being close rather than distant. As to the demand effect, it will be strong if ϕ is large because each new firm brings along many workers, and if A is small because immigrants have a large impact on the local market size.

We can therefore conclude that the demand effect dominates the competition effect when goods are bad substitutes (c small), increasing returns are intense (ϕ large), the farmers are unimportant (A small), and trade costs are low (τ small). Under such circumstances, the entry of new firms in one region would raise the operating profits of all firms, hence wages. *Higher operating profits and wages would attract more firms and workers, thus generating circular causation among locational decisions.* Agglomeration would then be sustainable as a spatial equilibrium.

This argument establishes a sufficient condition for agglomeration. Since the impact of firms' relocation on consumer surplus is always positive, agglomeration could still arise even when operating profits, hence wages, decrease with the size of the local market, because the demand effect is dominated by the competition effect. Furthermore, the same argument is likely to hold for most downward-sloping demand functions.

5. OPTIMALITY VERSUS EQUILIBRIUM

We now wish to determine whether or not such an agglomeration is socially optimal. 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. Observe that no distortion arises in the total number of varieties since N is determined by the factor endowment (L) and technology (ϕ) in the manufacturing sector and is, therefore, the same at both the equilibrium and optimum outcomes. Because our setting assumes transferable utility, the planner chooses λ in order to maximize the sum of individual indirect utilities:

$$(22) \quad W(\lambda) \equiv \frac{A}{2}[S_H(\lambda) + 1] + \lambda L[S_H(\lambda) + w_H(\lambda)] + \frac{A}{2}[S_F(\lambda) + 1] \\ + (1 - \lambda)L[S_F(\lambda) + w_F(\lambda)]$$

in which all prices have been set equal to marginal cost:

$$p_{HH}^0 = p_{FF}^0 = 0 \quad \text{and} \quad p_{HF}^0 = p_{FH}^0 = \tau$$

thus implying by (15) and (16) that operating profits are zero, and hence $w_H^0(\lambda) = w_F^0(\lambda) = 0$ for every λ , so that firms do not incur any loss. Hence, (22) becomes

$$(23) \quad W(\lambda) = C^0 \tau (\tau^0 - \tau) \lambda (\lambda - 1) L + \text{constant}$$

where

$$C^0 \equiv \frac{L}{2\phi^2} [2b\phi + c(A + L)]$$

and

$$\tau^0 \equiv \frac{4a\phi}{2b\phi + c(A + L)}$$

The function (23) is strictly concave in λ if $\tau > \tau^0$, and strictly convex if $\tau < \tau^0$. Furthermore, since 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 τ^0 .

Hence, we have the following proposition:

PROPOSITION 2. *If $\tau > \tau^0$, then the symmetric configuration is the optimum; if $\tau < \tau^0$, any agglomerated configuration is the optimum; if $\tau = \tau^0$, any configuration is an optimum.*

In accordance with intuition, it is socially desirable to agglomerate the manufacturing sector into a single region once trade costs are low, increasing returns in the manufacturing sector are strong enough ($d\tau^0/d\phi > 0$), and/or the output of this sector is sufficiently differentiated ($d\tau^0/d\gamma < 0$). In particular, the optimum is always dispersed when increasing returns vanish ($\phi = 0$).

A simple calculation shows that $\tau^0 < \tau^*$. This means that *the market yields an agglomerated configuration for a whole range ($\tau^0 < \tau < \tau^*$) of trade cost values for which it is socially desirable to have a dispersed pattern of activities*. Accordingly, when trade costs are low ($\tau < \tau^0$) 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 trade costs ($\tau^0 < \tau < \tau^*$), the market provides excessive agglomeration, thus justifying the need for an active regional policy in order to foster the dispersion of the modern sector on both the efficiency and equity grounds.

This discrepancy may be explained as follows: First, workers do not internalize the negative external effects they impose on the farmers who stay put in the workers'

region of origin, nor do they account for the impact of their migration decisions on the residents in their region of destination. Hence, even though the workers 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), so that 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, which tells us why $\tau^0 < \tau^*$.

We will return to the debate of spatial equilibrium versus optimum in Section 7.

6. THE IMPACT OF WORKERS' EXPECTATIONS ON THE AGGLOMERATION PROCESS

The adjustment process (17) is often used in new economic geography. Yet, the underlying dynamics are myopic because workers care only about their current utility level, thus implying that only history matters. This is a fairly restrictive assumption to the extent that *migration decisions are typically made on the grounds of current and future utility flows*. In addition, this approach has been criticized because it is not consistent with fully rational forward-looking behavior (Matsuyama, 1991). In this section, we want to see how the model presented above can be used to shed more light on the interplay between history and expectations in the formation of the economic space when migrants maximize the intertemporal value of their utility flows. In particular, we are interested in identifying the conditions under which, when initially regions host different numbers of workers, the common belief that workers will eventually agglomerate in the smaller region can reverse the historically inherited advantage of the larger region. Formally, we are interested in determining *the parameter domains for which there exists an equilibrium path consistent with this belief*, assuming that workers have perfect foresight (*self-fulfilling prophecy*).

For concreteness, let us consider the case in which initially region F is larger than H . The opposite case can be studied in a symmetric way. Therefore, we want to test the consistency of the belief that, starting from $t=0$, all workers will end up being concentrated in H at some future date $t=T$; that is, there exists $T \geq 0$ such that, given $\lambda_0 < 1/2$,

$$(24) \quad \begin{aligned} \dot{\lambda}(t) &\neq 0 \quad \text{and} \quad \lambda(t) < 1 && \text{for } t \in [0, T) \\ \dot{\lambda}(t) &= 0 \quad \text{and} \quad \lambda(t) = 1 && \text{for } t \geq T \end{aligned}$$

Let $V_H(t)$ and $V_F(t)$ be the instantaneous utility levels of a worker currently in regions H and F , respectively, at time $t \geq 0$.¹¹ Furthermore, let V^C be the instantaneous utility level in region H at $\lambda=1$. Then, under (24),

¹¹ To ease notation, when variables depend on time t through $\lambda(t)$, we only report their dependence on time as long as this does not lead to any ambiguity. Moreover, throughout this section, we keep on assuming that the consumption of the numéraire is always positive in both regions at any time. Formally, this implies that we assume away any form of intertemporal trade.

$$V_H(t) = V^C \quad \text{for all } t \geq T$$

Since workers have perfect foresight, 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 (Mussa, 1978). In other words, a migrant imposes a negative externality on the others by congesting the migration process. Specifically, we follow Krugman (1991a) and assume that the utility loss for a migrant at time t is equal to $|\dot{\lambda}(t)|/\delta$, where $\delta > 0$ is the speed of adjustment. Thus, under (24), the intertemporal utility of a worker who moves from F to H at time $t \in [0, T)$ is given by

$$(25) \quad u(t) = \int_0^t e^{-\rho s} V_F(s) ds + \int_t^T e^{-\rho s} V_H(s) ds + e^{-\rho T} V^C / \rho - e^{-\rho t} \dot{\lambda}(t) / \delta$$

where $\rho > 0$ is the rate of time preference.

We are now ready to characterize the equilibrium migration process. At the initial time 0, each worker residing in F decides at which date $t \geq 0$ to migrate from F to H . In so doing, she believes that at date T all workers will end up being in H . As argued by Fukao and Bénabou (1993) and Ottaviano (1999), for this belief to be consistent with the equilibrium outcome, a worker must be indifferent between moving at any date $t \geq 0$ or at the final expected date T . In the former case, she expects to receive (25). In the latter, she forecasts

$$(26) \quad u(T) = \int_0^T e^{-\rho s} V_F(s) ds + e^{-\rho T} V^C / \rho$$

that is, the limit of (25) as t approaches T where the moving cost disappears since migration stops. Notice that, since the assumed belief has all workers in H from T onwards, in both cases the worker expects a utility flow $V_H(t)$ for all $t \geq T$.

Subtracting (26) from (25), for each $t \geq 0$, we get

$$(27) \quad \begin{aligned} u(t) - u(T) &= \int_t^T e^{-\rho s} [V_H(s) - V_F(s)] ds - e^{-\rho t} \dot{\lambda}(t) / \delta \\ &= e^{-\rho t} [v_H(t) - v_F(t)] - e^{-\rho t} \dot{\lambda}(t) / \delta \\ &= e^{-\rho t} \Delta v(t) - e^{-\rho t} \dot{\lambda}(t) / \delta \end{aligned}$$

where

$$v_r(t) \equiv \int_t^T e^{-\rho(s-t)} V_r(s) ds + e^{-\rho(T-t)} V^C / \rho \quad r = H, F$$

is what residence in r buys to a worker who stays in H ($r = H$) or to a worker in F who plans to move to H at T ($r = F$) from t onwards, while $\Delta v(t) \equiv v_H(t) - v_F(t)$ and $\Delta v(T) = 0$.

Since in equilibrium a worker moving at t must be indifferent between migrating at that date or at any other date, until the final expected date T , along an equilibrium path it must be that $u(t) = u(T)$ for all $t \in [0, T)$, which, using (27), implies

$$(28) \quad \dot{\lambda}(t) = \delta \Delta v(t) \quad \text{for all } t \in [0, T)$$

with $\lambda(T) = 1$.

Furthermore, differentiating $\Delta v(t)$ with respect to time yields

$$(29) \quad \dot{\Delta v}(t) = \rho \Delta v(t) - \Delta V(t)$$

where $\Delta V(t) \equiv V_H(t) - V_F(t)$ stands for the expected instantaneous indirect utility differential flow given by (18). Hence, we obtain a system of two linear differential equations, instead of the first-order differential equation (17), with the terminal conditions $\lambda(T) = 1$, $\Delta v(T) = 0$.

Since $\lambda = 1/2$ implies $\Delta V = 0$, the systems (28) and (29) have always an interior steady state at $(\lambda, \Delta v) = (1/2, 0)$ that corresponds to the dispersed configuration. While for $\tau > \tau^*$ it is the only steady state; for $\tau < \tau^*$ two other steady states exist at $(\lambda, \Delta v) = (0, 0)$ and $(\lambda, \Delta v) = (1, 0)$. In the latter case, as in Fukao and Bénabou (1993), since all workers are concentrated within the same region, there is no reason to hold that

$$\dot{\Delta v}(T) = 0$$

in (29). In fact, by the terminal conditions, we have

$$\dot{\Delta v}(T) = -\Delta V(T) = -C\tau(\tau^* - \tau)/2 < 0$$

which ensures that the spatial equilibrium involves no worker in F . Therefore, the assumed belief (24) can be consistent only in the latter case, on which we concentrate from now on.

In order to identify the conditions under which the belief that workers will eventually agglomerate in the initially smaller region H can reverse the historically inherited advantage of the larger region F , we have to study the global stability of systems (28) and (29). In so doing, we exploit the fact that, since the system is linear, local and global stability properties coincide, and we focus on the former. The eigenvalues of the Jacobian matrix of the systems (28) and (29) are given by

$$(30) \quad \frac{\rho \pm \sqrt{\rho^2 - 4\delta C\tau(\tau^* - \tau)}}{2}$$

When $\tau < \tau^*$, two scenarios may arise. In the first one, $\rho > 2\sqrt{C\delta\tau(\tau^* - \tau)}$ so that the two eigenvalues are real and both positive. The steady state $(1/2, 0)$ is an unstable node, and there are two trajectories that steadily go to the endpoints, $(0, 0)$ or $(1, 0)$, depending on the initial spatial distribution of workers, say, λ_0 . In this case, *only history matters*: From any initial $\lambda_0 \neq 1/2$, there is a single trajectory that goes toward the closer endpoint as in the case where the dynamics are given by (17). This means that belief (24) is inconsistent with the equilibrium path, whence the myopic adjustment process studied in the previous section provides a

good approximation of the qualitative evolution of the economy under forward-looking behavior.

Things turn out to be quite different in the second scenario in which $\rho < 2\sqrt{C\delta\tau(\tau^* - \tau)}$. Since $C\tau(\tau^* - \tau) = 0$ at both $\tau = 0$ and $\tau = \tau^*$, the equation $C\tau(\tau^* - \tau) - \rho^2/4\delta = 0$ has two positive real roots in τ , denoted τ_1^e and τ_2^e , that are smaller than τ^* :

$$\tau_1^e \equiv \frac{\tau^* - D}{2} \quad \text{and} \quad \tau_2^e \equiv \frac{\tau^* + D}{2}$$

where

$$D \equiv \sqrt{(\tau^*)^2 - \rho^2/C\delta}$$

stands for the size of the domain of values of τ for which expectations matter; it shrinks as the discount rate ρ increases or as the speed of adjustment δ decreases. Indeed, for $\tau \in (0, \tau_1^e)$ as well as for $\tau \in (\tau_2^e, \tau^*)$, both eigenvalues are real positive numbers and the steady state $(1/2, 0)$ is an unstable node as before. However, for $\tau \in (\tau_1^e, \tau_2^e)$, they become complex numbers with a positive real part, so that the steady state is an unstable focus. The two trajectories spiral out from $(1/2, 0)$. Therefore, for any λ_0 close enough to, but different from, $1/2$, there are two alternative trajectories going in opposite directions. It is in such a case that expectations decide along which trajectory the system moves, so that belief (24) is self-fulfilling. In other words, *expectations matter for λ close enough to $1/2$, while history matters otherwise*. The corresponding domains are now described.

The range of values for which expectations matter, called the *overlap* by Krugman (1991a), can be obtained as follows: As observed by Fukao and Bénabou (1993), the system must be solved backwards in time starting from the terminal points $(0, 0)$ and $(1, 0)$. The first time the backward trajectories intersect the locus $\Delta v = 0$ allows for the identifications of the endpoints of the overlap:

$$\lambda^L \equiv \frac{1}{2}(1 - \Lambda) \quad \lambda^H \equiv \frac{1}{2}(1 + \Lambda)$$

where

$$\Lambda \equiv \exp\left(-\frac{\rho\pi}{\sqrt{4\delta C\tau(\tau^* - \tau) - \rho^2}}\right)$$

is the width of the overlap, which is an interval centered around $\lambda = 1/2$.

The overlap is nonempty as long as $\tau \in (\tau_1^e, \tau_2^e)$. Thus, the width of the overlap is increasing in δ , C , and τ^* , while it decreases with ρ . Moreover, it is \cap -shaped with respect to τ , reaching a maximum at $\tau = \tau^*/2$. Since $C\tau(\tau^* - \tau) > 0$ is the slope of ΔV and since this one measures the strength of the forward and backward linkages pushing towards agglomeration, we see that expectations matter more when such linkages are stronger. Consequently, we have shown the following result:

PROPOSITION 3. *Let λ_0 be the initial spatial distribution of workers. If $\tau < \tau^*$ and $\rho < 2\sqrt{C\delta\tau(\tau^* - \tau)}$, there exist $\tau_1^e \in (0, \tau^*/2)$, $\tau_2^e \in (\tau^*/2, \tau^*)$, $\lambda^L \in (0, 1/2)$, and*

$\lambda^H \in (1/2, 1)$ such that workers' beliefs about their future earnings influence the process of agglomeration if and only if $\tau \in (\tau_1^e, \tau_2^e)$ and $\lambda_0 \in [\lambda^L, \lambda^H]$.

Hence, history alone matters when τ and λ are large enough or small enough. In other words, the agglomeration process evolves as if workers were shortsighted when obstacles to trade are high or low and when regions are initially quite different. Instead, as long as obstacles to trade take intermediate values and regions are not initially too different, *the equilibrium is determined by workers' expectations and not by history.*¹²

The existence of the range (τ_1^e, τ_2^e) may be explained as follows: Suppose, indeed, that the economy is such that $\lambda_0 < 1/2$, and ask what is needed to reverse an ongoing agglomeration process currently leading towards $\lambda = 0$. If the evolution of the economy were to change direction, workers would experience falling instantaneous indirect utility flows for some time period as long as $\lambda < 1/2$. The instantaneous indirect utility flows would start growing only after λ becomes larger than $1/2$. Accordingly, workers would first experience utility losses followed by utility gains. Since the losses would come before the gains, they would be less discounted. This provides the root for the intuition behind Proposition 3. When the forward and backward linkages lead to substantial wage rises (that is, for intermediate values of τ), the benefits of agglomerating at $\lambda = 1$ can compensate workers for the losses they incur during the transition phase, thus making the reversal of migration possible. On the contrary, when these linkages get weaker (that is, for low or high values of τ), the benefits of agglomerating at $\lambda = 1$ do not compensate workers for the losses. As a consequence, *the reversal in the migration process may occur only for intermediate values of τ .*

As to the remaining comparative static properties of the overlap, they are explained by the fact that proximity to $\lambda = 0$ increases the time period over which workers bear losses, a large rate of time preference gives more weight to them, and a slow speed of adjustment extends the time period over which workers' well-being is reduced.

7. THE IMPACT OF URBAN COSTS

So far, we have assumed that the agglomeration of workers into a single region does not involve any agglomeration costs. Yet, it is reasonable to believe that a growing settlement in a given region will often take the form of an urban area, typically a city. In order to deal with such an aspect of the process of agglomeration, we extend the core-periphery model by adding the central variables suggested by urban economics (Fujita, 1989). In order to keep the analysis short, we go back to the myopic adjustment process of Section 4.

¹² When $\tau > \tau^*$, we know that the equilibrium path is inconsistent with workers' expectations under (24). And, indeed, the two eigenvalues are real and have opposite signs, so that the steady state is a saddle point. Under the assumption of perfect foresights, regardless of the initial distribution λ_0 , the system moves along a single stable trajectory toward the symmetric steady state. In other words, the system converges toward the dispersed configuration, thus implying that neither expectations nor history matter for the final outcome.

Space is now continuous and one-dimensional. Each region has a spatial extension and involves a linear city whose center is given but with a variable size. The city center stands for a central business district (CBD) in which all firms locate once they have chosen to set up in the corresponding region (see Fujita and Thisse, 1996, for various arguments explaining why firms want to be agglomerated in a CBD). The two CBDs are two remote points of the location space. Interregional trade flows go from one CBD to the other.

Housing is a new good in our economy and is described by the amount of land used by workers. While firms are assumed not to consume land, workers, when they live in a certain region, are urban residents who consume land and commute to the regional CBD in which manufacturing firms are located. Hence, unlike Krugman (1991b) but like Alonso (1964), Helpman (1998), and Tabuchi (1998), *each agglomeration has a spatial extension that imposes commuting and land costs on the corresponding workers*. For simplicity, workers consume a fixed lot size normalized to unity, while commuting costs are linear in distance, the commuting cost per unit of distance being given by $\theta > 0$ units of the numéraire. Without loss of generality, the opportunity cost of land is normalized to zero.

When λL workers live in H , they are equally distributed around the H -CBD. In equilibrium, since all workers residing in region H earn the same wage, they reach the same utility level. Furthermore, since they all consume one unit of land, the equilibrium land rent at distance $x < \lambda L/2$ from the H -CBD is given by

$$R^*(x) = \theta(\lambda L/2 - x)$$

Hence, a worker located at the average distance $\lambda L/4$ from the H -CBD bears a commuting cost equal to $\theta \lambda L/4$ and pays the average land rent $\theta \lambda L/4$ (Fujita, 1989; Papageorgiou and Pines, 1999). When the land rents go to absentee landlords, individual urban costs, defined by commuting cost plus land rent at each residence x , are given by $\theta \lambda L/2$. In order to avoid working with absentee landlords, we assume that all the land rents are collected and equally redistributed among the H -city workers.¹³ Consequently, the individual urban costs after redistribution are equal to $\theta \lambda L/4$.

Since the urban costs prevailing in regions H and F are not equal, the incentives to move from one region to the other are no longer given by (18). Indeed, we must account for the difference in urban costs between H and F , namely,

$$\lambda \theta L/4 - (1 - \lambda) \theta L/4 = (\lambda - 1/2) \theta L/2$$

which must be subtracted from (18) to obtain the actual utility differential:

$$\Delta V_u(\lambda) = [C\tau(\tau^* - \tau) - \theta L/2][\lambda - 1/2]$$

¹³ Assuming, as in Helpman (1998), that the total land rent across regions is equally redistributed among workers (hence controlling for the resulting fiscal externality) does not affect the utility differential, whereas the difference in urban costs becomes twice as large. Hence, θ is to be replaced by 2θ in the foregoing developments. This reduces the value of E and shrinks the interval (τ_1^u, τ_2^u) . As a result, Proposition 4 shows that agglomeration is less likely to occur.

As in Section 4, agglomeration is a spatial equilibrium when the slope of $\Delta V_u(\lambda)$ is positive. This is the case as long as τ falls within the two values

$$\tau_1^u \equiv \frac{\tau^* - E}{2} \quad \text{and} \quad \tau_2^u \equiv \frac{\tau^* + E}{2}$$

where $\tau_1^u, \tau_2^u \in (0, \tau^*)$, and

$$E \equiv \sqrt{(\tau^*)^2 - 2\theta L/C}$$

measures the domain of values of τ for which $\Delta V_u(\lambda) > 0$. Notice that such domain shrinks as the commuting cost per unit of distance θ increases. Consequently, we have the following proposition:

PROPOSITION 4. *If $2\theta L < C(\tau^*)^2$, there exist $\tau_1^u \in (0, \tau^*/2)$ and $\tau_2^u \in (\tau^*/2, \tau^*)$ such that agglomeration (dispersion) is the only stable equilibrium if and only if $\tau \in (\tau_1^u, \tau_2^u)$ ($\tau \notin (\tau_1^u, \tau_2^u)$). For $\tau = \tau_1^u$ or $\tau = \tau_2^u$, any distribution of workers is a spatial equilibrium. If $2\theta L > C(\tau^*)^2$, dispersion is the only spatial equilibrium.*

Thus, the existence of positive commuting costs within the regional centers is sufficient to yield dispersion when the trade costs are sufficiently low. This implies that, as trade costs fall, the economy involves dispersion, agglomeration, and redispersion. An increase (decrease) in the commuting costs fosters dispersion (agglomeration) by widening (shrinking) the left range of τ -values for which dispersion is the only spatial equilibrium. Also, sufficiently high commuting costs always yield dispersion.

It is interesting to point out that while dispersion arises for both high and low trade costs, this happens for very different reasons. In the former case, firms are dispersed as a response to the high trade costs they would incur by supplying farmers from a single agglomeration. In the latter, firms are dispersed as a response to the high urban costs workers would bear within a single agglomeration.

In the present context, we may assume that there are no farmers without annihilating the dispersion force. If $A = 0$, it is readily verified that $\tau_{\text{trade}} < \tau^*/2 < \tau_2^u$, so that dispersion does not arise when trade costs are high. Consequently, the economy moves from agglomeration to dispersion when trade costs fall, thus confirming the numerical results obtained by Helpman (1998).

Finally, it is worth revisiting the debate about the social desirability of the market outcome once we account for urban costs. Computing the first-order condition for the social optimum as in Section 5 in which we now account for the urban costs $\theta\lambda L/4$ in region H and $\theta(1 - \lambda)L/4$ in region F , we obtain the new critical values:

$$\tau_1^{uo} \equiv \frac{\tau^o - F}{2} \quad \text{and} \quad \tau_2^{uo} \equiv \frac{\tau^o + F}{2}$$

where $\tau_1^{uo}, \tau_2^{uo} \in (0, \tau^o)$, and

$$F \equiv \sqrt{(\tau^o)^2 - 2\theta L/C^o}$$

and we fall back on the condition obtained in Section 5 when $\theta = 0$. As a result, we have the following proposition:

PROPOSITION 5. *Assume $(\tau^0)^2 > 2\theta L/C^0$. Then, there exist $\tau_1^{uo} \in (0, \tau^0/2)$ and $\tau_2^{uo} \in (\tau^0/2, \tau^0)$ such that the agglomerated configuration is the optimum when $\tau \in (\tau_1^{uo}, \tau_2^{uo})$. When $\tau = \tau_1^{uo}, \tau_2^{uo}$, any configuration is an optimum. Otherwise, the symmetric configuration is the optimum.*

Observe that the domain for which agglomeration is optimal shrinks as the commuting cost θ increases.

The comparison between the optimum and equilibrium outcomes is less straightforward than it was in Section 5, with Figure 2 showing the different possible patterns. First, as in Section 5, it is readily verified that excessive agglomeration arises for intermediate values of the trade costs. Second, when urban costs are positive, the equilibrium may yield either suboptimal agglomeration or suboptimal dispersion, depending on the parameter values of the economy. In particular, if we set $A = 0$, the interval I_2 in Figure 2 disappears, so that inefficiency arises only a range of low trade costs for which “market forces do not generate enough agglomeration” (Helpman, 1998, p. 46).¹⁴

8. CONCLUDING REMARKS

Recent years have seen the proliferation of applications of the “Dixit–Stiglitz, iceberg, evolution, and the computer” framework for studying the impact of trade costs on the spatial distribution of economic activities. While these applications have produced valuable insights, they have often been criticized because they rely on a very particular research strategy.

We have proposed a different framework that is able not only to confirm those insights but also to produce new results that could barely be obtained within the standard one. Specifically, we have used this framework to deal with the following issues: (i) the welfare properties of the core-periphery model, (ii) the impact of expectations in shaping the economic space, and (iii) the effects of urban costs on the interregional distribution of activities. This suggests that our framework is versatile enough to accommodate other extensions.

So, we have shown that the main results in the literature do not depend on the specific modeling choices made, as often argued by their critics. In particular, the robustness of the results obtained in the core-periphery model against an alternative formulation of preferences and transportation seems to point to the existence of a whole class of models for which similar results would hold. However, we have also shown that those modeling choices can be fruitfully reconsidered once the aim is to shed light on different issues.

¹⁴ These results are a novel example of the ambiguous welfare properties of agglomerations that have been pointed out especially in urban economics. As argued recently by Papageorgiou and Pines (2000), such an ambiguity finds its origin in the simultaneous working of many potential sources of distortions such as various kinds of external (dis)economies, indivisibilities, and nonreplicabilities.

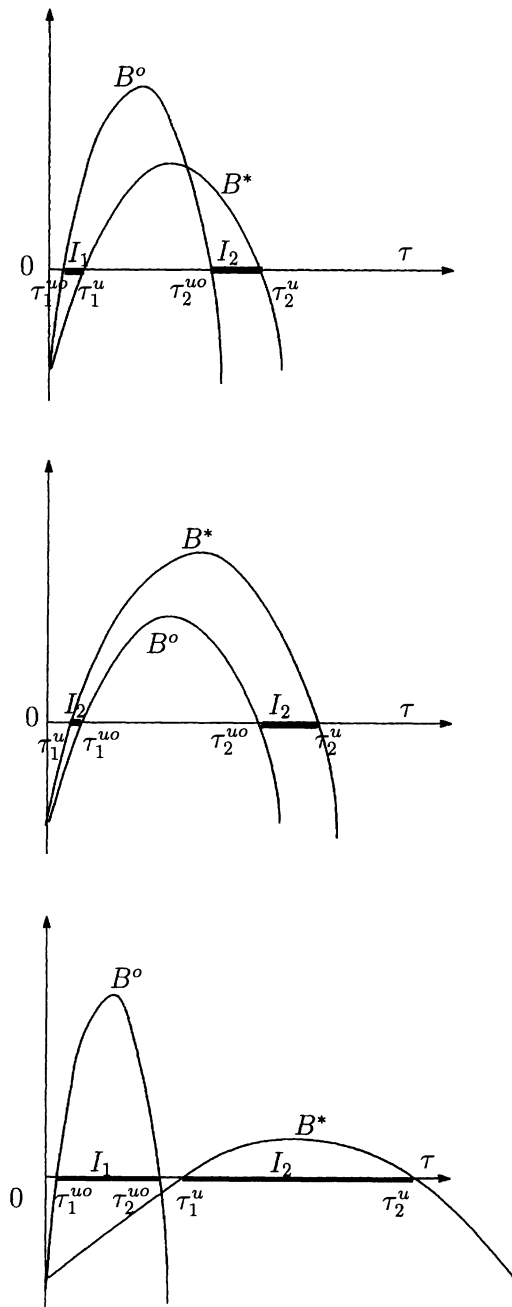


FIGURE 2

DIVERGENCE BETWEEN EQUILIBRIUM AND OPTIMUM: THE THREE CASES

The model used in this article still displays some undesirable features that should be remedied in future research. First, there is a fixed mass of firms regardless of the consumer distribution. Furthermore, by ignoring income effects, our setting has a strong partial equilibrium flavor.

APPENDIX

In the case of two varieties, the symmetric quadratic utility is given by

$$U(q_1, q_2) = \alpha(q_1 + q_2) - (\beta/2)(q_1^2 + q_2^2) - \gamma q_1 q_2 + q_0$$

In the case of $n > 2$ varieties, this expression is extended as follows:

$$\begin{aligned} U(q) &= \alpha \sum_{i=1}^n q_i - (\beta/2) \sum_{i=1}^n q_i^2 - (\gamma/2) \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + q_0 \\ &= \alpha \sum_{i=1}^n q_i - [(\beta - \gamma)/2] \sum_{i=1}^n q_i^2 - (\gamma/2) \sum_{i=1}^n \left(q_i \sum_{j=1}^n q_j \right) + q_0 \\ &= \alpha \sum_{i=1}^n q_i - [(\beta - \gamma)/2] \sum_{i=1}^n q_i^2 - (\gamma/2) \left(\sum_{i=1}^n q_i \right)^2 + q_0 \end{aligned}$$

When $\gamma \rightarrow \beta$, this utility boils to a standard quadratic utility for a homogenous good. Letting $n \rightarrow \infty$ and $q_i \rightarrow 0$, we obtain (1), in which N stands for the mass of varieties.

REFERENCES

- ALONSO, W., *Location and Land Use* (Cambridge, MA: Harvard University Press, 1964).
- , “Five Bell Shapes in Development,” *Papers of the Regional Science Association* 45 (1980), 5–16.
- ANDERSON, S. P., N. SCHMITT, AND J.-F. THISSE, “Who Benefits from Antidumping Legislation?,” *Journal of International Economics* 38 (1995), 321–37.
- DIXIT, A. K., “A Model of Duopoly Suggesting a Theory of Entry Barriers,” *Bell Journal of Economics* 10 (1979), 20–32.
- , AND J. E. STIGLITZ, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review* 67 (1977), 297–308.
- FUJITA, M., *Urban Economic Theory: Land Use and City Size* (Cambridge: Cambridge University Press, 1989).
- , AND J.-F. THISSE, “Economics of Agglomeration,” *Journal of the Japanese and International Economies* 10 (1996), 339–78.
- , P. KRUGMAN, AND A. VENABLES, *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge, MA: MIT Press, 1999).
- FUKAO, K., AND R. BÉNABOU, “History versus Expectations: A Comment,” *Quarterly Journal of Economics* 108 (1993), 535–42.
- GEYER, H. S., AND T. M. KONTULY, *Differential Urbanization: Integrating Spatial Models* (London: Arnold, 1996).
- HEAD, K., AND T. MAYER, “Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU,” *Weltwirtschaftliches Archiv* 136 (2000), 284–314.
- HELPMAN, E., “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, 1998), 33–54.
- HOTELLING, H., “Stability in Competition,” *Economic Journal* 39 (1929), 41–57.

- KOOPMANS, T. C., *Three Essays on the State of Economic Science* (New York: McGraw-Hill, 1957).
- KRUGMAN, P., "History versus Expectations," *Quarterly Journal of Economics* 106 (1991a), 651–67.
- , "Increasing Returns and Economic Geography," *Journal of Political Economy* 99 (1991b), 483–99.
- , *Development, Geography, and Economic Theory* (Cambridge, MA: MIT Press, 1995).
- , "Urban Concentration: The Role of Increasing Returns and Transport Costs," *International Regional Science Review* 19 (1996), 5–30.
- , "Space: The Final Frontier," *Journal of Economic Perspectives* 12 (1998), 161–74.
- , AND A. J. VENABLES, "Integration and the Competitiveness of Peripheral Industry," in C. Bliss and J. Braga de Macedo, eds., *Unity with Diversity in the European Community* (Cambridge: Cambridge University Press, 1990), 56–75.
- AND ———, "Globalization and the Inequality of Nations," *Quarterly Journal of Economics* 110 (1995), 857–80.
- LÖSCH, A., *Die Räumliche Ordnung der Wirtschaft* (Jena: Gustav Fischer Verlag, 1940). English translation: *The Economics of Location* (New Haven, CT: Yale University Press, 1954).
- MATSUYAMA, K., "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *Quarterly Journal of Economics* 106 (1991), 617–50.
- MCCALLUM, J., "National Borders Matter: Canada–US Regional Trade Patterns," *American Economic Review* 85 (1995), 615–23.
- MUSSA, M., "Dynamic Adjustment in the Heckscher–Ohlin–Samuelson Model," *Journal of Political Economy* 86 (1978), 775–91.
- OTTAVIANO, G. I. P., "Integration, Geography and the Burden of History," *Regional Science and Urban Economics* 29 (1999), 245–56.
- , AND D. PUGA, "Agglomeration in the Global Economy: A Survey of the 'New Economic Geography,'" *World Economy* 21 (1998), 707–31.
- PAPAGEORGIOU, Y. Y., AND D. PINES, *An Essay in Urban Economic Theory* (Dordrecht: Kluwer Academic Publishers, 1999).
- , AND ———, "Externalities, Indivisibility, Nonreplicability and Agglomeration," *Journal of Urban Economics* 48 (2000), 509–35.
- PHILIPS, L., *Applied Consumption Analysis* (Amsterdam: North-Holland, 1983).
- PUGA, D., "The Rise and Fall of Regional Inequalities," *European Economic Review* 43 (1999), 303–34.
- TABUCHI, T., "Agglomeration and Dispersion: A Synthesis of Alonso and Krugman," *Journal of Urban Economics* 44 (1998), 333–51.
- TRIONFETTI, F., "Public Procurement, Market Integration, and Income Inequalities," *Review of International Economics* 9 (2001), 29–41.
- VIVES, X., "Trade Association Disclosure Rules, Incentives to Share Information, and Welfare," *Rand Journal of Economics* 21 (1990), 409–30.
- WEI, S.-J., "Intra-national versus International Trade: How Stubborn Are Nations in Global Integration," National Bureau of Economic Research Working Paper No. 5531, 1996.