

An Economic Theory of Regional Clusters*

Paul Belleflamme[†], Pierre Picard[‡] and Jacques-François Thisse[§]

22 November 1999

Abstract

This paper investigates the impact of localization economies on firms' locations. It is known that such external effects lead to substantial cost reductions when firms are located together. However, when they are agglomerated, firms also face the prospects of tough price competition whose intensity can be relaxed through product differentiation. In addition, their access to isolated markets varies with the level of transport costs. As a result, there is a trade-off whose solution depends on the structural parameters of the economy. The market and optimal solutions are compared for the case of small and large groups of firms.

Keywords: cluster, trade, localization economy, market structure

JEL Classification: F12, R12

*We thank two referees and J. Brueckner for useful suggestions as well as M. Fujita and especially T. Tabuchi for stimulating discussions. The third author is grateful to the Fonds national de la recherche scientifique (Belgium) for financial support.

[†]Department of Economics, Queen Mary and Westfield College.

[‡]University of Manchester and CREW, Facultés Universitaires Notre-Dame de la Paix, Namur

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

1 Introduction

The purpose of this paper is to show that, in a world of globalization, location still matters although its impact on economic agents differs from what it was in the past. The modern paradox of location is well summarized in the following quotation by Porter [15, p.90]:

“In a global economy - which boasts rapid transportation, high-speed communication, and accessible markets - one would expect location to diminish in importance. But the opposite is true. The enduring competitive advantages in a global economy are often heavily local, arising from concentrations of highly specialized skills and knowledge, institutions, rivals, related business, and sophisticated customers.”

The main reason behind this paradox lies in the fact that technologies that can effectively be used in a given area often depend on many local factors. In the same vein, Prescott [17] observes that all existing theories of international income differences fail to explain the huge differences in living standards, and so probably because these theories do not integrate the diversity of local conditions fostering or deterring the adoption of new technologies. Similarly, although she follows quite a different approach, Saxenian [19] argues that the institutional and economic environment influencing the collective process of learning within a given area is probably as important as microeconomic linkages between firms and other economic agents.

Such differences between locales may be apprehended through Marshallian externalities since these externalities aim precisely at accounting for the benefits associated with the formation of different types of economic agglomerations at particular places [4]. The now standard classification of Marshallian externalities is between (i) localization economies, which refer to the benefits generated by the proximity of firms producing similar goods,

and (ii) urbanization economies, which account for all the advantages associated with the overall level of activity prevailing in a particular area. Both these external effects have been studied extensively in urban economics (see e.g., Henderson [7] and the references therein) and their existence is a well-documented fact.¹ From our point of view, it is worth stressing that the main distinctive feature of Marshallian externalities is the fact that their impact on economic agents is *local*, namely only the agents situated in the same area benefit from their positive impact.

In this paper, we therefore follow a well-established tradition in urban economics by assuming that firms belonging to the same sector benefit from a higher productivity when they locate together. For simplicity, we follow Chipman’s [1] modeling strategy by assuming that localization economies lead the marginal production cost prevailing in a locale to be a decreasing function of the number of similar firms established there. We use the “short-cut” associated with Marshallian externalities not only because of its convenience in view of the difficulty to study the details of the interplay between competition across locales and social interactions within each locale,² but also because we concur with Porter [16] for whom *clusters cannot be understood without explicit reference to competition and to the new role played by location in the global economy*.

Specifically, we see the formation and size of clusters as depending on the relative strength of three distinct forces: the magnitude of localization economies, the intensity of price competition, and the level of transport costs. It is well known from industrial organization that geographical proximity renders competition on the product market fiercer, thus inducing firms to

¹For recent contributions, see [2, 5, 6, 8 and 20].

²Social interactions may lead to the formation of agglomerations when firms enjoy spillovers (Ogawa and Fujita, [11]) and/or when learning-by-doing is a localized process (Soubeyran and Thisse, [21]). However, these contributions do not study the impact of competition on the existence and stability of clusters.

locate far apart [23]. This implies that firms' decisions to congregate or to separate depends on the relative intensity of localization economies and of price competition. This is not the end of the story, however. Even if price competition is relaxed through product differentiation, it is still true that firms want to be separated when transport costs (broadly defined in order to include all the impediments to trade) are high. Since the emergence of industrial clusters is generally confined to small geographical areas, it is reasonable to assume that the spatial consumer distribution is unaffected by firms' locational behavior. Hence, the cost reduction associated with the agglomeration may be more than offset by the fall in exports. Consequently, transport costs have to be sufficiently low for firms to gather. Collecting all these arguments together, we observe that *firms must be able to serve almost equally all markets* (globalization) *in order to enjoy the local advantages associated with the formation of a cluster* (localization).

Therefore, this paper can be viewed as an attempt to cast Porter's ideas about regional clusters within the realm of microeconomics. It is worth stressing from the outset that our approach differs in several respects from Krugman's (1991). First, the forces at work in our setting are different. On the one hand, while Krugman assumes monopolistic competition and pecuniary externalities fed by an expansion of local demands, we choose to focus on technological externalities and price competition. On the other hand, while Krugman's results hinge on workers' mobility, consumers' immobility is here a strong dispersion force. Second, although it can be cast within a general equilibrium framework, our model has a strong partial equilibrium flavor while Krugman works with a straight general equilibrium model. However, our model is well suited to the study of the formation of regional clusters in specific industries in which demand is dispersed and exogenous. By contrast, economic geography models pertain to the formation of regional imbalance at the level of large aggregates.

Previewing some of our main results, we may say that the formation of regional clusters seems to obey the same general principles: *full or partial agglomeration of firms into one region occurs when transport costs are low, when products are differentiated enough, and when localization economies are strong*. Some of these results are reminiscent of those obtained by Krugman [9] and Fujita *et al.* [3] in economic geography models. However, the differences are many. First, the bifurcation is smooth instead of being discontinuous as in Krugman. More precisely, we show that, as the level of transport costs (respectively, the intensity of localization economies) falls (respectively, rises), the dispersed equilibrium ceases to be stable but, unlike existing economic geography models, asymmetric clusters happen to be the only stable equilibria *despite* the fact that the spatial distribution of demand is assumed to be symmetric; full agglomeration arises only as the limiting case of a gradual process. This difference in results is due to the fact that, as explained in the foregoing, consumers' locations and, therefore, firms' demands are unaffected by firms' locational choices. Hence, the spatial agglomeration of firms is governed by lower and lower production costs, provided that products are sufficiently differentiated to relax price competition within the expanding cluster. By contrast, in Krugman and others, pecuniary externalities are expressed through a demand size effect. In other words, a process of agglomeration starts only when workers/consumers move together with firms, thus tending to make sudden the change in the spatial distribution of firms.

Second, our model is well suited for a welfare analysis that reveals some unsuspected results. For example, we show that, in the duopoly case, agglomeration is always socially optimal but may fail to take root in the region offering the largest potential for interactions; worse: firms may choose to locate apart. A similar result holds in the large group case: the market outcome is more dispersed than the optimum. Moreover, a second best analysis

in which the planner controls only firms' locations reveals that a degree of agglomeration of firms higher than that of the first best allows one to reduce the equilibrium prices in the large cluster.

We now describe the remaining of the paper. Since the underlying economic and social structure matters for the localization economies, it is reasonable to suppose that the intensity of these economies varies across regions. Thus, in Section 2, we consider the case of an oligopoly in which the production cost reduction firms may enjoy by being together changes with the region they choose. This setting provides a benchmark to study the strategic decisions of a small number of large firms in that each firm is aware that its locational choice affects not only its production cost but also its rival's. It captures some critical elements of the trade-off faced by large firms, such as those belonging to the German chemical clusters.

In Section 3, we move to the case of a large group of firms in which there is no explicit strategic interaction because each firm has a negligible impact on the others; however, each firm is aware that its locational choice has an impact on its production cost because it depends on where its competitors are located. This allows us to study the emergence of clusters involving a large number of small firms. This seems to fit the case of many industrial districts, that is, locales that accommodate a large number of small firms producing similar goods and that benefit from the localized accumulation of skills associated with workers residing in these locales [18]. Industrial districts seem to share one basic feature, namely the fact that knowledge is embodied in workers living within small geographical areas and who interact together through various social processes, such as informal discussions among workers in each firm, inter-firm mobility of skilled workers, the exchange of ideas within families or clubs, and bandwagon effects. The impact of such complex interactions can be studied through localization economies acting within the limits of a well-defined area. Note also that our approach is consistent with

the fact that, in many Italian industrial districts, workers stay put. Our concluding remarks are given in Section 4.³

2 Oligopoly and Regional Advantage

To keep matters simple, we consider an economy with two firms (say 1 and 2) producing each a differentiated variety. Both firms decide first to locate in either of two possible regions (say A and B) and then compete in prices. In order to focus on the pure impact of localization economies, we assume that both regions A and B are characterized by the same market conditions. We also assume that markets are segmented, that is, each firm sets a price specific to the market in which its product is sold. More precisely, in each region the demand functions for firms 1 and 2's variety are generated by a representative consumer who has the following quadratic utility function:

$$U(q_1, q_2) = \alpha(q_1 + q_2) - (\beta/2)(q_1^2 + q_2^2) - \delta q_1 q_2 + q_0 \quad (1)$$

where q_i ($i = 1, 2$) is the quantity of variety i and q_0 the quantity of numéraire she consumes. As usual, we have $\alpha > 0$ and $0 \leq \delta < \beta$. Her budget constraint is $y = p_1 q_1 + p_2 q_2 + q_0$.

Maximizing (1) subject to the budget constraint yields the standard linear inverse demand schedule $p_i = \alpha - \beta q_i - \delta q_j$ in the price domain where quantities are positive. For $\delta \neq \beta$, the demand function for variety i is given by:

$$q_i = a - b p_i + d(p_j - p_i) \quad (2)$$

³Before proceeding, we should like to mention a related paper by Soubeyran and Weber [22] that was brought to our attention recently. Like us these authors allow for Marshallian externalities and imperfect competition. Unlike us, they study a Cournot oligopoly in which there is a single market but any arbitrary given number n of regions in which firms can locate. Among other things, they show the existence of a location equilibrium with n regions when the market demand is linear.

where $a \equiv \alpha/(\beta + \delta)$, $b \equiv 1/(\beta + \delta)$, and $d \equiv \delta/[(\beta - \delta)(\beta + \delta)]$.

The demand system (2) can be interpreted as follows. Parameter d is an inverse measure of the degree of product differentiation between varieties: they are independent when $d = 0$ and perfect substitutes when $d \rightarrow \infty$. In other words, increasing the degree of product differentiation between varieties amounts to decreasing d . Parameter b gives the link between individual and industry demand (total demand becomes inelastic when $b \rightarrow 0$ as in the Hotelling model with firms located at the market endpoints).⁴

In order to export its product, each firm has to incur a constant unit transportation cost from one region to the other; this cost is given by t . The production cost structure of the firms depends on their *proximity* and is described by the following set of assumptions:

- When firms are located in different regions, their marginal cost of production is equal to $c > 0$.
- When firms are located in the same region, they benefit from some positive localization economy. This means that their marginal cost is reduced by a positive amount which is region-specific. More precisely, if both firms locate in region K ($K = A, B$), firm i 's cost is given by $c - \theta_K$.

In other words, we assume that firms experience the same reduction in their marginal cost when they locate together. However, this reduction is

⁴Assuming that all prices are identical and equal to p , we see that the aggregate demand for the differentiated product equals $2(a - bp)$ which is independent of d . Hence (2) has the desirable property that the market size in the industry does not change when the substitutability parameter d varies. More generally, it is possible to decrease (increase) d through a decrease (increase) in the parameter δ 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 $b + d$) than each cross price effect (as measured by d).

likely to depend on the region where they locate because the nature and intensity of non-market interactions between firms vary from one region to the other. Without loss of generality, we assume that the cost reduction is larger in region A than in region B : $\theta_B \leq \theta_A < c$.

We solve the game for its subgame-perfect Nash equilibria by backward induction. We start by solving the second stage of the game where two subgames must be considered according to whether the firms are located together or separately.

2.1 Interregional price competition

(i) Assume that both firms are located in region K . Let p_{iK} and q_{iK} respectively denote the price and quantity of the product sold by firm i in region K . Firm i 's problem consists in choosing the prices p_{iK} (the “home” price) and p_{iL} (the “foreign” price) that maximize its profit function defined as follows:

$$\begin{aligned} \Pi_i = & [p_{iK} - (c - \theta_K)][a - bp_{iK} + d(p_{jK} - p_{iK})] \\ & + [p_{iL} - (c - \theta_K) - t][a - bp_{iL} + d(p_{jL} - p_{iL})] \end{aligned} \quad (3)$$

A similar expression holds for firm j .

It is well known that this game has a unique Nash price equilibrium. Taking the first order conditions and solving for the system of four equations in four unknowns yields the following equilibrium prices:

$$\begin{aligned} p_{iK} = p_{jK} &= \frac{a + (b + d)(c - \theta_K)}{2b + d} \equiv p_K^h \\ p_{iL} = p_{jL} &= \frac{a + (b + d)(c + t - \theta_K)}{2b + d} \equiv p_K^f \end{aligned} \quad (4)$$

(we use the subscript K to refer to the case where both firms are located in region K as well as h and f to denote variables related to the home and foreign markets).

Equilibrium quantities are easily found as:

$$\begin{aligned} q_{iK} &= q_{jK} = \frac{(b+d)[a-b(c-\theta_K)]}{2b+d} \equiv q_K^h \\ q_{iL} &= q_{jL} = \frac{(b+d)[a-b(c+t-\theta_K)]}{2b+d} \equiv q_K^f \end{aligned} \quad (5)$$

Plugging (4) and (5) into (3), we obtain the equilibrium profits when firms are located together in region K as

$$\Pi_K = \frac{b+d}{(2b+d)^2} [[a-b(c-\theta_K)]^2 + [a-b(c+t-\theta_K)]^2] \quad (6)$$

(ii) Suppose now that firm i is located in region K and firm j in region $L \neq K$. Firm i 's profit function is now written as follows:

$$\begin{aligned} \Pi_i &= (p_{iK} - c)[a - bp_{iK} + d(p_{jK} - p_{iK})] \\ &\quad + (p_{iL} - c - t)[a - bp_{iL} + d(p_{jL} - p_{iL})] \end{aligned} \quad (7)$$

One obtains a similar expression for the other firm by substituting j for i , and K for L .

Taking the first order conditions and solving for the corresponding system of four equations yields the following equilibrium prices:

$$\begin{aligned} p_{iK} &= p_{jL} = \frac{a + (b+d)c}{2b+d} + \frac{(b+d)dt}{(2b+d)(2b+3d)} \equiv p_S^h \\ p_{iL} &= p_{jK} = \frac{a + (b+d)c}{2b+d} + \frac{2(b+d)^2t}{(2b+d)(2b+3d)} \equiv p_S^f \end{aligned}$$

(where the subscript S refers to the case where the firms are in separate locations).

Equilibrium quantities are then easily computed as:

$$\begin{aligned} q_{iK} &= q_{jL} = \frac{(b+d)(a-bc)}{2b+d} + \frac{(b+d)^2dt}{(2b+d)(2b+3d)} \equiv q_S^h \\ q_{iL} &= q_{jK} = \frac{(b+d)(a-bc)}{2b+d} - \frac{(b+d)(2b^2+4bd+d^2)t}{(2b+d)(2b+3d)} \equiv q_S^f \end{aligned}$$

Straightforward computations establish that, whether firms are located together or separately, equilibrium quantities and mark-ups are positive (meaning that we have an interior solution and that both firms export their variety) provided that

$$t < t_{trade}^D \equiv \frac{(2b + 3d)(a - bc)}{2b^2 + 4bd + d^2} > 0 \quad (8)$$

In what follows, we will assume that the latter condition is met. In other words, we assume that the transport cost t is low enough to allow firms to export their product whatever their location. Note that condition (8) becomes less stringent as products become more differentiated (i.e. as d decreases).

Collecting previous results, we derive the equilibrium profits when the firms locate separately as

$$\Pi_S = \frac{b + d}{(2b + d)^2} \left\{ \left[a - bc + \frac{(b + d)dt}{2b + 3d} \right]^2 + \left[a - bc - \frac{(2b^2 + 4bd + d^2)t}{2b + 3d} \right]^2 \right\} \quad (9)$$

2.2 Location equilibrium

In the first stage, firms 1 and 2 simultaneously choose their location. Comparing expressions (7) and (9), it is readily verified that $\Pi_K > \Pi_S$ if and only if $\theta_K > \theta_P(t)$, where

$$\theta_P(t) \equiv \frac{t}{2} - \left(\frac{a - bc}{b} \right) + \sqrt{\frac{(a - bc)(a - bc - bt)}{b^2} + \frac{(b + 2d)^2(2b + d)^2}{4b^2(2b + 3d)^2} t^2}$$

When $\theta_A > \theta_B$, three cases may arise.

Proposition 1 *For any triple (θ_A, θ_B, t) such that $\theta_A > \theta_B$, the outcome of the duopoly game takes one of the following forms.*

(i) *If $\theta_A < \theta_P(t)$, then $\Pi_S > \Pi_A > \Pi_B$ and the equilibrium involves dispersion.*

(ii) If $\theta_B < \theta_P(t) < \theta_A$, then $\Pi_A > \Pi_S > \Pi_B$ and the unique equilibrium involves agglomeration in region A.

(iii) If $\theta_P(t) < \theta_B < \theta_A$, then $\Pi_A > \Pi_B > \Pi_S$ and there are two equilibria in which there is agglomeration in region A or in region B (the latter being Pareto-dominated by the former).⁵

In words, we observe that, for a given value of the transportation cost, firms must be compensated for the increased competition that a common location implies by localization economies whose intensity is above some threshold level. Or, to put it differently, *when transport costs are sufficiently low, agglomeration is the market outcome because firms can benefit from production cost reductions by being together without losing much business in the other region.* It is worth noting that the threshold level on θ decreases when the degree of product differentiation rises (that is, when d falls). Indeed, more product differentiation relaxes price competition and, for the same level of localization economies, makes the agglomeration of firms more likely. This occurs because *a high degree of product differentiation allows firms to relax price competition when they are together*, thus making a joint location more attractive. On the other hand, $\theta_P(t)$ rises with the transportation cost because higher trade costs strengthen the benefits of geographical isolation.⁶

⁵If $\theta_A = \theta_B$, case (ii) does not arise.

⁶The inequalities above may be reinterpreted in the context where firm 1 is already located when firm 2 considers entering the market. If $\theta_P(t) < \theta_B < \theta_A$, firm 2 always wants to be with firm 1 regardless of its location. Hence, if firm 1 has chosen to locate in region B, the agglomeration will occur despite the fact that this region is less efficient. This result provides a simple illustration of the phenomenon of *trap* associated with the presence of Marshallian externalities and shows how history matters in the development of a particular region.

2.3 Welfare

Plugging the inverse demands into the utility (1) and using the definitions of b and d , we easily derive the general formulation of the consumers' surplus as

$$C = \frac{1}{2b(b+2d)} [(b+d)(q_1^2 + q_2^2) + 2dq_1q_2] \quad (10)$$

Let us adopt the following notation. Let C_L denote the surplus for the consumers in region L ($L = A, B$) and let the location of the firms be represented by K if both firms are in K ($K = A, B$), or S if they are in separate locations. Similarly, let $C(K) = C_A(K) + C_B(K)$ and $C(S) = C_A(S) + C_B(S)$ respectively denote the global consumer surplus if both firms are in K , or if they are in separate locations. From (10) and from the previous results, we can express the consumer surpluses in the different regions according to the location of the firms as follows:

$$\begin{aligned} C_A(A) &= (1/b)(q_A^h)^2; \quad C_B(A) = (1/b)(q_A^f)^2 \\ C_B(B) &= (1/b)(q_B^h)^2; \quad C_A(B) = (1/b)(q_B^f)^2 \\ C_A(S) &= C_B(S) = \frac{1}{2b(b+2d)} \{(b+d)[(q_S^h)^2 + (q_S^f)^2] + 2dq_S^h q_S^f\} \equiv C_S \end{aligned}$$

We start by considering a *first best* situation in which the planner is able to control both the locations of firms and their prices. Because of marginal cost pricing, the following quantities are sold:

$$\begin{aligned} q_K^h &= a - b(c - \theta_K), & q_K^f &= a - b(c + t - \theta_K), & K &= A, B; \\ q_S^h &= a - bc, & q_S^f &= a - b(c + t). \end{aligned}$$

Since firms earn zero profits, the global welfare is equal to the global consumer surplus. A simple calculation reveals that $C(A) \geq C(B) > C(S)$. That is,

Proposition 2 *If $\theta_A > \theta_B$, then the first best optimum always involves agglomeration in region A. If $\theta_A = \theta_B$, then the first best optimum involves agglomeration in region A or in region B.*

This implies that it is always socially desirable that firms be agglomerated. Yet, when the intensity of the localization economies in region A is not large ($\theta_P(t) > \theta_A$), strategic competition leads to more dispersion than the first best optimum. But the reverse does not hold since *the market never yields excessive agglomeration*.

When transport costs are low enough, the market equilibrium is likely to coincide with the first best location pattern. In this case, the efficiency loss arises only from the discrepancy between prices and marginal costs. Nevertheless, the market may well be at the origin of another efficiency loss in that agglomeration may arise in region B whereas it is socially desirable that firms be located together in region A ($\theta_P(t) < \theta_B < \theta_A$). In the previous section, we have shown that agglomeration in region B is Pareto-dominated and that agglomeration in A would prevail if firms can cooperate. As consequence, cooperation is welfare improving as it induces the right location choice.

Next, we consider a *second best* situation in which the planner is able to control the locations of firms but not their prices and quantities which are determined at the market equilibrium. Plugging equilibrium quantities into the above expressions, we have that the consumers in region K earn the following surpluses according to whether two, one or zero firm(s) are (is) located in their region:

$$\begin{aligned} C_K(K) &= \frac{(b+d)^2}{b(2b+d)^2} [a - b(c - \theta_K)]^2 \\ C_S &= \frac{(b+d)^2}{b(2b+d)^2} \left[(a - bc)(a - bc - bt) + \frac{(b+d)(4b^2 + 8bd + d^2)bt^2}{2(2b+3d)^2} \right] \\ C_K(L) &= \frac{(b+d)^2}{b(2b+d)^2} [a - b(c + t - \theta_L)]^2 \end{aligned}$$

Let us first adopt the point of view of local governments. From our assumption that $\theta_A \geq \theta_B$ (and assuming further that $t > \theta_A - \theta_B$), it is easily seen that the surpluses in case of common location are ranked as follows: $C_A(A) \geq C_B(B) > C_B(A) > C_A(B)$. Furthermore, when we compare

the consumer surpluses when firms are located either together or separately, we can establish the following two results: (i) $C_A(A) \geq C_B(B) > C_S$; (ii) $C_K(L) > C_S$ if and only if $\theta_L > \theta_R(t)$, where

$$\theta_R(t) \equiv t - \left(\frac{a - bc}{b} \right) + \sqrt{\frac{(a - bc)(a - bc - bt)}{b^2} + \frac{(b + d)(4b^2 + 8bd + d^2)}{2b(2b + 3d)^2} t^2}$$

These findings are summarized in the following proposition.

Proposition 3 *At the second best optimum at which firms sell at the equilibrium prices,*

(i) *consumers in region A or B are better off when both firms locate in their region than when only one does so;*

(ii) *consumers in region A or B are better off when no firm locates in their region than when one does if the intensity of localization economies in the other region exceeds some threshold which depends on the transport cost.*

Thus, when localization economies are strong, a local government should attract either the whole industry or no firms because the members of its constituency are worse off when only one firm locates in the corresponding region. Clearly, such an observation does not account for the possible welfare gain associated with the creation of jobs accompanying the installation of a new firm in this region. This also reveals a possible conflict of interest between workers who would find a job in the new company and the whole body of consumers living in the region in question who prefer to benefit from the lower price resulting from the formation of a cluster in the other region.

We now consider the point of view of the federal government. As far as firm's interests are concerned, we already know that total profits are higher when both firms locate in region K than when they separate provided that $\theta_K > \theta_P(t)$. Comparing global surpluses for the consumers for different

locations, we see that $C(K) > C_S$ if and only if $\theta_K > \theta_C(t)$, where

$$\theta_C(t) \equiv \frac{t}{2} - \left(\frac{a-bc}{b} \right) + \sqrt{\frac{(a-bc)(a-bc-bt)}{b^2} + \frac{(b+2d)(2b+d)^2}{4b(2b+3d)^2} t^2}$$

Let the global welfare W be defined as the sum of total profits and total consumer surplus. It can be shown that $W(K) > W(S)$ if and only if $\theta_K > \theta_W(t)$, where

$$\theta_W(t) \equiv \frac{t}{2} - \left(\frac{a-bc}{b} \right) + \sqrt{\frac{(a-bc)(a-bc-bt)}{b^2} + \frac{(3b+5d)(b+2d)(2b+d)^2}{4b(3b+d)(2b+3d)^2} t^2}$$

Some straightforward computations reveal that $\theta_P(t) > \theta_W(t) > \theta_C(t)$ for all t smaller than t_{trade}^D . It then appears that the second best outcome may involve agglomeration while the market selects dispersion but the reverse is never true. In addition, the above inequalities also mean that *the interests of the various economic groups may vastly diverge in the choice process of a location pattern for firms*. For instance, if for a given value of the transportation cost t we have that $\theta_W(t) < \theta_A < \theta_P(t)$, then firms choose separate locations while the federal government would prefer the firms to be agglomerated in region A at the second-best optimum. By contrast, when $\theta_C(t) < \theta_A < \theta_W(t)$, both the federal government and firms prefer separate locations but consumers as a whole are better off when agglomeration occurs in region A .

More conflict might even appear if the interests of consumers in each region are taken into account. It is indeed possible to have situations where $\theta_W(t) < \theta_A < \theta_R(t) < \theta_P(t)$. Then, firms choose separate location, in accordance with the interests of consumers in region B , but not with the interest of the consumers of region A and of the federal government (that would prefer agglomeration in A).

3 The Formation of Clusters with a Large Group of Firms

In this section, we consider an economy with a continuum $[0, 1]$ of firms producing each a differentiated variety. The representative consumer's utility function is now as follows (see the appendix for some details):

$$U(q_0; q(i), i \in [0, 1]) = \alpha \int_0^1 q(i) di - \frac{\beta - \delta}{2} \int_0^1 q(i)^2 di - \frac{\delta}{2} \int_0^1 \int_0^1 q(i) q(j) di dj + q_0 \quad (11)$$

where $q(i)$ is the quantity of variety $i \in [0, 1]$ and q_0 the quantity of the numéraire, while the parameters in (11) are such that $\alpha > 0$ and $\beta > \delta > 0$. In this expression, α is a measure of the size of the market since it expresses the intensity of preferences for the differentiated product with respect to the numéraire, whereas $\beta > \delta$ means that the representative consumer is biased toward a dispersed consumption of varieties, thus reflecting a love for variety. the quadratic utility function exhibits a preference for variety. Suppose, indeed, that the representative consumer consumes a total of $Q \equiv \int_0^1 q(i) di$ of the differentiated product, which is uniform on $[0, x]$ and zero on $(x, 1]$. Then, the density on $[0, x]$ is Q/x . Equation (11) evaluated for this consumption pattern is:

$$\begin{aligned} U &= \alpha \int_0^x \frac{Q}{x} di - \frac{\beta - \delta}{2} \int_0^x \left(\frac{Q}{x} \right)^2 di - \frac{\delta}{2} \int_0^x \int_0^x \left(\frac{Q}{x} \right)^2 di dj + q_0 \\ &= \alpha Q - \frac{\beta - \delta}{2x} Q^2 - \frac{\delta}{2} Q^2 + q_0 \end{aligned}$$

This expression is increasing in x since $\beta > \delta$ and, hence, is maximized at $x = 1$ where variety consumption is maximal. Finally, for a given value of β , the parameter δ expresses the substitutability between varieties: the higher δ , the closer substitutes the varieties.

The consumer is endowed with $\bar{q}_0 > 0$ units of the numéraire. Her budget

constraint can then be written as follows:

$$\int_0^1 p(i)q(i)di + q_0 = \bar{q}_0$$

where $p(i)$ is the price of variety i and q_0 her consumption of the numéraire. The initial endowment $\bar{q}_0$ is supposed to be large enough for the optimal consumption of the numéraire to be strictly positive at the market outcome. Solving the budget constraint for the numéraire consumption, plugging the corresponding expression into (11) and solving the first order conditions with respect to $q(i)$ yields

$$\alpha - (\beta - \delta)q(i) - \delta \int_0^1 q(j)dj = p(i), \quad i \in [0, 1]$$

Since $\beta > \delta$, the demand function for variety $i \in [0, 1]$ is:⁷

$$q(i) = a - bp(i) + d \int_0^1 [p(j) - p(i)]dj \quad (12)$$

where $a \equiv \alpha/\beta$, $b \equiv 1/\beta$ and $d \equiv \delta/\beta(\beta - \delta)$.

There are two regions A and B . When there are N_K firms in region K , firm i is able to produce the variety i at marginal cost $c_K(N_K)$. Of course, we have $N_A + N_B = 1$. Let t be the unit transport cost between the two regions. It is assumed that any firm finds it profitable to export. In order to ensure that this condition is met at any symmetric price equilibrium, we set $p(i) = p(j) = \max \{c_A, c_B\}$ in (12) and obtain

$$t < \frac{a}{b} - \max \{c_A, c_B\} \quad (13)$$

3.1 The equilibrium pricing strategy of a representative firm

We study here the process of competition between firms for a given spatial distribution (N_A, N_B) of firms. Since we have a continuum of firms, each

⁷Compare (2) and (12).

one is negligible in the sense that its action has no impact on the market. Hence, when choosing its prices, a firm in A accurately neglects the impact of its decision over the regional price indices. In addition, because firms sell differentiated varieties, each one has some monopoly power in that it faces a demand function with finite elasticity. All of this is in accordance with Chamberlin's large group competition where the effect of a price change by one firm has a significant impact on its own demand but only a negligible impact on competitors' demands. However, in order to determine its own equilibrium price, a firm must account for the distribution of all firms' prices through some aggregate statistics, given here by the price index. 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.⁸

Again we assume that firms compete in segmented markets. In the sequel, we focus on region A . Things pertaining to region B can be derived by symmetry. We suppose that the parameters are such that the equilibrium prices exceed costs and mark-ups are positive (meaning that we have an interior solution and that exportation occurs for all firms). A sufficient condition for this to hold will be given below.

The demand in region A for variety i is given by

$$q_A(i) = a - bp(i) + d \int_0^1 [p(j) - p(i)]dj$$

Assume that variety i is produced in A . The corresponding firm sells on both markets, i.e., quantity $q_{AA}(i)$ at price $p_{AA}(i)$ in market A , and quantity $q_{AB}(i)$ at price $p_{AB}(i)$ in market B . Thus, $p_{AA}(i)$ is the price in region A of variety i produced locally and $p_{AB}(i)$ the price of the same variety exported

⁸See Ottaviano and Thisse (1999) for more details

from A to B . We adopt the following notation:

$$P_{AA} \equiv \int_{i \in A} p_{AA}(i) di ; P_{AB} \equiv \int_{i \in A} p_{AB}(i) di$$

$$P_{BB} \equiv \int_{j \in B} p_{BB}(j) dj ; P_{BA} \equiv \int_{j \in B} p_{BA}(j) dj$$

Demands for firm i are then given by:

$$q_{AA}(i) = a - (b + d)p_{AA}(i) + d(P_{AA} + P_{BA})$$

and

$$q_{AB}(i) = a - (b + d)p_{AB}(i) + d(P_{AB} + P_{BB})$$

Firm i in A maximizes its profits defined by:

$$\Pi_A(i) = [p_{AA}(i) - c_A]q_{AA}(i) + [p_{AB}(i) - c_A - t]q_{AB}(i) \quad (14)$$

We first differentiate (14) with respect to prices $p_{AA}(i)$ and $p_{AB}(i)$ for a representative firm i to obtain the first order conditions. Integrating the corresponding expressions across firms i located in A , we obtain the following equations:

$$[2(b + d) - dN_A]P_{AA} - dN_AP_{BA} = N_A[a + (b + d)c_A] \quad (15)$$

$$[2(b + d) - dN_A]P_{AB} - dN_AP_{BB} = N_A[a + (b + d)(c_A + t)] \quad (16)$$

Through a similar process, we obtain two more equations for the firms located in B :

$$[2(b + d) - dN_B]P_{BB} - dN_BP_{AB} = N_B[a + (b + d)c_B] \quad (17)$$

$$[2(b + d) - dN_B]P_{BA} - dN_BP_{AA} = N_B[a + (b + d)(c_B + t)] \quad (18)$$

Since profit functions are concave in own price and varieties are symmetric, solving the system of equations (15)-(18) yields the equilibrium prices:

$$p_{AA} = \frac{c_A}{2} + \frac{2a + d[N_Ac_A + N_B(c_B + t)]}{2(2b + d)}$$

$$\begin{aligned}
p_{AB} &= \frac{c_A + t}{2} + \frac{2a + d[N_A(c_A + t) + N_B c_B]}{2(2b + d)} \\
p_{BB} &= \frac{c_B}{2} + \frac{2a + d[N_A(c_A + t) + N_B c_B]}{2(2b + d)} \\
p_{BA} &= \frac{c_B + t}{2} + \frac{2a + d[N_A c_A + N_B(c_B + t)]}{2(2b + d)}
\end{aligned}$$

As expected, the equilibrium prices depend on the distribution of firms between the two regions. They rise when the size of the local market, evaluated by a , gets larger or when the degree of product differentiation, inversely measured by d , increases provided that (13) holds. All these results agree with what is known in industrial organization and spatial pricing theory. By inspection, it is also readily verified that both local prices, p_{AA} and p_{BB} , increase with t because the local firms in A (B) are more protected against distant competitors, whereas the export prices, $p_{AB} - t$ and $p_{BA} - t$, decrease because it becomes more difficult for these firms to penetrate the distant market. Finally, both the prices charged by local and distant firms fall when the number of local firms in, say, region A increases, while holding the total number of firms constant, if and only if $c_A < c_B + t$. This occurs because the lower cost prevailing in A intensifies local price competition.

Using the first order conditions, it is easy to establish the following relationships between equilibrium prices and quantities: $q_{AA} = (b + d)(p_{AA} - c_A)$ and $q_{AB} = (b + d)(p_{AB} - c_A - t)$. The equilibrium profits of any firm located in region A are thus

$$\begin{aligned}
\Pi_A(N_A, N_B) &= (p_{AA} - c_A)q_{AA} + (p_{AB} - c_A - t)q_{AB} \\
&= (b + d)[(p_{AA} - c_A)^2 + (p_{AB} - c_A - t)^2] \\
&= \frac{b + d}{2(2b + d)^2} \{ [2(a - bc_A) - dN_B(c_A - c_B) - bt]^2 \\
&\quad + (b + dN_B)^2 t^2 \}
\end{aligned}$$

Similarly, the profits of any firm with located in region B are

$$\begin{aligned}\Pi_B(N_A, N_B) = & \frac{b+d}{2(2b+d)^2} \{ [2(a-bc_B) - dN_A(c_B - c_A) - bt]^2 \\ & + (b+dN_A)^2 t^2 \}\end{aligned}$$

In the remainder of the paper, it is assumed for simplicity that the localization economies obey the same law in each region:

$$c_K(N_K) = c - \theta N_K$$

where $0 < \theta < c$. We are then able to state the conditions under which the equilibrium prices and quantities are positive (meaning that we have an interior solution and that exportation occurs for all firms). It is readily checked that a sufficient condition is that $q_{AB} > 0$ in the limiting case where $N_A = 0$ (or that $q_{BA} > 0$ in the limiting case where $N_A = 1$), which translates as:

$$t < t_{trade}^C \equiv \frac{2(a-bc) - d\theta}{2b+d} \quad (19)$$

whose right hand side is positive by (13).

3.2 Location equilibrium

We can take advantage of the symmetry of the problem by setting $\Delta N \equiv N_A - N_B$. Thus, $N_A = (1/2)(1 + \Delta N)$, $N_B = (1/2)(1 - \Delta N)$, $c_A(N_A) + c_B(N_B) = 2c - \theta$, and $c_A(N_A) - c_B(N_B) = -\theta\Delta N$. Consequently, the equilibrium profits can be rewritten as follows (where D stands as a shortcut for $2b + d$):

$$\begin{aligned}\Pi_A(\Delta N) = & \frac{b+d}{8D^2} \{ [4a - 2b(2c + t - \theta) - d\theta(\Delta N)^2 + D\theta\Delta N]^2 \\ & + (D - d\Delta N)^2 t^2 \}\end{aligned} \quad (20)$$

$$\Pi_B(\Delta N) = \frac{b+d}{8D^2} \{ [4a - 2b(2c + t - \theta) - d\theta(\Delta N)^2 - D\theta\Delta N]^2 + (D + d\Delta N)^2 t^2 \} \quad (21)$$

Accordingly, the difference $\Delta\Pi(\Delta N)$ between the profits earned in each region is given by

$$\Delta\Pi(\Delta N) \equiv \Pi_A - \Pi_B = \frac{b+d}{2D} \Delta N \{ [4a - 2b(2c + t - \theta) - d\theta(\Delta N)^2] \theta - dt^2 \}$$

which can be rewritten as the following cubic function of ΔN :

$$\Delta\Pi(\Delta N) = -\frac{d\theta^2(b+d)}{2D} \Delta N [(\Delta N)^2 - X] \quad (22)$$

where

$$X \equiv \frac{[4a - 2b(2c + t - \theta)] \theta - dt^2}{d\theta^2} \quad (23)$$

We now ask whether for a given spatial distribution of firms, (N_A, N_B) , there is an incentive for a firm to relocate. A *location equilibrium* is such that when no locational deviation by a firm is profitable. This arises at an interior point $N_A \in (0, 1)$ when $\Delta\Pi(N_A) = 0$, or at $N_A = 0$ when $\Delta\Pi(0) \leq 0$, or at $N_A = 1$ when $\Delta\Pi(N) \geq 0$. In the first case, we have either a two identical clusters or two asymmetric clusters; in the last two cases, we have a single cluster. Given (22), the fully dispersed configuration ($\Delta N = 0$) is always an equilibrium.

If firms observe that a region offers a higher profits than the other, they want to move to that location. In other words, for any interior solution the driving force in the *profit differential* between A and B :

$$\dot{N}_A \equiv \frac{dN_A}{d\tau} = N_A \Delta\Pi(\Delta N) N_B$$

when τ is time. Since $\dot{N}_A = 0$ implies $\Delta\Pi(\Delta N) = 0$, or $N_A = 0$, or $N_B = 0$, any location equilibrium is such that $\dot{N}_A = 0$. When $0 < N_A < 1$, $\Delta\Pi$ positive implies some firms will move from B to A ; if it is negative, some will

go in opposite direction. An equilibrium is *stable* if, for any marginal deviation in the firm distribution from the equilibrium, the equation of motion above brings the firm distribution back to the original one. Therefore, a fully agglomerated configuration is always stable when it turns out to be an equilibrium, while an interior equilibrium is stable if and only if the slope of $\Delta\Pi(\Delta N)$ is negative in a neighborhood of this equilibrium.

Several kinds of equilibria may arise in this setting. Either all firms agglomerate in one region (corner solution) or they distribute themselves between the two regions (interior solution) in a way that equalizes profits. In the latter case, firms can spread evenly ($\Delta N = 0$) or unevenly across regions. The stable equilibria are now fully described.

Proposition 4 *The two-region economy has a single location equilibrium.*

This one involves:

- (i) *identical clusters ($\Delta N = 0$) if and only if $X \leq 0$;*
- (ii) *asymmetric clusters ($\Delta N = \pm\sqrt{X}$) if and only if $0 < X \leq 1$;*
- (iii) *a single cluster ($\Delta N = 1$) if and only if $1 < X$.*

Proof: When $X < 0$, the equation $\Delta\Pi = 0$ has a single real solution $\Delta N = 0$ which is therefore stable. When $X > 0$, there are three real solutions $\Delta N = 0$, and $\Delta N = \pm\sqrt{X}$. The nonzero solutions are the only stable equilibria if and only if $d(\Delta\Pi)/d(\Delta N)$ evaluated at $\Delta N = \pm\sqrt{X}$ is negative. This is so when

$$\frac{d(\Delta\Pi)}{d(\Delta N)} = -\frac{d\theta^2(b+d)}{2D}[3(\Delta N)^2 - X] \leq 0$$

which holds for $\Delta N = \pm\sqrt{X}$ if $0 < X \leq 1$. When $X > 1$, there is no asymmetric interior equilibrium and the only stable equilibria are such that $\Delta N = 1$ since $\Delta\Pi(N) \geq 0$ and $\Delta\Pi(0) \geq 0$. Q.E.D.

The stable equilibria are depicted in the $(X, \Delta N)$ plane of Figure 1.

Figure 1: The stable equilibria

Hence, *despite the symmetry of the setting there exist stable equilibria in which regions collect different numbers of firms.* In this case, the number of firms in region A is

$$N_A^* = \frac{1 \pm \sqrt{X}}{2}$$

Of course, the region which ends up with the larger number of firms is the one which has the larger initial share of firms, however small is the difference. This shows again that history matters for the geographical distribution of production. This occurs in the t -region defined by the interval $]t_1, t_2[$, t_1 being the solution to $X = 0$ and t_2 the solution to $X = 1$. Stated differently, the existence of localization economies may lead to the emergence of a polarized space, especially when transport costs are low.

Furthermore, as the size a of the market rises, the degree of asymmetry between the two clusters grows. This occurs because the relative impact of the localization economies rises with the market size. Consequently, *economic growth*, as measured by a expanding market, *should yield a more agglomerated pattern of production.* It is worth pointing out that such an increase in the agglomeration of firms arises although the spatial distribution of demand remains unchanged.⁹ However, this effect is damped by an increase in the extent of localization economies but amplified by an increase in product differentiation.

The impact of transport cost (t), product differentiation (d) and localization economies (θ) on the location equilibrium can be analyzed through the term X which is an increasing function of ΔN . When transport costs are negligible ($t \simeq 0$), firms agglomerate because both markets are almost

⁹A similar effect appears in Martin and Ottaviano (1996) in the context of a regional growth model in which technological spillovers are localized.

bunched into a single one ($p_{AA} \simeq p_{AB}$ and $p_{BB} \simeq p_{BA}$) so that geographical separation no longer yields monopoly rents. Accordingly, firms want to agglomerate in order to be able to enjoy the highest possible level of localization economies. To show this, it is sufficient to check that $X > 1$ when $t = 0$. Some simple manipulations show that this is equivalent to $(2t_{trade}^C + \theta)\theta > 0$ which holds since $t_{trade}^C > 0$.

Firms also want to agglomerate when products are very differentiated ($d \simeq 0$). Indeed, when $d = 0$ there is no need for firms to relax price competition by selecting distinct locations. In addition, since the product demand is identical in each region, no region provides any locational advantage with respect to transportation costs. Localization economies are thus the only active force, whence agglomeration arises. Formally, we observe that $X \rightarrow +\infty$ when $d \rightarrow 0$ since $2a - b(2c - t - \theta)$ is positive by (19).

Finally, when localization economies are weak ($\theta \simeq 0$), firms want to separate as much as possible to exploit the geographical isolation of each market. Formally, we have:

$$\lim_{\theta \rightarrow 0} X = \lim_{\theta \rightarrow 0} -t^2/\theta^2 = -\infty$$

so that two identical clusters are formed.

Comparative statics can be performed to determine the impact of the parameters of the economy on the relative size of clusters in the case of an asymmetric equilibrium ($0 < X < 1$). In particular, we have:

$$\frac{\partial X}{\partial t} = -\frac{2(b\theta + dt)}{d\theta^2} < 0 \quad (24)$$

$$\frac{\partial X}{\partial d} = -\frac{X}{d} - \frac{t^2}{d\theta^2} < 0 \quad (25)$$

$$\frac{\partial X}{\partial \theta} = -\frac{X}{\theta} + 2\frac{b}{d\theta} + \frac{t^2}{\theta^3} \quad (26)$$

Equation (24) shows that *a decrease in transport cost leads to more asymmetry between clusters*. This reveals that very low transportation costs are

likely to drive the economy towards more agglomeration in one region at the expense of the other. However, the economy moves smoothly from the fully dispersed pattern to the fully agglomerated pattern as t decreases from high to low values, a result that vastly differs from what it is observed in economic geography models [9, 12]. Similarly, as shown by (25), *more product differentiation leads to more agglomeration of firms within the large region*. This is now a standard result in many spatial models.

Equation (26) shows that an increase in the intensity of localization economies strengthens the tendency toward agglomeration provided that θ is small enough, that is

$$\theta < \frac{dt^2}{2(a - bc) - bt} \quad (27)$$

By contrast, when (27) does not hold we see that stronger localization economies generate more dispersion. This surprising result may be explained as follows. When production costs are not too high, firms in the small region price in the inelastic part of their demands, while firms in the large region price in the elastic part. By reducing production costs, an increase in θ thus intensifies competition much more in the large region than in the small one, leading some firms to move from the large region to the small one.

3.3 Welfare

It is readily verified that the consumer surplus is given by the following expression:

$$\begin{aligned} C &= \alpha \int_0^1 q(i) di - \frac{1}{2}(\beta - \delta) \int_0^1 [q(i)]^2 di - \frac{1}{2}\delta \left(\int_0^1 q(i) di \right)^2 - \int_0^1 p(i) q(i) di \\ &= \frac{1}{2}(\beta - \delta) \int_0^1 [q(i)]^2 di + \frac{1}{2}\delta \left(\int_0^1 q(i) di \right)^2 \\ &= \frac{1}{2b(b + d)} \left[b \int_0^1 [q(i)]^2 di + d \left(\int_0^1 q(i) di \right)^2 \right] \end{aligned}$$

3.3.1 First Best

To begin with, we assume that the planner is able to impose prices equal to marginal costs as well as to choose firms' locations. Since firms earn zero profits, the relevant welfare measure is the sum of regional consumer surpluses $C_G = L = A, B \sum C_L$ where

$$C_L = \frac{1}{2(b+d)} [(q_{LL}^*)^2 N_L + (q_{KL}^*)^2 N_K] + \frac{d}{2b(b+d)} [(q_{LL})^2 N_L^2 + (q_{KL})^2 N_K^2 + 2(q_{LL} N_L)(q_{KL} N_K)] \quad (28)$$

where $L = A, B$ denotes the region in which consumption occurs while $K = A, B$ and $K \neq L$ stands for the other region. After having replaced prices and quantities by their first best values, we obtain

$$C_G = -\frac{A_1}{4}(\Delta N)^4 + Y \frac{A_1}{2}(\Delta N)^2 + A_0$$

in which

$$A_1 \equiv d\theta^2 > 0$$

and A_0 are constant and

$$Y \equiv \frac{(3b+d)\theta^2 + 4\theta(a-bc-bt/2) - dt^2}{2d\theta^2} \quad (29)$$

The first best locational pattern is then obtained by maximizing C_G with respect to ΔN .

Proposition 5 *The first best allocation involves:*

- (i) *identical clusters if and only if $Y \leq 0$;*
- (ii) *asymmetric clusters if and only if $0 < Y < 1$;*
- (iii) *a single cluster if and only if $1 \leq Y$.*

Proof. We must find $\Delta N \in [-1, 1]$ that maximizes C_G . When the maximum is interior, we must have:

$$\frac{\partial C_G}{\partial(\Delta N)} = -A_1 \Delta N [(\Delta N)^2 - Y] = 0 \quad (30)$$

and

$$\frac{\partial^2 C_G}{\partial(\Delta N)^2} = -A_1 [3(\Delta N)^2 - Y] < 0$$

When $Y \leq 0$, there is a unique maximizer given by $\Delta N = 0$ since the second order condition is always satisfied. When $0 < Y < 1$, C_G reaches a minimum at $\Delta N = 0$ and is maximized at $\Delta N = \pm\sqrt{Y} < 1$. Finally, when $Y \geq 1$, the welfare function C_G is maximized if $\Delta N = \pm 1$. Q.E.D.

In the case of asymmetric clusters, the socially optimal distribution of firms (N_A^o, N_B^o) is given by

$$N_A^o = \frac{1 \pm \sqrt{Y}}{2}$$

This proposition implies that *the first best solution displays a pattern similar to that arising when firms are free to choose prices and locations at the market equilibrium*. Using (23) and (29), we obtain the following relation between X and Y :

$$Y = \frac{b+d}{2d} + \frac{X}{2} \quad (31)$$

Given (24) and (25), both X and Y then move in the same direction when t or d changes so that Y is a decreasing function of t and d :

$$\frac{\partial Y}{\partial t} = \frac{1}{2} \frac{\partial X}{\partial t} < 0 \quad (32)$$

$$\frac{\partial Y}{\partial d} = -\frac{b}{2d^2} + \frac{1}{2} \frac{\partial X}{\partial d} < 0 \quad (33)$$

In words, *lower transport costs and/or more product differentiation yield more asymmetry between clusters in the first best*. Using (31), it is readily verified that the impact of a or θ on Y is similar to the impact on X as discussed in 3.2.

Furthermore, it follows from (31) that $Y > 0$ when $X = 0$ and that $Y = X$ for a value of X that exceeds one. Consequently, we have the following result.

Proposition 6 *The first best optimum never involves less agglomeration than the market equilibrium.*

Thus, as in the duopoly case in which there is never too much agglomeration from the social point of view, we observe that, with a large group of firms, *the large cluster never involves too many firms in equilibrium*. In particular, the planner sets up more asymmetric clusters than what arises at the market solution. This requires some explanations. At the optimum, prices are set at the lowest level and locations are chosen as to maximize the benefits of agglomeration and to minimize total transport costs. By contrast, at the market equilibrium, firms take advantage of spatial separation in order to relax price competition, thus making higher profits. These two effects combine to generate the foregoing discrepancy between the market and optimal solutions.

Unless dispersion corresponds to both the equilibrium and the optimum, the difference between regional surpluses generates *a conflict between regions about firms' locations*. Indeed, the region with the larger cluster benefits from larger localization economies, and thus lower prices, as well as from lower transportation costs on its imports (through less varieties and smaller quantities). This occurs because the planner focuses only upon global efficiency and not on interregional equity. This makes sense when lump sum transfers compensating the consumers of the less industrialized region are available. However, when such redistributive instruments are not available, a trade-off between global efficiency and interregional equity arises.

3.3.2 Second Best

Consider now a situation in which the planner is able to control firms' locations but not their prices, which are determined at the market equilibrium. Assume again that her purpose is to maximize the global surplus which is now given by $W_G = C_A + C_B + P_A + P_B$. The consumer surplus C_K in region $K = A, B$ is computed by plugging the market equilibrium quantities into expression (28). The producer surplus P_K in region $K = A, B$ is given by

the sum of profits made in K and L by the firms located in K :

$$P_K = \frac{N_K}{b+d} [(q_{KK}^*)^2 + (q_{KL}^*)^2]$$

To determine the second best, it is sufficient to maximize W_G . The first order condition yields a cubic function with properties similar than those obtained in the first best when Y is replaced by Z as given by:

$$\begin{aligned} Z &\equiv \frac{4(3b+d)\theta[4a-2b(2c+t-\theta)] + 4b(3b+d)\theta^2 - d(8b+3d)(t^2-\theta^2)}{2d\theta^2(8b+3d)} \\ &= \frac{3(2b+d)^2 + d(4b+d)(t/\theta)^2}{2d(8b+3d)} + \frac{2(3b+d)}{8b+3d}X \end{aligned}$$

Consequently, the second best allocation involves (i) identical clusters if and only if $Z \leq 0$; (ii) asymmetric clusters if and only if $0 < Z < 1$; (iii) a single cluster if and only if $1 \leq Z$. The comparison between the second best and the market outcomes is very similar to that made above provided that X and Z move in the same direction. In particular, it is readily verified that Proposition 6 still holds, that is $Z > X$, since $Z > 0$ when $X = 0$ while $X = Z$ for a value of X exceeding 1.

A related question we ask is whether the second best induces more agglomeration than the first best. This turns out to be true because

$$Z - Y = \frac{(4b+d)[4(a-bc) - 2bt + 3\theta b]}{2d\theta(8b+3d)}$$

which is positive since $4(a-bc) > 2bt$ by (13). Therefore, *the planner's best response to the loss of control on prices is to take advantage of the localization economies by having more firms in the large cluster*. This surprising result is to be understood as follows. The planner's objective is now to dampen too high prices through the control of locations, especially when varieties are very differentiated. In order to achieve this goal, she chooses to expand the cluster in the larger region because, in so doing, both price competition is

intensified and localization economies are made stronger. This confirms what was observed in spatial competition theory, namely that price competition is a strong dispersion force (Fujita and Thisse, 1996).

4 Concluding Remarks

We have shown how the impact of localization economies rises as transport costs between regions fall. This suggests that agglomeration is more likely to occur in the global economy because firms are able to enjoy a higher level of localization economies while still being able to sell a substantial fraction of their output on distant markets. Among other things, our analysis sheds light on the conditions under which a firm may benefit from the presence of local competitors despite the fact that agglomeration makes price competition fiercer.

We have also uncovered a market size effect. When the desirability of the product rises, more firms tend to locate within the same cluster whose relative size increases at the expense of the other. This occurs because the relative impact of the localization economies rises with the market size. Consequently, economic growth, expressed here by an expanding market, could well lead to more geographical concentration in larger clusters.¹⁰ It is worth pointing out that such an increase in the agglomeration of firms arises although the spatial distribution of demand remains unchanged. However, one expects the total number of firms, which is here normalized to one, to increase as a result of growth. In this case, the impact on agglomeration is ambiguous and depends on the many features of the economy.

Unlike what many regional analysts and planners would argue, the optimal configuration involves a more imbalanced distribution of firms than

¹⁰This agrees with the analysis of the interplay between agglomeration and growth developed by Martin and Ottaviano (1996).

the market outcome. This holds both in the duopoly and large group cases. If localization economies tend to become more and more important in advanced economic sectors, as suggested by the growing role played by knowledge spillovers in research and development, the observed regional imbalances in the geographical distribution of high tech activities may not correspond to a wasteful allocation of resources. On the contrary, the size of the existing clusters could well be too small. In the same vein, we have shown that the second best is even more agglomerated than the first best because geographical concentration reduces the wasteful effects of imperfect competition, thus making the case of agglomeration in advanced and specialized activities even stronger. It is our belief that these welfare comparisons are fairly robust because the forces acting behind them seem to be general. If so, one should expect the resulting imbalance in the geographical distribution of differentiated activities to generate a social trade-off between efficiency and spatial equity, especially in countries characterized by a low mobility of their labor force.

This needs qualification, however, because we have focussed upon a single issue in this paper: the connection between regional clusters and the global economy. And, indeed, our approach suffers from several drawbacks. The most severe is probably the absence of non-market institutions that seem to play a central role in the working of real world clusters. No better: we had to restrict ourselves to two regions because of technical reasons. More work is called for here to cope with the case of several regions. However, the results obtained by Soubeyran and Weber [22] suggest that the prospects are good enough.

References

- [1] J.S. Chipman, External economies of scale and competitive equilibrium, *Quarterly Journal of Economics*, **85**, 347–385 (1970).
- [2] M.P. Feldman and D.B. Audretsch, Innovation in cities: science-based diversity, specialization and localized competition, *European Economic Review*, **43**, 409-429 (1999).
- [3] M. Fujita, P. Krugman, and A.J. Venables “The Spatial Economy. Cities, Regions and International Trade,” Cambridge (Mass.), MIT Press (1999).
- [4] M. Fujita and J.-F. Thisse, Economics of agglomeration, *Journal of the Japanese and International Economies*, **10**, 339-378 (1996).
- [5] E. Glaeser, H.D. Kallal, J.A. Scheinkman, and A. Shleifer, Growth in cities, *Journal of Political Economy*, **100**, 1126-1152 (1992).
- [6] G. H. Hanson, Localization economies, vertical organization, and trade, *American Economic Review*, **86**, 1266-1278 (1996).
- [7] J.V. Henderson, “Urban Development. Theory, Fact and Illusion,” Oxford, Oxford University Press (1988).
- [8] J. V.Henderson, Kuncoro A., and M. Turner, Industrial development in cities, *Journal of Political Economy*, **103**, 1066-1090 (1995).
- [9] P.R. Krugman, Increasing returns and economic geography, *Journal of Political Economy*, **99**, 483-499 (1991).
- [10] Ph. Martin and G.I.P. Ottaviano, Growth and agglomeration, CEPR Discussion Paper N°1529 (1996).

- [11] H. Ogawa and M. Fujita, Equilibrium land use patterns in a non-monocentric city, *Journal of Regional Science*, **20**, 455-475 (1980).
- [12] G.I.P. Ottaviano and J.-F. Thisse, Agglomeration and trade revisited, CEPR Discussion Paper N°1903 (1998).
- [13] G.I.P. Ottaviano and J.-F. Thisse, Monopolistic competition, multi-product firms and optimum product diversity, CORE Discussion Paper N°9903 (1999).
- [14] M.E. Porter, “The Competitive Advantage of Nations,” New York, Free Press (1990).
- [15] M.E. Porter, Clusters and the new economics of competition, *Harvard Business Review*, December, 77-90 (1998).
- [16] M.E. Porter, “On Competition,” A Harvard Business Review Book (1998).
- [17] E.C. Prescott, Nedded: a theory of total factor productivity, *International Economic Review*, **39**, 525-51 (1998).
- [18] F. Pyke, G. Becattini, and W. Sengenberger, “Industrial Districts and Inter-firm Cooperation in Italy,” Geneva, International Institute for Labour Studies, (1990).
- [19] A. Saxenian “Regional Advantage: Culture and Competition in Silicon Valley and Route 128,” Cambridge (Mass.), Harvard University Press (1994).
- [20] P.J. Smith, Do knowledge spillovers contribute to U.S. state output and growth? *Journal of Urban Economics*, **45**, 331-353 (1999).

- [21] A. Soubeyran and J.-F. Thisse, Learning-by-doing and the development of industrial districts, *Journal of Urban Economics*, **45**, 156-176 (1999).
- [22] A. Soubeyran and S. Weber, Localization of activities: an endogenous geographical choice, GREQAM Document de travail 99A02 (1999).
- [23] J. Tirole, "The Theory of Industrial Organization," Cambridge (Mass.), MIT Press (1988).

Appendix

In the case of two varieties, we know that the quadratic utility is given by (1):

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

In the case of $n > 2$ varieties, (1) is extended as follows:

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

Letting $n \rightarrow \infty$ and $q_i \rightarrow 0$, we then obtain (11) in which the unit interval stands for the set of varieties.