Agglomeration and Trade Revisited*

Gianmarco Ottaviano† and Jacques-François Thisse‡

19 June 1999

Abstract

The purpose of this paper 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 adopt a forward-looking approach to the dynamics of migration in the process of agglomeration instead of the myopic Marshallian model used so far in this literature.

Keywords: agglomeration, trade, self-fulfilling expectations, monopolistic competition.

J.E.L. Classification: F12, L13, R13.

*We should like to thank R. Baldwin, G. Duranton, M. Fujita, P. Jehiel, A. Linbeck, K. Middelfart-Knarvik, T. Mori, D. Peeters, D. Pines, K. Stahl as well as participants to seminars at Kyoto University, the Institute of Industrial Economics in Stockholm, CEPR ERWIT in Rotterdam, and SITE Summer Workshop in Stanford for very useful discussions and suggestions. The usual disclaimer applies. The first author is grateful to the European Commission for financial support through the TMR programme and the second author to the Fonds national de la recherche scientifique (Belgium) for funding.

†Universita degli Studi di Bologna, CORE, Université catholique de Louvain, and CEPR.

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

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 homogeneous 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, their potential has not been really explored as recognized by Krugman (1998, p.164) himself:

> To date, the new economic geography has depended heavily on the tricks summarized in Fujita, Krugman and Venables (1999) with the slogan "Dixit-Stiglitz, icebergs, evolution, and the computer"

The main tool used in 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. Because utility is described by a constant elasticity of substitution (CES) function and because each firm is negligible, the demand for any specific variety has itself a constant elasticity. 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. This is not innocuous in that it implies that any increase in the price is accompanied with a proportional increase in transport cost, which seems both unrealistic and undesirable. The use of both CES function and iceberg cost leads to a convenient setting in which demands have a constant elasticity. However, such a result conflicts with research in spatial competition where it is shown that demand elasticity varies with distance. Third, the stability analysis used to select spatial equilibria rests on ad hoc adjustment processes in which the location of production factors is driven by differences in current returns. Despite some superficial analogy with evolutionary game theory, this approach neglects the role of expectations, thus placing itself outside mainstream economics (Krugman, 1991a; Matsuyama, 1991).[1] In addition, because locational decisions are often made once-and-for-all, future flows of earnings should also be taken into account. Last, although models are based on very specific assumptions, they are often beyond the reach of analytical resolution so that authors have to appeal to numerical investigations.[2]

The purpose of this paper is to propose an alternative setting which allows one to go beyond the limits above. In particular, we present a model of agglomeration and trade that displays the main features of the work of Krugman (1991b), while being sufficiently tractable to be solved by means of analytical tools. Moreover, the model will be shown to permit a full-fledged forward looking analysis of the role of history and expectations in the emergence of economic clusters.

To this end, we consider a popular model, also used in industrial organization (Dixit, 1979; Vives, 1990) and international trade (Anderson *et al.,* 1995; Krugman and Venables, 1990) as well as in demand analysis (Phlips, 1983), that is, the *quadratic utility* model. It is well known that this model allows one to capture in a neat way the idea of product differentiation in imperfect markets.

Our model is very much in the tradition of the new economic geography. However it differs from the existing literature in several major respects. First, the functional form used for the utility is not the same as that used by Krugman and others, though it also encapsulates a preference for variety. This shows that the main tendencies toward agglomeration are robust against alternative specifications of demand, so that our results strengthen the existing literature which no longer depends on a particular specification of preferences. Second, closed analytical solutions are obtained. This means that we do not have to appeal to numerical resolution and this gives more robustness to the results. Third, since each parameter of the functions we use has a unique meaning in equilibrium, our model leads to clear-cut comparative static results and is likely to be easier to test than those based on Dixit-Stiglitz. Fourth, our model allows for asymmetry between own price and cross price effects. In particular, the price elasticity is no longer equal to the elasticity of substitution and both may vary with prices. Fifth, transportation is modeled as a costly activity that uses other resources than the transported good itself.

[1] See, however, Ottaviano (1999) for the analysis of a special case.

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

This is a more natural approach than using the iceberg transport cost and is in accordance with what has accomplished in spatial pricing theory. Last, the equilibrium concept used is broader than that employed by Dixit and Stiglitz in that it agrees with Chamberlin while yielding prices depending on all the fundamentals of the market (see also Spence, 1976, for a related approach). This permits a more detailed analysis of the forces at work as well as more explicit connections to the industrial organization literature. In the same spirit, since each parameter of our model has a univocal meaning, it is likely to be easier to test than models based on Dixit-Stiglitz.

Finally, our model allows us to determine the exact domain for 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 transport 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.*

In view of this discussion, it seems fair, therefore, to claim that our model may be considered as an alternative to the 'Dixit-Stiglitz, iceberg, evolution, and the computer' framework which has served as the main tool-box in economic geography.

The organization of the paper reflects what we have said in the foregoing. The model is presented in the next section while the equilibrium prices are determined in Section 3. The corresponding process of agglomeration is analyzed in Section 4 by using the Marshallian approach in the stability analysis. In Section 5, we show how our model can be used to compare history and expectations in the emergence of an agglomeration. Section 6 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 perfectly 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, however, that *the role of factor A is to capture the idea that some inputs* (such as land or some services) *are nontradable 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 homogeneous. Consumers have a positive initial endowment of this good[3] which 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 which is produced using L as the only input under increasing returns to scale and imperfect competition.

As in most economic geography models, 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. In addition, since each firm sells a differentiated variety, it faces a downward sloping demand, so that all the unknowns are described by density functions.[4] There are no scope economies so that, due to increasing returns to scale, there is a one-to-one relationship between firms and varieties.

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. This is a significant departure from Krugman's analysis in that this specification allows for a straightforward connection between economic geography and spatial pricing and location theories.

Preferences are identical across individuals and described by the following quasi-linear utility function which is supposed to be symmetric in all varieties:

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

[3]This is different from the assumption made in standard models of economic geography where consumers have a positive endowment in labor only. It implies that income equals wage. We will see below why we depart here from this assumption.

[4]So our model, as the continuum version of the Dixit-Stiglitz model used by Fujita *et al.* (1999), is very much in the same spirit as mainstream urban economics where quantities and prices are also expressed by density functions (see, e.g. Fujita, 1989).

$$-\gamma \int_0^N \int_0^N q(i)q(j)didj + q_0$$

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 $\beta \geq \gamma > 0$. In this expression, α is a measure of the size of the market (since it expresses the intensity of preferences for the differentiated product), whereas a large value for β means that the representative consumer is biased toward a dispersed consumption of varieties, thus reflecting a love for variety.[5] For a given value of β, the parameter γ expresses the substitutability between varieties: the higher γ, the closer substitutes the varieties. Finally, we assume that the initial endowment $\overline{q}$ in the numéraire is large enough for the consumption of the numéraire to be strictly positive at the market and optimal solutions.

In Ottaviano and Thisse (1999), we identify a simple condition on the parameters of (1) for our representative consumer to have a taste for variety:

$$\beta > 2\gamma \tag{2}$$

In other words, the quadratic utility function exhibits a preference for variety when the product is differentiated enough. Clearly, the smaller is γ and/or the larger is β, the stronger is the preference for variety.

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

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

where y is the individual's labor income, $p(i)$ is the price of variety i while the price of the agricultural good is normalized to one. The initial endowment $\overline{q}_0$ is supposed to be large enough for the optimal consumption of the numéraire to be strictly positive for each individual regardless of his labor income.

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

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

[5]The intuition behind this interpretation is very similar to the the one that stands behind the Herfindahl index used to measure industrial concentration. Controlling for the total amount of the differentiated good consumption, the absolute value of the quadratic term in (1) increases with concentration of consumption on few varieties, thus decreasing utility.

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

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

where $a \equiv \alpha/(\beta + N\gamma)$, $b \equiv 1/(\beta + N\gamma)$ and $c \equiv \gamma/\beta(\beta + N\gamma)$.

As usual, the parameter c represents the degree of product differentiation among varieties: they are independent when $c = 0$ and perfect substitutes when $c = b$. In other words, 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 $Na - bpN$ 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. Finally, the condition (2) becomes $b > 2c$.

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

$$\begin{aligned} V(y; p(i), i \in [0, N]) &= -a \int_0^N p(i)di + \frac{b + cN}{2} \int_0^N [p(i)]^2 di \\ &\quad -c \int_0^N \int_0^N p(i)p(j)didj + y + \overline{q}_0 \end{aligned} \tag{4}$$

While Krugman (1991b) and others define a Cobb-Douglas preference on the homogeneous and differentiated goods with CES subutility, we assume instead a *quasi-linear preference with a quadratic subutility.*[6] This difference in preferences explains why we make a different assumption about individuals' initial endowments. In Krugman, a positive endowment in a good different from labor would imply an income higher than wage, thus yielding a much more complex framework. By contrast, here a positive endowment is assumed in order for each individual to be able to consume at the maximum of his quadratic subutility. Observe also that both utilities correspond to two rather extreme cases: the former assumes a unit elasticity of substitution,

[6] A similar assumption is made by Spence (1976) in studying monopolistic competition.

the latter an infinite elasticity between the differentiated product and the numéraire. Finally, since the marginal utility of the numéraire is constant and equal across all consumers, our model can be used in welfare comparisons and is known not to generate any perverse effect when it is cast into a general equilibrium framework. Though this model has considerable merits, we must acknowledge the fact that it has a strong partial equilibrium flavor to the extent that it eliminates income effects.

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, $w_A^H = w_A^F = 1$. Technology in manufacturing requires ϕ units of L in order to produce any amount of a variety, i.e. 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, makes sense here unlike in Dixit and Stiglitz (1977) because our preferences imply that firms use an absolute markup instead of a relative one when choosing prices.

In the analysis below, we will encounter several conditions involving the parameters a, b, c and τ. Throughout the paper, we will assume that the least demanding condition regarding these three parameters, that is,

$$\tau \leq a/b \tag{5}$$

is always satisfied. This condition must hold for a spatial monopolist producing at zero marginal cost to sell in the foreign market, regardless of whom pays for the transport cost.

Labor market clearing implies that

$$n_H = \lambda L/\phi \tag{6}$$

and

$$n_F = (1-\lambda)L/\phi \tag{7}$$

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 $\phi \to 0$ (or $L \to \infty$), the mass of varieties becomes arbitrarily large. In addition, (6) and (7) show that the larger labor market is also the region accommodating the larger mass of firms. In order to ease the burden of notation, we can set $L = \phi$ by appropriately choosing the unit of L so that $N = 1$.

Entry and exit are free so that profits are zero in equilibrium. Hence, (6) and (7) 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. By (6) and (7), the demand and supply of workers in each region are equal. As in Krugman (1991b), the corresponding equilibrium wages are then determined by a bidding process between firms which ends when no firm can earn a strictly positive profit at the equilibrium market prices.

By assumption, firms compete in segmented markets, that is, each firm sets a price specific to the market in which its product is sold. Indeed, even for very low transport costs, there are many good reasons to believe that firms will succeed anyway to segment markets (Horn and Shy, 1996), while empirical work confirms the assumption that international markets are still very segmented (see, e.g. McCallum, 1995). 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 Roy's identity, demands for a representative firm located in H arising in region H (q^{HH}) and region F (q^{HF}) are given respectively by:

$$q^{HH} = a - (b + c)p^{HH} + cP^{H} \tag{8}$$

and

$$q^{HF} = a - (b + c)p^{HF} + cP^{F} \tag{9}$$

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 and P^F can be interpreted as the price index prevailing in region H and F, respectively.

A representative firm in H maximizes its profits defined by:

$$\Pi^{H} = p^{HH}q^{HH}(p^{HH})(A/2+\lambda L) + (p^{HF}-\tau)q^{HF}(p^{HF})[A/2+(1-\lambda)L] - \phi w^{H} \tag{10}$$

where $A/2$ stands for the number of farmers in each region.

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 (8) and (9)). 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 non-negligible impact on its behavior.*

Our model thus provides an alternative way to tackle monopolistic competition, more appealing than Dixit and Stiglitz because some degree of interaction among firms is involved at the market level. As a result, the equilibrium prices will depend on all the key aspects of the market instead of being given by a simple relative mark-up 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:

$$p_D^{HH} = \frac{1}{2}\frac{2a + \tau c(1-\lambda)}{2b + c} \tag{11}$$

$$p_D^{FF} = \frac{1}{2}\frac{2a + \tau c\lambda}{2b + c} \tag{12}$$

$$p_D^{HF} = p_D^{FF} + \frac{\tau}{2} \tag{13}$$

$$p_D^{FH} = p_D^{HH} + \frac{\tau}{2} \tag{14}$$

Unlike the (spatialized) Dixit-Stiglitz model, *the equilibrium prices under monopolistic competition depend on the firm distribution between regions*. In particular, both the prices charged by 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 H, say, 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 (5) holds. All these results are in accordance with what is known in industrial organization and spatial pricing theory, and are more explicit than in the Dixit-Stiglitz model.

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

$$p_D^{HF} - p_D^{HH} = \tau\frac{b+\lambda c}{2b+c}, \text{ ...which is equal to...}\frac{\tau}{2} \text{ when } \lambda = 1/2$$

$$p_D^{FH} - p_D^{FF} = \tau\frac{b+(1-\lambda)c}{2b+c}, \text{...which is equal to...}\frac{\tau}{2} \text{ when } \lambda = 1/2$$

It is well known that a monopolist facing a linear demand absorbs exactly one-half of the transport cost. Hence, we see that monopolistic competition leads to more (less) freight absorption than monopoly when the foreign market is the small (large) one: in their 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_D^{HH} (resp. p_D^{FF}) is increasing in τ because the local firms in H (F) are more protected against foreign competition while $p_D^{HF} - \tau$ (resp. $p_D^{FH} - \tau$) is decreasing because it is now more difficult for these firms to sell on the foreign market. 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.

Substracting the unit transport cost τ from the prices set on the distant markets, i.e. (13) and (14), we see that firms' prices net of transport costs are positive regardless of the workers' distribution if and only if

$$\tau < \tau_{trade}^{D} \equiv \frac{2a}{2b+c} \tag{15}$$

The same condition must hold for consumers in F (H) to buy from firms in H (F), i.e. for the demand (9) evaluated at the prices (11) and (12) to be positive for all λ. From now on, condition (15) is assumed to hold. Consequently, there is intra-industry trade and reciprocal dumping, as in Anderson *et al.* (1995).

Finally, local sales rise with τ because of the higher protection enjoyed by the local firms but exports fall for the same reason.

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

$$\Pi^{HH} = (b + cN)(p_D^{HH})^2(A/2 + \lambda L) \tag{16}$$

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

$$\Pi^{HF} = (b + cN)(p_D^{HF} - \tau)^2[A/2 + (1 - \lambda)L] \tag{17}$$

Increasing λ has two opposite effects on Π^{HH}. First, the equilibrium price (11) falls as well as the quantity of each variety bought by each consumer living in region H. However, the total population of consumers residing in this region is now larger so that the profits made by a firm located in H 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 individual demand effect.*

The equilibrium wage prevailing in region H may be obtained by evaluating $(\Pi^{HH} + \Pi^{HF})/L$ at the equilibrium prices but, unfortunately, the corresponding expression turns out to be especially cumbersome. For the analysis developed in section 5, it is sufficient to study their behavior in the vicinity of $\lambda = 1/2$. Differentiating (16) and (17) with respect to λ yields

$$\frac{d\Pi^{HH}}{d\lambda} \Big|_{\lambda=1/2} = \frac{(b+c)}{8(2b+c)^2}(4a + c\tau)(2aL - cA\tau) \tag{18}$$

and

$$\frac{d\Pi^{HF}}{d\lambda}\mid_{\lambda=1/2}= -\frac{(b+c)}{8(2b+c)^2}(4a-c\tau-4b\tau)(2aL-cA\tau-cL\tau-2bL\tau) \quad (19)$$

The inspection of (18) and (19) reveals that w^H is not necessarily monotonic with respect to λ because $dw^H/d\lambda$ evaluated at $\lambda = 1/2$, obtained from the sum of (18) and (19), changes sign and is $\cap$-shaped with respect to τ. This is so because more workers in H also means more firms located in this region, thus making the final impact on the local wage ambiguous. Standard analysis shows that both (18) and (19) are $\cap$-shaped; (18) is positive for low τ and negative for large τ while (19) is negative for low and high values of τ and positive for intermediate values. As a result, four domains of τ are to be considered as τ rises from zero: (i) (18) is positive and (19) negative; (ii) (18) and (19) are both positive; (iii) (18) is negative but (19) is still positive; and (iv) both (18) and (19) are negative. Since (18) and (19) describe two concave parabolas, the locus describing $dw^H/d\lambda$ evaluated at $\lambda = 1/2$ is also a concave parabola with a positive intercept, a single maximizer arising at

$$\tau_{\max} = \frac{a}{4b+c+cA/L}$$

and a positive zero, corresponding to the highest wage prevailing at a dispersed equilibrium, given by

$$\tau_0 = \frac{a}{2b+c+cA/L} > \tau_{\max}$$

Hence the relocation of some workers (and firms) in region H depresses the local wage when τ exceeds τ_0, the more so the higher τ. On the other hand, for τ lower than τ_0, more workers in H increases the local wage. However, the marginal impact of an increase of λ above 1/2 upon the equilibrium wage reaches its peak at $\tau_{\max}$. This implies that such a positive impact becomes weaker and weaker as τ gets closer to zero.

4 When do we observe agglomeration?

We now ask whether for a given spatial distribution of skilled workers, $(\lambda, 1-\lambda)$, there is an incentive for them to migrate and, if so, what direction the flow of migrants will take. A *spatial equilibrium* arises when no skilled worker may get a higher utility level in the other region[7]:

$$\Delta V(\lambda) \equiv V_H(\lambda) - V_F(\lambda) = 0 \quad \text{for } \lambda \in (0,1)$$

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 such an equilibrium, 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 (6) and (7) remain valid for the new distribution of workers. Wages are then adjusted in each region for each firm to earn zero profits everywhere. In other words, the driving force in the migration process is workers' indirect utility differential between H and F:

$$\dot{\lambda} \equiv d\lambda/dt = \lambda(1-\lambda)\Delta V(\lambda) \tag{20}$$

when t is time. Clearly, a location 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 opposite direction.

An equilibrium is *stable* for (20) if, for any marginal deviation from the equilibrium, the equation of motion above brings the distribution of firms 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 negative in a neighborhood of this point.

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, (11) and (14) imply that the equilibrium prices of all varieties sold in H are lower. Both effects generate a higher indirect utility. (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 leads more firms to locate there. In other words, both backward and forward linkages

[7]Note that our dynamics is identical to that of Krugman since comparing real wages is equivalent to comparing indirect utility levels in his model.

are present in our model, though the constitutive effects are somewhat richer than those uncovered by Krugman.

The indirect utility differential is obtained by plugging the equilibrium prices (11)-(14) and, using (6) and (7), the equilibrium wages for the workers into (4):

$$\begin{aligned} \Delta V \equiv V^H - V^F = & \frac{\tau(b+c)}{2(2b+c)^2}(\lambda - 1/2)\{12a(b+c) \\ & -\tau[2b(3b+2c+cA/L)+c^2(A/L-1)]\} \end{aligned} \tag{21}$$

It follows immediately from this expression that $\lambda = 1/2$ is always an equilibrium. For $\lambda \neq 1/2$, the indirect utility differential has always the same sign as $\lambda - 1/2$ if and only if the curly bracketed term is positive. This is always the case when the coefficient of τ in (21) is nonpositive, that is,

$$A/L \leq \frac{c^2 - 2b(3b+2c)}{c(2b+c)} \tag{22}$$

which means that the population of workers is relatively large. In this case, the advantages of having a large home market dominate the disadvantages incurred while supplying a distant periphery as long as $\tau \leq \tau^D_{trade}$.

More interesting is the case when (22) does not hold because there exists a critical value of τ below (above) which the sign of ΔV is positive (negative):

$$\tau^D \equiv \frac{12a(b+c)}{2b(3b+2c+cA/L)+c^2(A/L-1)} \tag{23}$$

which is always positive since (22) is violated:

$$A/L > \frac{c^2 - 2b(3b+2c)}{c(2b+c)} \tag{24}$$

It remains to determine when τ^D is lower than τ^D_{trade} for the analysis to be consistent. This is so if and only if

$$A/L > \frac{7c^2 + 14bc + 6b^2}{c(2b+c)} > 1 \tag{25}$$

which means that the population of farmers is relatively large. Note that (25) is more demanding than (24).

When τ^D exists and when $\tau < \tau^D$, 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 transportation rate is low enough*, as in Krugman (1991a) and for similar reasons. Also, when increasing returns are stronger, as expressed by higher values of $L = \phi$, it follows from (23) that τ^D rises. This means that *the agglomeration of the manufacturing sector is more likely, the stronger are the increasing returns at the firm's level.* Observe, finally, that τ^D increases with product differentiation if and only if $A > L$ which is granted by (25). In other words, *more product differentiation fosters agglomeration.*

In contrast, for large transport costs, that is, when $\tau > \tau^D$, it is straightforward to see that the symmetric configuration is the only stable equilibrium. When τ^D exists, it corresponds to both the critical value of τ below which symmetry is no longer stable; the break point and the sustain point, that is, the value below which agglomeration is stable, are identical; this follows from the fact that (21) is linear in λ.

Proposition 1 *Assume that discriminatory pricing prevails and that $\tau < \tau^D_{trade}$. Two cases may arise:*

(i) When (22) holds, any stable equilibrium involves agglomeration.

(ii) When (22) does not hold and (25) is satisfied, we have the following: if $\tau > \tau^D$, then the symmetric configuration is the only stable equilibrium with trade; if $\tau < \tau^D$ there are two stable equilibria corresponding to the agglomerated configurations with trade; if $\tau = \tau^D$ there is a continuum of equilibria.

When $\tau^D_{trade} > \tau^D$, 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 urban region accommodating the manufacturing sector only imports the homogeneous good from the rural region.

Observe that c very small implies that $\tau^D_{trade} < \tau^D$ so that agglomeration always arises under trade. This is confirmed by the special case where varieties are independent in consumers' preferences. Indeed, (15) becomes $\tau < a/b$ while $\tau < \tau^D$ reduces to $\tau < 2a/b$. This says that the indirect utility differential is positive for all admissible values of τ. In this case, we know that there is always agglomeration in equilibrium (unless the initial value of λ is 1/2).

5 The impact of workers' expectations on the agglomeration process

The adjustment process (20) is often used in economic geography models. Yet, the underlying dynamics is myopic because workers care only about their current utility level, thus implying that only history matters. This strikes us as a pretty naive 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.

Since workers have perfect foresights, the easiest way to generate a non bang-bang migration behavior is to assume that workers face a costly intertemporal migration decision. When moving from one region to the other, migrants incur a utility loss which depends on the rate of migration $\dot{\lambda}$ because a migrant imposes a negative externality on the others by congesting the migration process (Mussa, 1978). In order to capture the hardships linked to large migration flows, we follow Krugman (1991a) and assume that the utility loss for a migrant is equal to $\left|\dot{\lambda}\right|/\delta$, where $\delta \in (0, +\infty)$ has the same meaning as in (20).

Following Fukao and Bénabou (1993) as well as Ottaviano (1999), we now define the utility of a worker residing in region H as

$$v^H(t) = \int_t^T e^{-\rho(s-t)} V^H(s) ds + e^{-\rho(T-t)} v^H(T) \tag{26}$$

where T is the first time when all workers are established into a single region and ρ the discount rate, while a similar expression holds for $v^F(t)$. Hence the motion equation (20) is to be replaced by

$$\dot{\lambda} = \delta \Delta v \equiv \delta(v^H - v^F)$$

Since workers are free to choose where to reside and since the individual migration cost at time t is equal to $\left|\dot{\lambda}\right|/\delta$, for a worker in region H to stay put it must be that the utility in this region (weakly) exceeds the utility in region F minus the migration cost:

$$v^H(t) \geq v^F(t) - \left|\dot{\lambda}(t)\right|/\delta \quad \text{where the equality holds when } \dot{\lambda}(t) < 0 \tag{27}$$

whereas a similar expression holds for someone living in region F. This means that $v^H(t) - v^F(t)$ stands for the private value for a worker to be in H instead of F. In the sequel, we consider the case of segmented markets only, but other pricing rules could be similarly analyzed.

Assuming an interior solution for (27), we easily get

$$\dot{\lambda} = \delta \Delta v \tag{28}$$

while differentiating (26) yields

$$\dot{\Delta v} = \rho \Delta v - \Delta V \tag{29}$$

where $\Delta V \equiv V^H - V^F$ stands for the instantaneous indirect utility differential flow given by (21). Hence we obtain a system of two linear differential equations instead of the first order differential equation (20). Since ΔV is linear in λ, we can simplify notation by defining two constants η_0 and η_1 such that $(\eta_0 + \eta_1 \lambda) \equiv \Delta V$ where

$$\eta_1 \equiv \frac{\tau(b+c)}{2(2b+c)^2}\{12a(b+c) - \tau[2b(3b+2c+cA/L) + c^2(A/L-1)]\}$$

It then follows from (23) that $\eta_1 > 0$ if and only if $\tau < \tau^D$ so that agglomeration occurs as long as $\eta_1 > 0$.

Since $\lambda = 1/2$ implies $\Delta V = 0$, this system has a steady state at $(\lambda, \Delta v) = (1/2, 0)$ which corresponds to the dispersed configuration. Consider now its stability. The eigenvalues of the Jacobian matrix of this system evaluated at $(1/2, 0)$ are given by

$$\frac{\rho \pm \sqrt{\rho^2 - 4\delta\eta_1}}{2} \tag{30}$$

When $\eta_1 < 0$ (that is, $\tau > \tau^D$), the two eigenvalues are real and have opposite signs. Then, the steady state is a saddle point so that one always converges towards the dispersed configuration, thus implying that neither history nor expectations matter for the final outcome.

Assume now that $\eta_1 > 0$ (that is, $\tau < \tau^D$) so that agglomeration must occur. Two cases may now arise. In the first one, $0 < \eta_1 < \rho^2/4\delta$ so that the two eigenvalues are positive. The steady state $(1/2, 0)$ is an unstable node and there are two trajectories that steadily go to the endpoints $(0, 0)$ and $(1, 0)$, depending on the initial conditions. In this case, *only history matters*:

from any initial $\lambda \neq 1/2$, there is a single trajectory that goes towards the closer endpoint, as in the case where the dynamics is given by (20). In this case, the myopic adjustment process studied in the above section provides a good approximation of the evolution of the economy under forward looking behavior.

Things turn out to be quite different when $\eta_1 > \rho^2/4\delta$. The two eigenvalues are complex and have a positive real part so that the steady state is an unstable focus. The two trajectories now spiral out from $(1/2, 0)$. Therefore, for any initial λ 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 is going to move. 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 both history and 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:

$$\lambda_L \equiv \frac{1}{2}\left[1 - \exp\left(-\frac{\rho\pi}{\sqrt{4\delta\eta_1 - \rho^2}}\right)\right]$$

$$\lambda_H \equiv \frac{1}{2}\left[1 + \exp\left(-\frac{\rho\pi}{\sqrt{4\delta\eta_1 - \rho^2}}\right)\right]$$

Clearly, the overlap is an interval centered around $\lambda = 0.5$ whose width is:

$$\Lambda \equiv \exp\left(-\frac{\rho\pi}{\sqrt{4\delta\eta_1 - \rho^2}}\right)$$

It exists as long as $\eta_1 > \rho^2/4\delta$. Thus, the width of the overlap is increasing in η_1 and δ, while it decreases with ρ. Since the parameter $\eta_1 > 0$ is the slope of ΔV, it expresses the strength of the forward and backward linkages pushing towards agglomeration. Since Λ rises with η_1, we see that expectations matter more when the forward and backward linkages are stronger.

We know that expectations matter for values of τ such that $\eta_1 > \rho^2/4\delta$. Since $\eta_1 = 0$ when $\tau = 0$ and $\tau = \tau^D$, the equation $\eta_1 - \rho^2/4\delta = 0$ has two positive roots in τ, denoted τ_1^e and τ_2^e, smaller than τ^D. Furthermore, since η_1 is concave in τ, the role of expectations matter when τ falls in between these two roots. Consequently, we have shown the following result:

Proposition 2 *Workers' expectations about their future earnings influence the process of agglomeration if and only if* $\tau \in [\tau_1^e, \tau_2^e]$.

Hence, history alone matters when τ is large enough or small enough. In other words, the agglomeration process evolves as if workers were short-sighted for high and low values of τ. Moreover, the domain of values of τ for which expectations matter shrinks when the discount rate ρ gets larger or when the speed of adjustment δ decreases. When τ belongs to this domain, *the equilibrium is determined by workers' expectations and not by history as long as regions are not initially too different.*

The existence of range $[\tau_1^e, \tau_2^e]$ for intermediate values of τ can be explained in terms of the issues discussed at the end of Section 3. There we showed that the positive effect of workers' immigration on local wages reaches a maximum at $\tau_{\max} > 0$. In other words, the 'complementarity' of workers' migration decisions reaches its maximum strength for a positive level of transport costs below which it becomes weaker and weaker as one gets closer to free trade ($\tau = 0$). This is crucial to understand the circumstances under which a whirling trajectory is viable.

Suppose, indeed, that the economy is such that $\lambda(0) > 1/2$ and ask what is needed to reverse the agglomeration process 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 that period. 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 2. When complementarity of workers locations leads to substantial wage rises (that is, for intermediate values of τ), the benefits of agglomerating at $\lambda = 0$ can compensate workers for the losses they incur during the transition phase, thus making the reversal of migration possible. On the contrary, when complementarity of workers locational decisions gets weaker (that is, for low or high values of τ), the benefits of agglomerating at $\lambda = 0$ do not compensate workers for the losses.

As a consequence, *the reversal in the migration process may occur only for intermediate values of* τ. Clearly, proximity to the endpoint increases the time period over which workers bear losses because it takes more time to reach $\lambda = 1/2$. As observed before, a large rate of time preference gives more weight to the losses, while a slow speed of adjustment extends the time period over which workers bear losses.

6 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 shown that *an alternative model is able not only to confirm those insights but also to produce new results that could barely be obtained within the standard framework*. In particular, we have proposed a more realistic description of the process of spatial competition in which, as opposed to the standard model, the pricing decisions of firms are affected by the spatial location of their competitors as well as by the distribution of demand. This has allowed us to come up with a richer set of results regarding the impact of the fundamentals of the economy on the equilibrium prices. This model has then been used to deal with the intrinsically dynamic issue of the lock-in effect of historical events on the spatial distribution of activities.

So, on the one hand, 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. Note, especially, that the robustness of the results obtained in the core-periphery model against alternative formulations of preferences and transportation modeling suggests there is a whole class of models for which similar results would hold. On the other hand, we have also shown that these modeling choices are to be reconsidered once the aim is to shed light on other relevant issues.

The model used in this paper still displays some undesirable features such as a fixed mass of firms regardless of the consumer distribution. In particular, as said above, by ignoring income effects, our setting has a strong partial equilibrium flavor to which it should be remedied in future research. Furthermore, the concentration of workers within a single region does not generate any negative external effects such as the crowding of the housing market, pollution and crime, as in Helpman (1998). It is reasonable to conjecture that the existence of such effects could lead to the emergence of stable asymmetric equilibria in which the benefits of agglomeration are just balanced by the negative external effects in the larger region.

References

[1] Anderson, S.P., N. Schmitt and J.-F. Thisse (1995) Who Benefits from Antidumping Legislation ? *Journal of International Economics* 38, 321-337.

[2] Dixit, A.K. (1979) A Model of Duopoly Suggesting a Theory of Entry Barriers, *Bell Journal of Economics* 10, 20-32.

[3] Dixit, A.K. and J.E. Stiglitz (1977) Monopolistic Competition and Optimum Product Diversity, *American Economic Review* 67, 297-308.

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

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

[6] Fujita, M. and J.-F. Thisse (1996) Economics of Agglomeration, *Journal of the Japanese and International Economies* 10, 339-378.

[7] Fukao, K. and R. Bénabou (1993) History versus Expectations: A Comment, *Quarterly Journal of Economics* 108, 535-542.

[8] 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).

[9] Horn, H. and O. Shy (1996) Bundling and International Market Segmentation, *International Economic Review* 737, 51-69.

[10] Hotelling, H. (1929) Stability in Competition, *Economic Journal* 39, 41-57.

[11] Koopmans, T.C. (1957) *Three Essays on the State of Economic Science* (New York: McGraw-Hill).

[12] Krugman, P. (1991a) History versus Expectations, *Quarterly Journal of Economics* 106, 651-667.

[13] Krugman, P. (1991b) Increasing Returns and Economic Geography, *Journal of Political Economy* 99, 483-499.

[14] Krugman, P. (1995) *Development, Geography, and Economic Theory* (Cambridge (Mass.): MIT Press).

[15] Krugman, P. (1998) Space: The Final Frontier, *Journal of Economic Perspectives* 12, 161-174.

[16] Krugman, P. and A.J. Venables (1990) 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).

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

[18] Matsuyama, K. (1991) Increasing Returns, Industrialization, and Indeterminacy of Equilibrium, *Quarterly Journal of Economics* 106, 617-650.

[19] McCallum, J. (1995) National Borders Matter: Canada-US Regional Trade Patterns, *American Economic Review* 85, 615-623.

[20] Mussa, M. (1978) Dynamic Adjustment in the Heckscher-Ohlin-Samuelson Model, *Journal of Political Economy* 86, 775-791.

[21] Ottaviano, G.I.P. (1999) Integration, Geography and the Burden of History, *Regional Science and Urban Economics* 29, 245-256.

[22] Ottaviano, G.I.P. and J.-F. Thisse (1999) Monopolistic Competition, Multiproduct Firms and Optimum Product Diversity, CORE Discussion Paper n° 9919.

[23] Phlips, L. (1983) *Applied Consumption Analysis* (Amsterdam: North-Holland).

[24] Puga, D. (1999) The Rise and Fall of Regional Inequalities, *European Economic Review* 43, 303-334.

[25] Spence, A.M. (1976) Product Selection, Fixed Costs and Monopolistic Competition, *Review of Economic Studies* 43, 217-236.

[26] Vives, X. (1990) Trade Association Disclosure Rules, Incentives to Share Information, and Welfare, *Rand Journal of Economics* 21, 409-430.