8

Industrial Agglomeration under Marshallian Externalities

8.1 INTRODUCTION

The work of Marshall has been very influential in regional and urban studies and has led to the concept of *Marshallian externalities*, which aim at accounting for the benefits associated with cluster formation. As seen in Chapter 1, these benefits arise because of

1. the formation of a highly specialized labor force and the development of new ideas, both based on the accumulation of human capital and face-to-face communications (the latter has been analyzed in Chapter 6);
2. the availability of specialized input services (see Chapter 4); and
3. the existence of modern infrastructure (see Chapter 5).[1]

From the point of view of this chapter, the main distinctive feature of these externalities is that they affect only the agents belonging to the same geographical area. They do not spread over other regions or, more precisely, their impact on distant regions may be considered negligible.[2]

The now standard classification of Marshallian externalities is attributed to Hoover (1936, chap. 6): (1) the *localization economies*, which are defined as the benefits generated by the proximity of firms producing similar goods; and (2) the *urbanization economies*, which are defined by all the advantages associated with the overall level of activity prevailing in a particular area. These external effects have been studied extensively from the empirical standpoint (see, e.g., Henderson 1988, chap. 5, and the references therein).[3] For example, Hanson (1996, 1266) accurately points out that "the fact that New York City remains a major apparel producer is perhaps the most persuasive evidence one can find of localization economies."

The evidence shows that both types of external effects are at the origin of several specialized and prosperous areas. According to Jacobs (1969), urbanization economies are predominant, whereas Porter (1998, chap. 7) has argued that the main reason for the success of industrial clusters in the global economy

is the presence of strong localization economies. The conclusion to the debate over localization versus urbanization economies is an empirical one and, for the time being, it seems hard to foresee what the outcome will be. However, one may conjecture that the results are likely to depend on the nature of the industry as well as on the size of the area in question.

To a large extent, the idea of localization economies also explains the growth and success of *industrial districts*, that is, regions that accommodate many small firms producing similar goods and that benefit from the localized accumulation of skills associated with workers residing in these places (Becattini 1990). Some industrial districts are engaged in high-tech activities (Saxenian 1994), but others are involved in more traditional, labor-intensive activities, many of which can be found in the "Third Italy" (Pyke et al. 1990, chaps. 4 and 5): Sassuolo specializes in ceramic tiles, Prato is known for textiles, shoes are made in Montegranaro, and wooden furniture in Nogara.

In any case, the combination of various factors turns out to be essential for the localized accumulation of various types of knowledge within a region.[4] For example, Feldman and Florida (1994, 226) concluded a detailed study of the geography of innovation in the United States by noticing that

> innovation is no longer the province of the inventor, the risk-taking entrepreneur, the insightful venture capitalist, or the large resource-rich corporation. Innovation instead has its sources in a broader social and spatial structure – a landscape of agglomerated and synergistic social and economic institutions welded into a technological infrastructure for innovation.

The foregoing discussion as well as early work in urban economics (Henderson 1988, chap. 5) leads us to consider Marshallian externalities as factors positively affecting the local productivity through the accumulation of some particular inputs available in the same area. Ever since the pioneering works of Thünen ([1826], 1966), Marshall ([1890], 1920), and Weber ([1909], 1929), such external effects have been at the heart of most explanations of economic agglomerations, but they tend to be invoked without much being said about their underpinnings. The analyses provided in previous chapters should shed light on what lies at the origin of such external economies. In this chapter, our research strategy is to view Marshallian externalities as technological externalities without accounting for the details of their specificity. This shortcut allows us to explore the common implications of Marshallian externalities for the spatial distribution of production activities.

The purpose of this chapter is twofold. First, we want to show how the introduction of externalities into the neoclassical theory of regional economics has affected the convergence property. We will see that even a small departure from the standard framework is sufficient for the most typical market outcome to involve a core–periphery structure in which one region is more prosperous than the other. Stated differently, there is no longer convergence between nations

or regions. Second, we wish to discuss the optimal strategies of firms facing the possibility of enjoying lower costs inside a cluster or of establishing their plants in spatially separated markets in order to enjoy more market power. The main result here is the emergence of an asymmetric distribution of firms in an otherwise symmetric environment.

In Section 8.2, we consider a simple setting with one mobile factor and one immobile factor developed by Michel, Perrot, and Thisse (1996). As we know, the neoclassical model predicts that the mobility of one production factor suffices to yield regional convergence. The situation may change dramatically if there is a production externality whose local intensity depends on the quantity of the mobile factor gathered into one region. We will see that such an externality may lead to *the expansion in wealth of the region* with the initial advantage at the expense of the other, thus providing an example of the circular causation process proposed by Myrdal (1957, chap. 3) to explain economic development. Similar results will be obtained in more specific contexts in Chapter 9.[5]

In the next two sections, we take a similar line of reasoning but focus on a microeconomic setting in which firms compete in price while being able to benefit from Marshallian externalities in their location choices (Belleflamme, Picard, and Thisse 2000). Specifically, we follow Chipman (1970) by assuming that firms belonging to the same sector benefit from a higher productivity when they locate together. We want to highlight how these external effects interact together with the dispersion forces generated by market competition within the global economy to lead to cluster formations of different sizes. Indeed, if localization economies are obviously an agglomeration force, it is also well known that geographical proximity renders price competition on the product market fiercer (Section 7.3). In other words, firms have an incentive to separate from one another to enjoy local market power. Consequently, whereas firms enjoy low costs when they concur in their locational choices, they can sell their products at higher prices when they are dispersed.

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 far apart when transport costs are high. Because the spatial distribution of demand is supposed to be unaffected by the locations and sizes of clusters, the cost reduction associated with the agglomeration may be more than offset by the fall in exports. By contrast, firms could enjoy higher profits by being local monopolists. Consequently, transport costs have to be low for firms to congregate. In other words, firms must be able to serve almost all markets equally (globalization) to enjoy the local advantages associated with the formation of a cluster (localization). Consequently, the formation of an industrial cluster appears to depend on the relative strength of three distinct forces: the magnitude of localization economies, the intensity of price competition, and the level of transport costs. In such a context, the number of firms is likely to matter, and we therefore consider two very different market structures, that is,

competition within a *small group* (Section 8.3) and within a *large group* of firms (Section 8.4).

It is reasonable to believe that the intensity of localization economies varies across regions. This is why we consider in Section 8.3 an oligopoly in which the production cost reduction firms may enjoy by being together changes with the region in which they set up. Such a setting provides a benchmark to study the decisions of a small number of large firms behaving strategically because each firm is aware that its locational choice affects not only its production cost but also its rivals'. This setting also captures some critical elements of the trade-off faced by large firms in designing their competitive strategy such as those belonging to the German chemical clusters.

In Section 8.4, we move to the case of a large group of firms in which there is no explicit strategic interaction: each firm has a negligible impact on the others but is aware that its locational choice affects its production cost because this factor depends on where its competitors are located. This market structure will allow us to study the emergence of industrial clusters involving a large number of small firms, such as those existing in Italy.

8.2 FACTOR MOBILITY AND AGGLOMERATION ECONOMIES

Consider an economy formed by two regions $r = A, B$. There are one product and two production factors. The product is costless to transport. One factor is immobile, whereas the other responds to market disequilibrium by moving into the more attractive region. Although other interpretations are possible, it is convenient to think of the mobile factor as the *skilled* workers and the immobile one as the *unskilled* workers. Indeed, empirical studies suggest that skilled workers are more mobile than the unskilled between distant locations (e.g., Shields and Shields 1989; SOPEMI 1998) perhaps because education generates human capital that is easily transferable to other regions.[6]

In region r, the output Y_r is produced according to the production function

$$Y_r = E(H_r)F(H_r, L_r),$$

where H_r stands for the number of skilled in region r, with $H = H_A + H_B$, whereas L_r represents the number of unskilled with $L = L_A + L_B$. Both H and L are fixed. The total number of workers in region r is denoted by $P_r = H_r + L_r$. Because we wish to focus on the pure effects of the forces at play, we consider the case of symmetric regions, and thus $L_A = L_B = 1$.

The production function, common to the two regions, consists of two parts: (1) a neoclassical production function $F(H_r, L_r)$ with constant returns and diminishing marginal product and (2) an externality function $E(H_r)$ that has the nature of a Hicksian shift factor (as in Section 4.4). For notational simplicity, we set

$$F(H_r, 1) \equiv f(H_r),$$

where $f'(H_r) > 0$, $f''(H_r) < 0$, whereas the Inada condition $f'(0) = \infty$ is supposed to hold. In the present context, it seems natural to assume that the intensity of the externality increases with the number (or mass) of skilled residing in the same region, and thus E is continuously increasing on $[0, H]$. Such a production externality is a force fostering the agglomeration of the skilled into a single region. However, increasing the number of skilled in a given region decreases their marginal productivity gross of the externality effect. As a result, the wage the skilled workers earn in the large region may eventually decline, thus leading to a dispersed pattern of production.

The regional output markets are perfectly competitive. There are no transportation costs, therefore, the price of the output is the same in the two regions and is normalized to 1. Similarly, regional factor markets are perfectly competitive, and thus the regional wage of each factor is given by its regional marginal productivity. Because each firm in region r is negligible to the market, it considers the externality $E(H_r)$ as given. Therefore, since both regional markets for the skilled are perfectly competitive, we have

$$w_r^H = E(H_r)f'(H_r) \quad r = A, B. \tag{8.1}$$

Workers have identical preferences. The utility of a j-worker ($j = H, L$) living in region r ($r = A, B$) is given by

$$U_r^j = u\left(w_r^j\right) + e_r(P_r),$$

where $u(w_r^j)$ is the indirect utility derived from the wage w_r^j, and $e_r(P_r)$ is a consumption externality depending on the total population residing in region r.[7] We assume that the consumption externality has the same functional form across the two groups of workers and is given by

$$e_r(P_r) = v\left(P_r, \frac{P_r}{S_r}\right),$$

where P_r/S_r is the density of population in region r, and S_r the area of this region. Note that S_r need not be the physical size of the region; for example, S_r could be the given amount of any public infrastructure or natural amenity consumed by workers. As for the unskilled, we assume that the two regions have identical endowments: $S_A = S_B = S$.

The function v is the same across regions so that none of them has a structural advantage. We also assume that $v_1' > 0$ and $v_2' < 0$. The first inequality ($v_1' > 0$) expresses a *conviviality effect* of the type analyzed in Section 6.2: the larger the population in region r, the higher the potential for social interaction. However, as discussed in Chapters 3 and 5, the positive effect of a larger population in region r is counterbalanced by *crowding effects* expressed by the second inequality ($v_2' < 0$). For example, a larger population living in a given area is likely to face a higher pollution and crime rate that reduces the workers' welfare.

Hence, there is a trade-off because the first effect encourages the agglomeration of skilled workers, whereas the second fosters their dispersion. When additional insights are required, we use the following specification for the consumption externality:

$$v(P_r) = aP_r - b(P_r/S)^2, \tag{8.2}$$

where a and b are two positive constants expressing the relative importance of the amenity and crowding effects in workers' well-being.

Let us now describe the migration behavior of the skilled. Migration is driven by the *utility differential*:

$$\dot{H}_A = \left[u\left(w_A^H\right) + e(P_A)\right] - \left[u\left(w_B^H\right) + e(P_B)\right], \tag{8.3}$$

for, as will be seen, H_A and H_B are always positive. In this case, the utility a skilled worker obtains in a given region depends on the regional wage, which varies with the number of skilled workers but also with the various consumption externalities generated inside this region by the local population. Various types of externalities are thus operative.

When regional amenities are weakly affected by the migration of the skilled, it is convenient to assume that they are the same in both regions, in which case migration is governed by the *wage differential*:

$$\dot{H}_A = u\left(w_A^H\right) - u\left(w_B^H\right). \tag{8.4}$$

8.2.1 Migrants as Labor Suppliers

We suppose here that the migration of the skilled workers has no significant impact on residents' well-being perhaps because a and b are very small.

Because $H_B = H - H_A$, we have $\dot{H}_B = -\dot{H}_A$ so that we can restrict ourselves to studying the dynamics of H_A. Substituting (8.1) into (8.4) yields the following equation of motion:

$$\begin{aligned}\dot{H}_A &= u[E(H_A)f'(H_A)] - u[E(H - H_A)f'(H - H_A)] \\ &\equiv \phi(H_A).\end{aligned} \tag{8.5}$$

As seen in the foregoing, two opposite effects at work. The equilibrium distribution of skilled workers is the outcome of this trade-off. Because $f'(0) = \infty$ and u is monotonically increasing, it follows that $\dot{H}_A > 0$ when $H_A \to 0$ and $\dot{H}_A < 0$ when $H_A \to H$. Furthermore, the function ϕ defined in (8.5) is continuous in H_A. The theorem of intermediate values therefore implies that at least one equilibrium distribution of the skilled exists at which $\dot{H}_A = 0$ and $0 < H_A < 1$. The preceding boundary conditions mean that all equilibria are interior. In other words, the immobility of the unskilled and the asymptotic

property of the production function f imply that all the skilled never concentrate into a single region.

A sufficient condition for a unique, and therefore stable, equilibrium is that ϕ be monotone decreasing. This is so when the function $E(x)f'(x)$ is strictly decreasing in x because u is strictly increasing. This condition means that an increase in the number of skilled in a given region is such that the positive effect of the production externality is always dominated by the negative effect of the marginal productivity decrease, or

$$\frac{E'(x)}{E(x)} < \frac{-f''(x)}{f'(x)} \quad x > 0.$$

This is clearly so when there is no production externality, as in the standard neoclassical model. Such a situation is depicted in Figure 8.1, where the equilibrium distribution of the skilled is symmetric and globally stable.

When there are multiple equilibria, not all equilibria are stable. Accordingly, the emerging equilibrium depends on the initial distribution of skilled workers for the path followed by the economy is unique. To illustrate, assume that the production externality is given by

$$E(H_r) = \exp(\varepsilon H_r) \quad r = A, B, \tag{8.6}$$

where ε is a positive constant. The production function is of the Cobb–Douglas type:

$$F(H_r, L_r) = H_r^{\alpha} L_r^{1-\alpha} \quad \text{or} \quad f(H_r) = H_r^{\alpha} \quad r = A, B \tag{8.7}$$

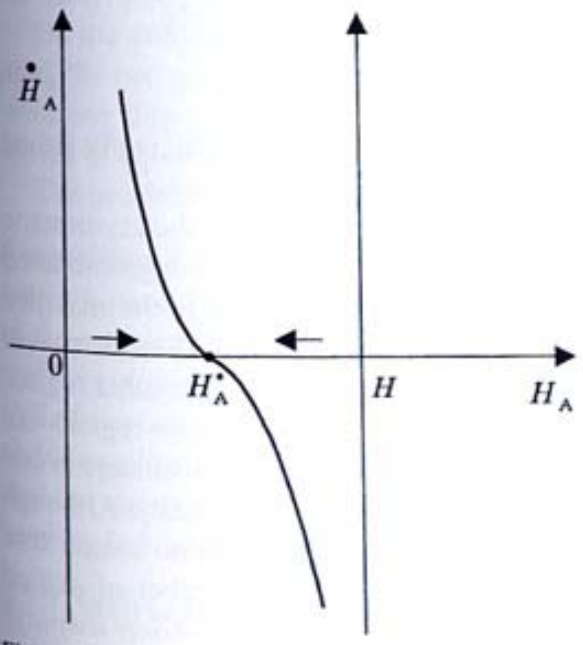


Figure 8.1: The symmetric equilibrium is stable.

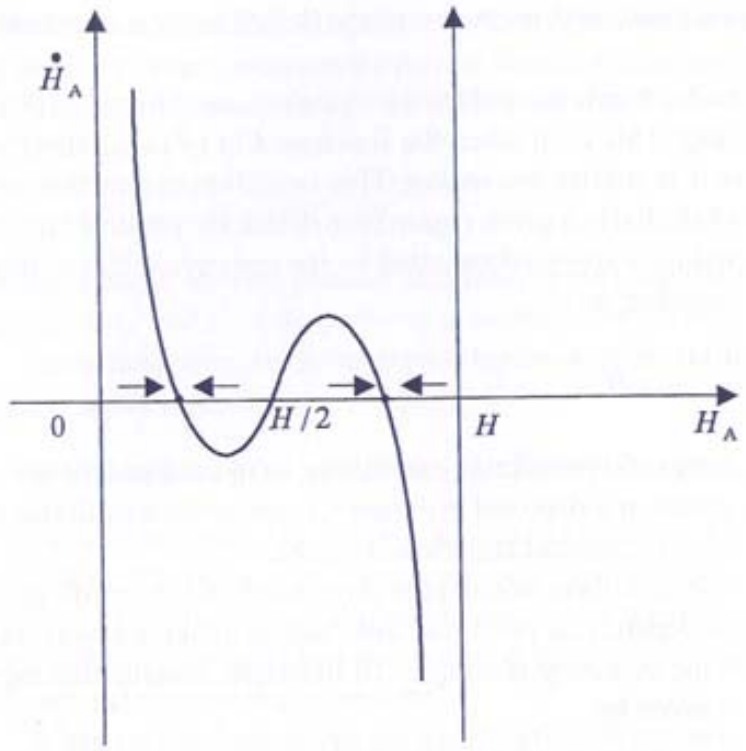


Figure 8.2: The partially agglomerated equilibria are stable.

with $0 < \alpha < 1$, whereas the utility u is

$$u\left(w_r^H\right) = \log\left(w_r^H\right) \quad r = A, B. \tag{8.8}$$

It is shown in Part A of the chapter appendix that there *three* interior equilibria exist (see Figure 8.2) if and only if the following condition holds:

$$\varepsilon H > 2(1-\alpha), \tag{8.9}$$

meaning that the intensity of the production externality is sufficiently strong compared with the share of the unskilled in the production process.

In this case, the symmetric equilibrium is unstable, whereas the asymmetric equilibria are stable. An asymmetric equilibrium involves more agglomeration when the production externality is stronger, when the cost share of the unskilled is weaker, or both. The region with the larger share of skilled workers will attract more and more workers and will grow at the expense of the other region, which loses a fraction of its skilled workers even when the two regions are initially almost similar. This process shows how a small initial advantage in one region is magnified through the action of the production externality. Although the skilled always earn the same wage in both regions, this is no longer true for the unskilled at each peripheral region: the larger the number of skilled in a region, the higher the wage of the unskilled. Thus, *the economy displays a core–periphery structure when the strength of the production externality is sufficiently strong.* In this case, there is a stable equilibrium in which regions

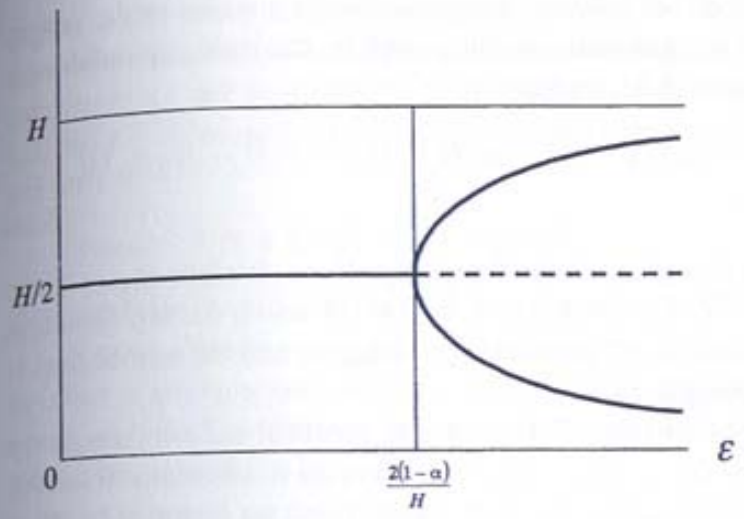


Figure 8.3: Bifurcation diagram of equilibria.

are ex post dissimilar although they are ex ante almost similar. It is worth noting that the reasons for preventing all the skilled from concentrating into a single region are the immobility and the "essentiality" of the unskilled (in the sense of the Inada limit condition). The structural changes in the spatial equilibrium in terms of the production externality parameter ε are depicted in Figure 8.3. Furthermore, although the unskilled initially earn (more or less) the same wage, those residing in the core region end up with higher wages, and those living in the periphery with lower wages, than had been the case initially. Hence, we have here a neat example in which agglomeration generates inequalities within the same group of individuals.

On the contrary, when ε, α, or both are sufficiently small for (8.9) not to hold, the two peripheral equilibria disappear, and there is a single equilibrium involving dispersion. Therefore, at $\varepsilon H = 2(1-\alpha)$ there is a discontinuity in the set of equilibria.

The preceding discussion is summarized in the following proposition.

Proposition 8.1 *Assume two regions with the same endowment of unskilled who are immobile and essential, whereas regional amenities are independent of the population size. If the production externality is not too strong, a unique stable equilibrium exists, and this equilibrium involves equal dispersion of the skilled workers. However, if the production externality is sufficiently strong, any stable equilibrium involves more than half of the skilled workers in one region, and the wage of the unskilled in the region having the larger number of skilled workers is higher.*

8.2.2 Migrants as Labor Suppliers and Amenity Consumers

We now assume that the migration of the skilled affects the level of amenities in the two regions. Accordingly, the (indirect) utility that an individual obtains

in a given region depends not only on the regional wage but also on the various consumption externalities generated in this region by the local population.

Substituting (8.1) into (8.3), we get

$$\begin{aligned}\dot{H}_A &= [u(H_A) + e_A(H_A + 1)] - [u(H - H_A) + e(H - H_A + 1)] \\ &\equiv \psi(H_A),\end{aligned} \tag{8.10}$$

where $u(x) \equiv u[E(x)f'(x)]$. As in the foregoing, $\dot{H}_A > 0$ when $H_A \to 0$ and $\dot{H}_A < 0$ when $H_A \to H$. Furthermore, ψ is continuous on $[0, H]$. Hence, at least one equilibrium exists, all equilibria are interior, and the equilibrium is globally stable when unique.

Because we consider the case of regions with identical endowments, it can readily be verified that $H_A^* = H_B^* = H/2$ is always an equilibrium. However, this equilibrium may be unstable. To study its behavior, we linearize (8.10) in the neighborhood of $H/2$ and obtain

$$\dot{H}_A \simeq 2m(H_A - H/2)[u'(H/2) + e'(H/2 + 1)], \tag{8.11}$$

where m is a positive constant. It then follows from (8.11) that the symmetric equilibrium is stable if and only if

$$u'(H/2) + e'(H/2 + 1) < 0. \tag{8.12}$$

This means that the marginal effect of increasing the number of skilled in any region is negative at the symmetric equilibrium. More precisely, $u'(H/2) < 0$ suggests that the equilibrium wage of the skilled decreases with their number when these workers are equally distributed between the two regions, whereas $e'(H/2 + 1) < 0$ implies that the marginal crowding effect dominates the conviviality effect when the population is evenly split. As before, regional wages are equal inside each group of workers for the symmetrical equilibrium.

When the opposite inequality holds in (8.12), the symmetric equilibrium is unstable, thus implying that the market equilibrium involves some concentration of the skilled even though both regions are a priori identical. Either the production externality or the conviviality effect destroys the stability of the symmetric equilibrium and leads to outcomes in which regions are ex post different. In particular, at any stable asymmetric equilibrium, each type of worker earns different wages in different regions. In other words, persistent wage differentials may well occur even when regions are ex ante identical. For the skilled, this result occurs because earning differential compensates for amenities and congestion differences. This is compatible with an equilibrium state in which the skilled have the same utility level. However, for the unskilled, wage difference reflects differences in marginal productivities as a result of the imbalance in the spatial distribution of skilled workers.

To get a richer characterization of the equilibrium, we return to the special case given by (8.6)–(8.8) as well as by (8.2). It is shown in Part A of the chapter

appendix that there are at most three equilibria. The stability properties of these equilibria are similar to those discussed in Section 8.2.1, and equilibrium configurations are qualitatively comparable to those depicted in Figures 8.1, 8.2, and 8.3. Because both regions are ex ante identical, three equilibria exist if and only if

$$\varepsilon + a - b(H+2)/S^2 > 2(1-\alpha)/H. \tag{8.13}$$

This condition reduces to (8.9) when there is no consumption externality ($a = b = 0$). Clearly, inequality (8.13) is likely to hold when the conviviality effect is strong relative to the crowding effect, the production externality is strong, or both. Note that we may have three equilibria even though there is no production externality ($\varepsilon = 0$).

In equilibrium, we must have

$$\{a - b[(H+1)/S)^2]\}(2H_A - H) = \log w_A^H - \log w_B^H.$$

Because the derivative of this expression with respect to H_A is equal to

$$2\{a - b[(H+1)/S)^2]\} = e'[(H+1)/2]$$

the wage differential varies monotonically with H_A. Consequently, when there are three equilibria, the two stable asymmetric equilibria involve wage dispersion for both types of workers.

We may summarize this discussion as follows:

Proposition 8.2 *Assume two regions with the same endowment of unskilled who are immobile and essential. When the production externality is sufficiently weak, the crowding effect is strong enough, or both, there is a single stable equilibrium that involves full dispersion of the skilled workers. Otherwise, there are several stable equilibria such that more than one half of the skilled are in one region and the wage within each group also varies between regions.*

A few general principles emerge from this analysis.

1. Technological externalities may be sufficient to generate agglomeration and a core–periphery structure even when the economy is perfectly competitive.
2. The geographical immobility of one group of workers (or of one production factor such as land) may be sufficient to preclude the concentration of production in a small number of regions. Indeed, if this group of workers is essential for the production process, the law of diminishing returns is applicable to the other groups, thus reducing the benefit these groups may have from being fully agglomerated. This result fosters the dispersion of production. It is only when the agglomeration externalities are sufficiently large that substantial regional discrepancies may arise.

3. When amenities vary with the population size, the equality of wages between regions does not necessarily hold in equilibrium. Labor mobility aims at equilibrating the combination of wages and external effects.
4. The existence of a production externality implies that the global product is maximized when both groups of workers are concentrated in a single region. As expected, the immobility of the unskilled leads to efficiency losses – at least when amenities are not affected negatively by population size. However, global product as a social objective may be rather a poor goal because narrowing regional discrepancies in per capita incomes may also be a policy objective (Michel, Pestieau, and Thisse 1983). In this perspective, the "emptiness" of a region would appear as a negative outcome.
5. Even when both skilled and unskilled workers are mobile, agglomeration externalities and crowding operate in opposite directions. Thus, the market equilibrium crucially depends on the relative strengths of these two effects, and making predictions about the spatial equilibrium therefore appears to be hard.[8]

8.3 OLIGOPOLY, LOCALIZATION ECONOMIES, AND REGIONAL ADVANTAGE

Consider an industry with two firms (say 1 and 2) each producing a differentiated variety of the same product. 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 here and in the next section that regions A and B are characterized by the same market conditions. This implies that consumers stay put and do not change their locations with those of the firms considered. More precisely, in each region the demand functions for the two varieties are generated by a representative consumer who has the following quadratic utility function:

$$U(q_1, q_2) = \alpha(q_1 + q_2) - (\beta/2)\left(q_1^2 + q_2^2\right) - \delta q_1 q_2 + z, \tag{8.14}$$

where q_i $(i = 1, 2)$ is the quantity of variety i, and z is the quantity of numéraire consumed. As usual, we have $\alpha > 0$ and $0 \leq \delta < \beta$. The consumer's budget constraint is $y = p_1 q_1 + p_2 q_2 + z$.

Maximizing (8.14) 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

$$q_i = a - bp_i + d(p_j - p_i), \tag{8.15}$$

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

The demand system (8.15) may then be interpreted as follows. Parameter d is an inverse measure of the degree of product differentiation between varieties:

they are independent when $d = 0$ (i.e., $\delta = 0$) and perfect substitutes when $d \to \infty$ (i.e., $\beta = \delta$). In other words, increasing the degree of product differentiation between varieties amounts to decreasing d.[9] The own-price effect is stronger (as measured by $b + d$) than each cross-price effect (as measured by d). Finally, parameter b gives the link between individual and industry demand (total demand becomes perfectly inelastic when $b \to 0$, as in the Hotelling model).

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 units of the numéraire. 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 a constant $c > 0$.
- When firms are located in the same region, they benefit from some positive localization economies. This means that their marginal cost is reduced by a positive amount that may be region-specific. More precisely, if both firms locate in region $r = A, B$, firm i's cost is given by $c - \theta_r$.

In other words, firms experience the same reduction in their marginal production cost when they locate together. However, this reduction may depend on the region where they locate because the nature and intensity of nonmarket interactions between firms may vary from one region to the other. For example, the stock of skills accumulated in a region cannot be moved to another one. A worker operating in a repeated production context continuously uncovers the many facets of the available techniques and gradually adjusts her behavior so as to improve productivity over time. However, workers are heterogeneous not only because they have different skills (by nature as well as by nurture) but also because they face different experiences and have different abilities of learning. When they live in the same locale, they can share their knowledge through various types of social interactions which, in turn, increases their productivity when they are combined within firms. Such features are specific to particular regions. Without loss of generality, it is assumed 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. Two different subgames are to be considered according to whether the firms are located together or separately. In both cases, it is supposed that markets are segmented, that is, each firm sets a price specific to the market in which its product is sold (see Section 7.3.3). Indeed, there are many good reasons to believe that firms want to use spatial separation to segment their market (Horn and Shy 1996; Thisse and Vives 1988), and empirical work confirms the assumption that international, not to say interregional, markets are still very segmented (see, e.g., McCallum 1995; Head and Mayer 2000; Wolf 2000).

8.3.1 Interregional Price Competition

1. Assume that both firms are located in region r. Let p_{ir} and q_{ir}, respectively, denote the price and quantity of the product sold by firm i in region r. Firm i's problem entails choosing the prices p_{ir} (the home price) and p_{is} (the foreign price) that maximize its profit function defined as follows:

$$\pi_i = [p_{ir} - (c - \theta_r)][a - bp_{ir} + d(p_{jr} - p_{ir})] + [p_{is} - (c - \theta_r) - t][a - bp_{is} + d(p_{js} - p_{is})]. \tag{8.16}$$

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, we obtain the following equilibrium prices:

$$p_{ir} = p_{jr} = \frac{a + (b + d)(c - \theta_r)}{2b + d} \equiv p_r^h \tag{8.17}$$

$$p_{is} = p_{js} = \frac{a + (b + d)(c - \theta_r + t)}{2b + d} \equiv p_r^f,$$

where we use the subscript r to refer to the case in which both firms are located in region r and use h and f to denote variables related to the home and foreign markets.

Equilibrium quantities such as the following are easily found:

$$q_{ir} = q_{jr} = \frac{(b + d)[a - b(c - \theta_r)]}{2b + d} \equiv q_r^h \tag{8.18}$$

$$q_{is} = q_{js} = \frac{(b + d)[a - b(c - \theta_r + t)]}{2b + d} \equiv q_r^f.$$

It can readily be checked that $q_r^h = (b + d)(p_r^h - c + \theta_r)$ and $q_r^f = (b + d)(p_r^f - c + \theta_r - t)$. Thus, the equilibrium quantities and markups are positive (meaning that we have an interior solution and that both firms export to region s) provided that[10]

$$t < \frac{a - bc}{b} - \theta_r. \tag{8.19}$$

Plugging (8.17) and (8.18) into (8.16), we obtain each firm's common equilibrium profits when they are located together in region r as follows:

$$\pi_r = \frac{b + d}{(2b + d)^2}\{[a - b(c - \theta_r)]^2 + [a - b(c - \theta_r + t)]^2\}. \tag{8.20}$$

2. Suppose now that firm i is located in region r and firm j in region $s \neq r$. Firm i's profit function is now written as follows:

$$\pi_i = (p_{ir} - c)[a - bp_{ir} + d(p_{jr} - p_{ir})] + (p_{is} - c - t)[a - bp_{is} + d(p_{js} - p_{is})]. \tag{8.21}$$

Again, a similar expression holds for firm j.

Taking the first-order conditions and solving the corresponding system of four equations, we obtain the following equilibrium prices:

$$p_{ir} = p_{js} = \frac{a + (b+d)c}{2b+d} + \frac{(b+d)dt}{(2b+d)(2b+3d)} \equiv p_S^h$$

$$p_{is} = p_{jr} = \frac{a + (b+d)c}{2b+d} + \frac{2(b+d)^2 t}{(2b+d)(2b+3d)} \equiv p_S^f,$$

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

It can readily be verified that price differentials between domestic and foreign markets are as follows:

$$p_{is} - p_{ir} = p_{jr} - p_{js} = \frac{b+d}{2b+3d} t < t.$$

Consequently, when firms are separate there is *intraindustry trade* as well as *reciprocal dumping* (Brander and Krugman 1983; Anderson, Schmitt, and Thisse 1995), whereas arbitrage is never profitable.

Equilibrium quantities are then easily computed as follows:

$$q_{ir} = q_{js} = \frac{(b+d)(a-bc)}{2b+d} + \frac{(b+d)^2 dt}{(2b+d)(2b+3d)} \equiv q_S^h$$

$$q_{is} = q_{jr} = \frac{(b+d)(a-bc)}{2b+d} - \frac{(b+d)(2b^2+4bd+d^2)t}{(2b+d)(2b+3d)} \equiv q_S^f.$$

As above, one can check that $q_S^h = (b+d)(p_S^h - c)$ and $q_S^f = (b+d)(p_S^f - c - t)$. Using these findings, we can express the condition for an interior solution as

$$t < t_{\text{trade}} \equiv \frac{(2b+3d)(a-bc)}{2b^2+4bd+d^2}. \tag{8.22}$$

In what follows, we assume that condition (8.22) is met. Moreover, it is easy to show that condition (8.22) is more stringent than (8.19) for $r = A, B$. In other words, we assume that the transport cost t is low enough to allow firms to export

Table 8.1. First-stage game

	A	B
A	Π_A, Π_A	Π_S, Π_S
B	Π_S, Π_S	Π_B, Π_B

their product whatever their location. Note that condition (8.22) 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

$$\pi_S = \frac{b+d}{(2b+d)^2}\left\{\left[a - bc + \frac{(b+d)dt}{2b+3d}\right]^2 + \left[a - bc - \frac{(2b^2+4bd+d^2)t}{2b+3d}\right]^2\right\}. \tag{8.23}$$

8.3.2 Spatial Equilibrium

In the first stage, firms 1 and 2 simultaneously choose their location. Thus, they play the game depicted in Table 8.1.

Comparing expressions (8.21) and (8.23), one can readily verify that $\pi_r > \pi_S$ if and only if $\theta_r > \theta_P(t)$, where

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

It can readily be verified that $\theta_P(t)$ is an increasing function of t and d.

When $\theta_A > \theta_B$, three cases may arise as follows:

Proposition 8.3 *Assume that (8.22) holds. For any triple (θ_A, θ_B, t) such that $\theta_B < \theta_A$, the outcome of the game takes one of the following forms:*

1. *If $\theta_A < \theta_P(t)$, then $\pi_S > \pi_A > \pi_B$, and the equilibrium is unique (up to a permutation) and involves dispersion.*
2. *If $\theta_B < \theta_P(t) < \theta_A$, then $\pi_A > \pi_S > \pi_B$ and the unique equilibrium involves agglomeration in region A.*
3. *If $\theta_P(t) < \theta_B < \theta_A$, then $\pi_A > \pi_B > \pi_S$ and there are two equilibria involving agglomeration in region A or in region B (the former equilibrium Pareto-dominates the latter).*[11]

In words, for a given value of the transportation cost, this means that firms must be compensated by a sufficient amount of localization economies for the

increased competition resulting from a common location. Or, to put it differently, *when transport costs are low, agglomeration is the market outcome because each firm can benefit from a production cost reduction by being together without losing much business in the other region.* It is worth noting that the threshold level $\theta_P(t)$ decreases when the degree of product differentiation rises (that is, when d falls). Indeed, more product differentiation relaxes price competition and makes the agglomeration of firms more likely. On the other hand, $\theta_P(t)$ rises with the transportation cost because higher trade costs strengthen the benefits of geographical isolation. If there were no localization economies ($\theta_A = \theta_B = 0$), equilibrium would always involve dispersion.

The preceding inequalities may be reinterpreted in the context in which 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 is in region B for some historical reasons, the agglomeration will occur there even though this region is less efficient. This result provides another illustration of the phenomenon of *lock-in effect* associated with the presence of Marshallian externalities and shows once more how history matters in the development of a particular region.[12]

8.3.3 Welfare

Plugging the linear inverse demands into function (8.14) and using the definitions of b and d, we can readily derive the general formulation of the consumers' surplus as follows:

$$C = \frac{1}{2b(b+2d)}\left[(b+d)\left(q_1^2+q_2^2\right)+2dq_1q_2\right]. \tag{8.24}$$

We adopt the following notation. Let C_s denote the surplus for the consumers in region $s = A, B$, and let the location of the firms be represented by r if both firms are in $r = A, B$, or S if they are in separate locations. Similarly, let $C(r) = C_A(r) + C_B(r)$ and $C(S) = C_A(S) + C_B(S)$, respectively, denote the total consumer surplus when both firms are in region r or when they are separate. Then, (8.24) yields

$$C_r(r) = (1/b)\left(q_r^h\right)^2 \qquad r = A, B \quad \text{and} \quad C_r(s) = (1/b)\left(q_s^f\right)^2 \qquad r \neq s$$

$$C_r(S) = \frac{1}{2b(b+2d)}\left\{(b+d)\left[\left(q_S^h\right)^2+\left(q_S^f\right)^2\right]+2dq_S^hq_S^f\right\} \equiv C_S, \quad r = A, B.$$

We start by considering the *first best* situation in which the planner is able to control both the locations of firms and their prices, whereas lump-sum transfers are available. Because of marginal cost pricing, the following quantities are

supplied:

$$q_r^h = a - b(c - \theta_r) \quad \text{and} \quad q_r^f = a - b(c - \theta_r + t) \qquad r = A, B$$
$$q_S^h = a - bc \quad \text{and} \quad q_S^f = a - b(c + t).$$

Because firms earn zero profits, the global welfare is equal to the global consumer surplus. A simple calculation reveals that $C(A) \geq C(B) > C(S)$. In other words, we have the following:

Proposition 8.4 *Assume that (8.22) holds. 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 proposition implies that it is always socially desirable that firms be agglomerated when trade occurs. Yet, when the intensity of the localization economies in region A is not large ($\theta_P(t) > \theta_A$), strategic competition leads to dispersion. However, the reverse does not hold because the market never yields excessive agglomeration of firms in this setting.

When transport costs are low enough, the market location outcome 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 is at the origin of another efficiency loss when agglomeration arises in region B, whereas it is socially desirable that firms located together in region A ($\theta_P(t) < \theta_B < \theta_A$).[13] This happens when one firm first establishes itself in a region that offers a low potential for the development of localized externalities.

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 preceding expressions, we have that the consumers in region r earn the following surpluses according to whether (1) two firms are located in their region, or (2) one or zero firms is located in region r:

$$C_r(r) = \frac{(b+d)^2}{b(2b+d)^2}[a - b(c - \theta_r)]^2$$

$$C_r(s) = \frac{(b+d)^2}{b(2b+d)^2}[a - b(c - \theta_s + t)]^2$$

$$C_S = \frac{(b+d)^2}{b(2b+d)^2}\left[(a - bc)(a - bc - bt) + \frac{(b+d)(4b^2 + 8bd + d^2)bt^2}{2(2b+3d)^2}\right].$$

Let us first adopt the viewpoint of local governments. From our assumptions that $\theta_A \geq \theta_B$ and $t > \theta_A - \theta_B$, it is easily seen that consumer surpluses in the case of common location are ranked as follows: $C(A) \geq C_B(B) > C_B(A) > C_A(B)$.

Furthermore, if we compare the consumer surpluses when firms are located either together or separately, we can establish the following two results: (1) $C_A(A) \geq C_B(B) > C_S$; (2) $C_r(L) > C_S$ if and only if $\theta_s > \theta_R(t)$, 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 8.5 *Assume that $t > \theta_A - \theta_B$. The second best optimum is such that*

1. *consumers in region $r = A, B$ are better off when both firms locate in their region than when only one does so;*
2. *if the intensity of localization economies in region r exceeds some threshold $\theta_R(t)$, even consumers in region $s \neq r$ are better off when no firm locates in their region than when one does.*

Thus, a local government should not necessarily strive to attract a firm when localization economies are strong because such a policy, when it is effective, can make the members of its constituency worse off. Of course, such considerations do not account for the welfare gain associated with the creation of jobs accompanying the establishment of a new firm in this region. This analysis 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.[14]

We now consider the global welfare. As far as firm's interests are concerned, we already know that total profits are higher when both firms locate in region r than when they separate provided that $\theta_r > \theta_P(t)$. Comparing global surpluses for the consumers for different locations, we see that $C(r) > C(S)$ if and only if $\theta_r > \theta_C(t)$, 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 $W(r) > W(S)$ if and only if $\theta_r > \theta_W(t)$, 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}. It then appears that the second best outcome may involve agglomeration even though the market selects dispersion, but the reverse is never true.

In addition, the preceding inequalities also mean that the interests of the various economic groups may be vastly different in the choice of a locational 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, whereas 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 in which $\theta_W(t) < \theta_A < \theta_C(t) < \theta_P(t)$. Then, firms choose separate locations 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 (which would prefer agglomeration in A).

In summary, the existence of localization economies may explain why oligopolistic firms want to be together despite the fiercer competition triggered by a common location. As in Chapter 7, such a clustering often appears to be globally efficient and may even lead to an increase in the welfare of consumers leaving the region that does not accommodate the cluster. However, this gain in efficiency is obtained at the expense of a higher degree of spatial inequity. In addition, the welfare analysis of the various groups involved reveals the existence potential conflicts of interest, thus suggesting that policy recommendations are not easy to make once it is recognized that lump-sum transfers are not available.

8.4 THE FORMATION OF INDUSTRIAL CLUSTERS UNDER LOCALIZATION ECONOMIES

In this section, we consider an industry with a continuum of a given size M of firms producing each a variety of a (horizontally) differentiated product. The representative consumer's utility function is now as follows:[15]

$$U(z; q(i), i \in [0, M]) = \alpha \int_0^M q(i)di - \frac{\beta - \delta}{2} \int_0^M [q(i)]^2 di - \frac{\delta}{2}\left[\int_0^M q(i)di\right]^2 + z, \tag{8.25}$$

where $q(i)$ is the quantity of variety $i \in [0, M]$ and z the quantity of the numéraire, whereas the parameters in (8.25) are such that $\alpha > 0$ and $\beta > \delta > 0$. In this expression, α is a measure of the intensity of preferences for the

differentiated product with respect to the numéraire, whereas $\beta > \delta$ means that the representative consumer has a love for variety. Suppose, indeed, that a consumer consumes a total of $Mq \equiv \int_0^M q(i)di$ of the differentiated product such that the consumption is uniform on $[0, x]$ and zero on $(x, M]$. Then, the density on $[0, x]$ is Mq/x. Equation (8.25) evaluated at this consumption pattern is

$$\begin{aligned} U &= \alpha \int_0^x \frac{Mq}{x} di - \frac{\beta - \delta}{2} \int_0^x \left(\frac{Mq}{x}\right)^2 di \\ &\quad - \frac{\delta}{2} \int_0^x \int_0^x \left(\frac{Mq}{x}\right)^2 didj + z \\ &= \alpha Mq - \frac{\beta - \delta}{2x} M^2 q^2 - \frac{\delta}{2} M^2 q^2 + z, \end{aligned} \tag{8.26}$$

which is increasing in x and, hence, maximized at $x = M$, where variety consumption is maximal. We may then conclude that, when $\beta > \delta$, the quadratic utility function exhibits a preference for variety.[16] Finally, for a given value of β, the parameter δ expresses the substitutability between varieties: the higher δ is, the closer substitutes are the varieties.

The consumer is endowed with $\bar{z} > 0$ units of the numéraire. Her budget constraint can then be written as follows:

$$\int_0^M p(i)q(i)di + z = \bar{z},$$

where $p(i)$ is the price of variety i. The initial endowment $\bar{z}$ is supposed to be sufficiently large for the optimal consumption of the numéraire to always be strictly positive at the market outcome. Solving the budget constraint for the numéraire consumption, plugging the corresponding expression into (8.25), and taking the first-order conditions with respect to $q(i)$, we obtain

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

which gives the demand function for variety $i \in [0, M]$:[17]

$$\begin{aligned} q(i) &= a - bp(i) + d \int_0^M [p(j) - p(i)]dj \\ &= a - (b + dM)p(i) + dP, \end{aligned} \tag{8.27}$$

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

$$P = \int_0^M p(j)dj$$

so that P/M may be interpreted as the price index. As in the preceding section, the parameter d represents an inverse measure of the degree of product differentiation among varieties: decreasing d amounts to increasing the degree of product differentiation among the given set of varieties. Furthermore, the own-price effect is stronger (as measured by $b + dM$) than each cross-price effect (as measured by d), thus allowing for different elasticities of substitution between pairs of varieties. Because the mass of firms is fixed and given, we may assume without substantial loss of generality that $M = 1$, and thus $a \equiv \alpha/\beta$, $b \equiv 1/\beta$, and $d \equiv \delta/\beta(\beta - \delta)$.[18] The more general case in which M may differ from 1 will be used in Chapter 9.

As in the preceding section, there are two regions that have the same market conditions. It is also assumed that the two regions are homogeneous with respect to the law governing localization economies: when there are M_r firms in region r, firm i is able to produce the variety i at marginal cost $c_r \equiv c(M_r)$. Of course, we must have $M_A + M_B = 1$. Let t be the unit transport cost between the two regions expressed in terms of the numéraire. It is assumed that

$$t < \frac{a}{b} - \max\{c(M_A), c(M_B)\}, \tag{8.28}$$

meaning that even a firm facing the higher marginal cost always finds it profitable to export to the other region.

8.4.1 Equilibrium Pricing

We study here the process of competition between firms for a given spatial distribution (M_A, M_B) of firms. Because 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 accurately neglects the impact of its decision over the prices selected by the other firms. In addition, because firms sell differentiated varieties, each firm has some monopoly power because it faces a demand function with finite elasticity. All of this is in accordance with Chamberlin's (1933) large group competition in which 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, a firm must account for the distribution of the firms' prices through some aggregate statistics, given here by the *price index*, to find its equilibrium price. Consequently, our market solution is given by a Nash equilibrium with a continuum of players in which prices are interdependent: each firm neglects its impact on the market but is aware that the market as a whole has a nonnegligible impact on its behavior.

Again, we assume that firms compete in segmented markets. In what follows, we suppose that the parameters of the economy are such that the equilibrium prices exceed costs and markups are positive (meaning that we have an interior

solution and that exportation occurs for all firms). A sufficient condition for this assumption to hold will be given below.

Assume that variety i is produced in $r = A, B$. The corresponding firm sells its output on both markets, that is, quantity $q_{rr}(i)$ at price $p_{rr}(i)$ on market r, and quantity $q_{rs}(i)$ at price $p_{rs}(i)$ on market $s \neq r$. Thus, $p_{rr}(i)$ is the price in region r of variety i produced locally, whereas $p_{rs}(i)$ the price of the same variety exported from r to s. We adopt the following notation:

$$P_{rr} \equiv \int_{i \in M_r} p_{rr}(i)\,di \quad \text{and} \quad P_{rs} \equiv \int_{i \in M_r} p_{rs}(i)\,di \qquad s \neq r.$$

Demands for firm i located in r are then given by

$$q_{rr}(i) = a - (b+d)p_{rr}(i) + d(P_{rr} + P_{sr}) \qquad s \neq r$$

and

$$q_{rs}(i) = a - (b+d)p_{rs}(i) + d(P_{rs} + P_{ss}) \qquad s \neq r.$$

Firm i in r maximizes its profits defined by

$$\pi_r(i) = [p_{rr}(i) - c_r]q_{rr}(i) + [p_{rs}(i) - c_r - t]q_{rs}(i) \qquad s \neq r.$$

Deriving the first-order conditions for profit maximization with respect to prices and summing them for all firms located in r, we obtain the following expressions:

$$[2(b+d) - dM_r]P_{rr} - dM_r P_{sr} = M_r[a + (b+d)c_r] \qquad s \neq r \tag{8.29}$$

$$[2(b+d) - dM_A]P_{rs} - dM_r P_{ss} = M_r[a + (b+d)(c_r + t)] \qquad s \neq r. \tag{8.30}$$

Varieties being symmetric, in equilibrium we have

$$p_{rs}(i) = p_{rs} = P_{rs}/M_r.$$

Because profit functions are concave in own price, solving the system of equations (8.29)–(8.30) yields the equilibrium prices

$$p^*_{rr} = \frac{c_r}{2} + \frac{2a + d[M_r c_r + M_s(c_s + t)]}{2(2b+d)}$$

$$p^*_{rs} = \frac{c_r + t}{2} + \frac{2a + d[M_r(c_r + t) + M_s c_s]}{2(2b+d)} \qquad s \neq r.$$

As expected, the equilibrium prices depend on the distribution of firms between the two regions. Prices 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 (8.28) holds. All these results agree with what is known in industrial organization and spatial pricing theory. By inspection, it can also be readily verified that the local price, p_{rr}^*, increases with t because the local firms in r are more protected against distant competitors, whereas the export price, $p_{rs} - t$, decreases 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 (or mass) of local firms in region r increases, while the total number of firms is held constant, if and only if $c_r < c_s + t$. This occurs because the lower cost prevailing in r intensifies local price competition.

Using the first-order conditions, we can easily establish the following relationships between equilibrium prices and quantities:

$$q_{rr} = (b+d)(p_{rr} - c_r) \quad \text{and} \quad q_{rs} = (b+d)(p_{rs} - c_r - t).$$

The equilibrium profits of any firm located in region $r \neq s$ are thus

$$\begin{aligned}\pi_r(M_r, M_s) &= (p_{rr} - c_r)q_{rr} + (p_{rs} - c_r - t)q_{rs} \\ &= (b+d)[(p_{rr} - c_r)^2 + (p_{rs} - c_r - t)^2] \\ &= \frac{b+d}{2(2b+d)^2}\{[2(a - bc_r) - dM_s(c_r - c_s) - bt]^2 \\ &\quad + (b + dM_s)^2 t^2\}.\end{aligned}$$

In the remainder of the section, it is assumed that localization economies obey the same law in each region $r = A, B$ given by

$$c_r(M_r) = c - \theta M_r$$

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 can readily be checked that a sufficient condition is that $q_{rs} > 0$ in the limiting case in which $M_r = 0$,

$$t < t'_{\text{trade}} \equiv \frac{2(a - bc) - d\theta}{2b + d}, \tag{8.31}$$

whose right-hand side is supposed to be positive. Note that (8.28) is satisfied when (8.31) holds.

8.4.2 Spatial Equilibrium

We can take advantage of the symmetry of the problem by setting $\Delta M = M_A - M_B$. Thus, $M_A = (1/2)(1 + \Delta M)$, $M_B = (1/2)(1 - \Delta M)$, $c_A(M_A) + c_B(M_B) = 2c - \theta$, and $c_A(M_A) - c_B(M_B) = -\theta \Delta M$. Consequently, the

equilibrium profits can be rewritten as follows (where D stands for $2b + d$):

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

$$\pi_B(\Delta M) = \frac{b+d}{8D^2}\{[4a - 2b(2c + t - \theta) - d\theta(\Delta M)^2 - D\theta\Delta M]^2 + [D + d(\Delta M)]^2 t^2\}.$$

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

$$\Delta\pi(\Delta M) = \frac{b+d}{2D}\Delta M\{[4a - 2b(2c + t - \theta) - d\theta(\Delta M)^2]\theta - dt^2\},$$

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

$$\Delta\pi(\Delta M) = \frac{d\theta^2(b+d)}{2D}\Delta M[\Lambda - (\Delta M)^2], \tag{8.32}$$

where Λ is a constant given by

$$\Lambda \equiv \frac{[4a - 2b(2c + t - \theta)]\theta - dt^2}{d\theta^2}, \tag{8.33}$$

whereas its slope is

$$\frac{d(\Delta\pi)}{d(\Delta M)} = -\frac{d\theta^2(b+d)}{2D}[3(\Delta M)^2 - \Lambda]. \tag{8.34}$$

The distribution $M_A \in [0, 1]$ is a *spatial equilibrium* when no firm may earn a higher profit by changing location. This arises at an interior point $M_A \in (0, 1)$ when

$$\Delta\pi(2M_A - 1) = 0$$

or at $M_A = 0$ when $\Delta\pi(-1) \leq 0$, or at $M_A = 1$ when $\Delta\pi(1) \geq 0$. In the first case, we have either a *fully dispersed configuration* or a *partially agglomerated configuration*; in the last two cases, we have a *fully agglomerated configuration*.

Given (8.32), the symmetric configuration ($\Delta M = 0$) is always an equilibrium. However, because other spatial equilibria exist, it is convenient to use stability as a selection device because an unstable equilibrium is unlikely to happen. To study the stability of a spatial equilibrium, we must define how firms behave away from the equilibrium. In the present context, the dynamics is fairly natural. If firms observe that a region offers higher profits than the other, they want to move to that location. In other words, the driving force in the *profit*

differential between A and B,

$$\dot{M}_A = M_A \Delta\pi(\Delta M) M_B,$$

where $\dot{M}_A$ is the time-derivative of M_A. If $\Delta\pi$ is positive and if $0 < M_A < 1$, 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 from the equilibrium, the equation of motion above brings the distribution of firms back to the original one. Therefore, a fully agglomerated configuration is always stable when it turns out to be an equilibrium, whereas an interior equilibrium is stable if and only if the slope of $\Delta\pi(\Delta M)$ is nonpositive in a neighborhood of this equilibrium. Clearly, a spatial equilibrium is a steady-state for the equation $\dot{M}_A = 0$, and conversely.

Several kinds of stable equilibria may arise in the present 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 ($\Delta M = 0$) or unevenly across regions. These equilibria are now fully characterized.

Proposition 8.6 *The two-region economy always has a single (up to a permutation) stable spatial equilibrium. This one involves the following:*

1. *symmetry ($\Delta M = 0$) if and only if $\Lambda \leq 0$;*
2. *partial agglomeration ($\Delta M = \pm\sqrt{\Lambda}$) if and only if $0 < \Lambda < 1$;*
3. *full agglomeration ($\Delta M = \pm 1$) if and only if $1 \leq \Lambda$.*

Proof On the basis of (8.32), the proof involves three steps.

1. When $\Lambda < 0$, the equation $\Delta\pi = 0$ has a single real solution $\Delta M = 0$ at which the slope is negative. Thus, $\Delta M = 0$ is the unique equilibrium and is stable.
2. When $\Lambda \geq 0$, the equation $\Delta\pi = 0$ has three real solutions ($\Delta M = 0$ and $\Delta M = \pm\sqrt{\Lambda}$) and one of the following two cases must arise:
   A. The asymmetric solutions are equilibria if and only if $0 < \Lambda < 1$. They are stable because the slope of $\Delta\pi$ evaluated at $\Delta M = \pm\sqrt{\Lambda}$ is nonpositive as $0 < \Lambda < 1$.
   B. When $\Lambda > 1$, there is no asymmetric interior equilibrium, and the only stable equilibria are such that $\Delta M = \pm 1$ because either $\Delta\pi(1) > 0$ or $\Delta\pi(-1) > 0$.

Q.E.D.

The stable equilibria are represented by the bold lines and the unstable ones by the dotted lines in Figure 8.4.

Despite the symmetry of the economy, *industrial clusters involving different numbers of firms may emerge in equilibrium.* In this case, the number of

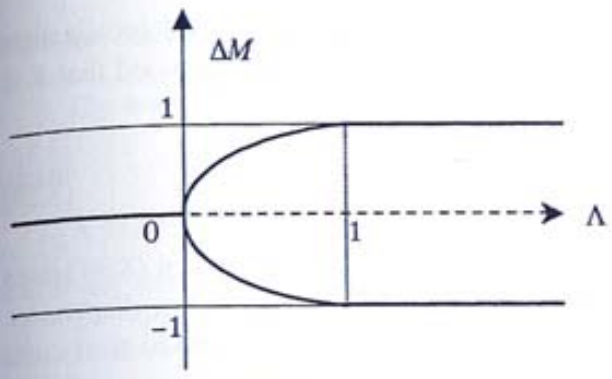


Figure 8.4: Stable equilibria.

firms in region r is

$$M_r^* = \frac{1 \pm \sqrt{\Lambda}}{2}.$$

Of course, the region ending up with the larger number of firms is the one that has the larger initial share of firms regardless of how small the difference is. This shows again that history (here the initial conditions) matters for the geographical distribution of production. In other words, the existence of localization economies may lead to the emergence of a polarized space, especially when transport costs are low.

Because ΔM depends only upon Λ, the impact of each parameter on the spatial equilibrium can be analyzed through Λ given by (8.33). First, as the intensity of preference for the output of the industry (a) rises, the degree of asymmetry between the two clusters grows. Second, we have

$$\frac{\partial \Lambda}{\partial t} = -\frac{2(b\theta + dt)}{d\theta^2} < 0 \tag{8.35}$$

$$\frac{\partial \Lambda}{\partial d} = -\frac{\Lambda}{d} - \frac{t^2}{d\theta^2} < 0 \tag{8.36}$$

$$\frac{\partial \Lambda}{\partial \theta} = -\frac{d\theta^2\Lambda + 2\theta^2 b - t^2 d}{d\theta^3} = -\frac{\Lambda}{\theta} + 2\frac{b}{d\theta} + \frac{t^2}{\theta^3}. \tag{8.37}$$

Equation (8.35) shows that a decrease in transport cost leads to more asymmetry between clusters. This finding reveals that lower transportation costs are likely to drive the economy toward more agglomeration in one region at the expense of the other. It is worth pointing out that this process is smooth unlike what is observed in the core–periphery model we will study in the next chapter. Similarly, as shown by (8.36), more product differentiation leads to more agglomeration of firms within the large region. As seen in Chapter 7, this is a standard result in many spatial competition models.

Equation (8.37) shows that an increase in the intensity of localization economies strengthens the tendency toward agglomeration provided that θ is small enough, that is,

$$\theta < \frac{2dt^2}{4(a-bc)-bt}, \tag{8.38}$$

where the right-hand side is positive by (8.28). In contrast, when (8.38) 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 cluster price in the inelastic part of their demands, whereas firms in the large cluster price in the elastic part. By reducing production costs, an increase in θ thus intensifies competition much more in the large cluster than in the small one, leading some firms to move from the large to the small cluster. In sum, localization economies do not have a monotone impact on the firm distribution.

8.4.3 Welfare

The research strategy is similar to the one used above. To begin with, we assume that the planner is able to impose prices equal to marginal costs as well as to choose firms' locations. Because firms earn zero profits, the only relevant welfare measure is the global consumer surplus $C_G = \sum_{s=A,B} C_s$. It can readily be verified that the consumer surplus is given by the following expression:

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

and thus

$$\begin{aligned} C_s &= \frac{1}{2(b+d)}\left[(q_{ss}^*)^2 N_s + (q_{rs}^*)^2 N_r\right] \\ &\quad + \frac{d}{2b(b+d)}\left[N_s^2(q_{ss}^*)^2 + N_r^2(q_{rs}^*)^2 + 2(N_s q_{ss}^*)(N_r q_{rs}^*)\right], \end{aligned} \tag{8.39}$$

where the first subscript denotes the region in which consumption occurs and the second the region in which the varieties are produced ($r, s = A, B$ and $r \neq s$).

After having replaced prices and quantities by their first best values, we obtain

$$C_G = -\frac{A}{4}(\Delta M)^4 + \frac{A}{2}\Gamma(\Delta M)^2 + B,$$

where

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

and B are constants and

$$\Gamma \equiv \frac{(3b+d)\theta^2 + 4\theta(a - bc - bt/2) - dt^2}{2d\theta^2} \tag{8.40}$$
$$= \frac{b+d}{2d} + \frac{\Lambda}{2}.$$

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

Proposition 8.7 *The first best allocation involves*

1. *full dispersion if and only if* $\Gamma \leq 0$;
2. *partial agglomeration if and only if* $0 < \Gamma < 1$;
3. *full agglomeration if and only if* $1 \leq \Gamma$.

Proof

1. We seek $\Delta M \in [-1, 1]$, maximizing C_G. When the maximum is interior, we have

$$\frac{\partial C_G}{\partial(\Delta M)} = -A\Delta M[(\Delta M)^2 - \Gamma] = 0 \tag{8.41}$$

and

$$\frac{\partial^2 C_G}{\partial(\Delta M)^2} = -A[3(\Delta M)^2 - \Gamma] < 0.$$

When $\Gamma \leq 0$, there is a unique maximizer given by $\Delta M = 0$ because the second-order condition is satisfied.

2. When $0 < \Gamma < 1$, C_G reaches a minimum at $\Delta M = 0$ and is maximized at $\Delta M = \pm\sqrt{\Gamma}$, where the second-order condition holds.
3. Finally, when $1 \geq \Gamma$, the welfare function C_G is maximized at $\Delta M = \pm 1$. Q.E.D.

Thus, the first best solution displays a pattern similar to that arising when firms are free to choose prices and locations at the market equilibrium. In the case of asymmetric clusters, the socially optimal distribution of firms (M_A^0, M_B^0) is given by

$$M_A^0 = \frac{1 \pm \sqrt{\Gamma}}{2}.$$

It is worth noting that Γ is a strictly decreasing function of t and d as shown by differentiating (8.40) with respect to t and d,

$$\frac{\partial \Gamma}{\partial t} = \frac{1}{2}\frac{\partial \Lambda}{\partial t} < 0 \tag{8.42}$$

$$\frac{\partial \Gamma}{\partial d} = -\frac{b}{2d^2} + \frac{1}{2}\frac{\partial \Lambda}{\partial d} < 0, \tag{8.43}$$

and thus *lower transport costs, more product differentiation, or both yield more asymmetry between optimal clusters.* From (8.40), it can readily be verified that the impact of a or θ on Γ is similar to the impact on Λ, as discussed in Section 8.4.2.

Furthermore, it follows from (8.40) that $\Gamma > 0$ when $\Lambda = 0$ and that $\Gamma = \Lambda$ for a value of Λ that exceeds 1. Consequently, we have the following result:

Proposition 8.8 *The first best optimum never involves a lower degree of agglomeration than the market equilibrium.*

Thus, as in the duopoly case, we observe that, with a large group of firms, the large cluster never involves too many firms in equilibrium from the efficiency viewpoint. In particular, the planner sets up more asymmetric clusters than arise at the market solution unless the market outcome corresponds to full agglomeration. This requires some explanation. At the optimum, prices are set at the marginal cost level whereas locations are chosen to maximize the difference between the benefits of agglomeration and total transport costs. By contrast, at the market equilibrium, firms take advantage of their spatial separation to relax price competition and, hence, to make higher profits. These two effects combine to generate the discrepancy between the market and optimal solutions. This confirms what we saw in Section 7.3, that is, price competition is a strongly dispersion force.

Unless full 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 on global efficiency and not on interregional equity. This policy is a sensible one 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.

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 the planner's purpose is to maximize the global surplus. The

consumer surplus C_s in region $s = A, B$ is computed by plugging the market equilibrium quantities into expression (8.39). The producer surplus Π_r in region $r = A, B$ is given by the sum of profits made in r and s by the firms located in r:

$$\Pi_r = \frac{M_r}{b+d}\left[(q_{rr}^*)^2 + (q_{rs}^*)^2\right].$$

To determine the second best, it is therefore sufficient to maximize the following welfare function:

$$W_G = C_A + C_B + \Pi_A + \Pi_B.$$

The first-order condition yields a cubic function with properties similar to those obtained in the first best when Γ is replaced by Θ as given by

$$\Theta \equiv \frac{4(3b+d)\theta[4a - 2b(2c+t-\theta)] + 4b(3b+d)\theta^2 - d(8b+3d)(t^2-\theta^2)}{2d\theta^2(8b+3d)}$$

$$= \frac{3(2b+d)^2 + d(4b+d)(t/\theta)^2}{2d(8b+3d)} + \frac{2(3b+d)}{8b+3d}\Lambda.$$

Consequently, the second-best allocation involves (1) identical clusters if and only if $\Theta \leq 0$; (2) asymmetric clusters if and only if $0 < \Theta < 1$; (3) a single cluster if and only if $1 \leq \Theta$.

The comparison between the second best and the market outcomes is very similar to that made above provided that Λ and Θ move in the same direction. In particular, it can readily be verified that Proposition 8.7 still holds, that is, $\Theta > \Lambda$ because $\Theta > 0$ when $\Lambda = 0$, whereas $\Lambda = \Theta$ for a value of Λ 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

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

which is negative, for $4(a-bc) > 2bt$ by (8.28). This implies that 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. Hence, we have the following:

Proposition 8.9 *When there is partial agglomeration, the second best involves more asymmetry between clusters than the first best.*

This surprising result is to be understood as follows. The planner's objective is now to damper too high prices by controlling locations, especially when varieties are very differentiated. To achieve this goal, the planner chooses to expand the cluster in the larger region because, in so doing, price competition is

intensified and localization economies are made stronger. This confirms once more that price competition is a strongly dispersion force.

The results obtained in this section confirm and extend those derived in the previous section. In particular, we have shown that asymmetric clusters are likely to emerge when transport costs become sufficiently low. Moreover, forcing a more equalitarian distribution of firms is likely to increase the inefficiency of the economy because price competition acts as a fairly strong dispersion force. Such results run against the conventional wisdom and point to a potential conflict between efficiency and spatial equity in the spatial distribution of production.

8.5 CONCLUDING REMARKS

The different models studied in this chapter allow us to shed some light on the main principles governing the spatial organization of a multiregional economy. In this perspective, one of the most striking results is that *a small initial advantage may lead to the emergence of a strongly polarized space* once we explicitly account for the existence of localized production externalities, natural amenities, or both. This effect is magnified when the mobility of factors or the transportability of products are high, or both. Indeed, either of the first two possibilities allows the localized externalities to display their full impact.

In addition, although they are active at the local level only, localization economies potentially appear to be a strong agglomeration force in shaping the interregional landscape (think of the Silicon Valley). This is especially true when transport costs between regions are low, varieties are very differentiated, or both. In these cases, agglomeration occurs because firms are able to enjoy a higher level of localization economies while they are still able to sell a substantial fraction of their output on distant markets. As in Chapters 6 and 7, in which we dealt with urban landscapes, low transport costs yield more agglomeration, but now this occurs at the interregional level. Our analysis has also identified a reason explaining how and when a firm may benefit from the presence of local competitors. Another reason will be discussed in the next chapter.

That more product differentiation fosters more agglomeration is not new. This fact agrees with what we have seen in Chapter 7 as well as with what we will see in the next chapter. The reason is always the same: a higher degree of product differentiation allows firms to relax price competition, thus permitting any existing agglomeration force to dominate the dispersion force. This force appears to be crucial in very different spatial settings and is likely to be strong in modern economies. Thus, the same causes lead to the same effects, although the geographical scale as well as the forces at work are very different. Interestingly, the same tendency holds at the first best optimum.

When the desirability for the output of the industry rises, more firms tend to locate within the same cluster, whose relative size increases at the expense of the

other firms. Consequently, economic growth, as measured by the relative importance of the differentiated good to the numéraire, tends to foster more geographical concentration of production.[17] It is worth pointing out that such an increase in the agglomeration of firms arises even though the spatial distribution of demand remains unchanged. However, when more firms enter as a result of growth, the impact on agglomeration depends on the many features of the economy.

Unlike what many regional analysts and planners would argue, the optimal configuration tends to involve a more unbalanced distribution of firms than the market outcome, both in the duopoly and large group cases. If localization economies become increasingly 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 signify a wasteful allocation of resources. On the contrary, the size of the existing clusters could well be too small.

In this chapter, we have stressed the role of nonmarket interactions in the formation of industrial clusters. Without such technological externalities ($E(H_r) = 1$ in 8.2, $\theta_r = 0$ in 8.3, and $\theta = 0$ in 8.4), the models studied would always have generated the same outcome, namely, dispersion. Therefore, for the settings considered in this chapter, these externalities have been critical in showing the emergence of industrial agglomerations.

In the next chapter, we will move to the study of pecuniary externalities. Several reasons explain this alternative research strategy. First, even though technological and pecuniary externalities are natural components of any complete explanation of economic clustering, one major intellectual advantage is presented by the latter. Because they rely on market interactions, the origin of pecuniary externalities is clearer. In contrast, as discussed in Chapter 1, technological externalities often have the nature of a black box, thus leading to results that may have a vague meaning. In addition, although only technological externalities affect efficiency in an otherwise perfect market, with imperfect competition pecuniary externalities also matter in that changes in prices affect the deadweight loss due to existing distortions. Hence, both types of externalities matter when evaluating the efficiency of the market outcome. However, the inefficiencies associated with pecuniary externalities can be better understood because their origin is easier to identify. Last, it should be kept in mind that modeling Marshallian external effects by means of a technological externality may lead to inappropriate policy recommendations because the "proxy" is too rough (Abel-Rahman and Fujita 1990).

APPENDIX

A. We study the examples presented in Sections 8.2.1 and 8.2.2 in which we have $E(H_r) = \exp(\varepsilon H_r)$, $f(H_r) = H_r^a$, $u(w_r^h) = \log w_r^h$ and $e(P_r) = aP_r - b(P_r/S)^2$ for $r = A, B$. Because the model of Section 8.2.1 is a special case of

the one of 8.2.2 in which both a and b are set equal to zero, it is sufficient to consider the general case.

Using the expressions above, we can readily verify that

$$\dot{H}_A = \psi(H_A) = \log\left(w_A^H / w_B^H\right) + e(H_A + 1) - e(H - H_A + 1).$$

Because $w_r^H = \exp(\varepsilon H_r) f'(H_r) = \alpha \exp(\varepsilon H_r) H_r^{\alpha-1}$ for $r = A, B$, we have

$$\begin{aligned}\psi(H_A) = {} & \varepsilon(2H_A - H) - (1-\alpha)\log[H_A/(H - H_A)] \\ & + e(H_A + 1) - e(H - H_A + 1).\end{aligned}$$

Differentiating this expression with respect to H_A yields

$$\begin{aligned}\psi'(H_A) = {} & 2(\varepsilon + a) - (1-\alpha)/H_A - (1-\alpha)/(H - H_A) \\ & - 2b(H+2)/S^2,\end{aligned} \tag{A.1}$$

whose derivative is

$$\psi''(H_A) = (1-\alpha)/H_A^2 - (1-\alpha)/(H - H_A)^2.$$

Because $\psi'''(H_A)$ is strictly negative for $0 < H_A < H$, the function $\psi''(H_A)$ is strictly decreasing on $(0, H)$. Moreover, we have $\psi''(0) = \infty$ and $\psi''(H) = -\infty$. Therefore, there is a unique value $H_A^* = H/2$ such that $\psi''(H_A) = 0$. In other words, H_A^* is the unique maximizer of $\psi'(H_A)$ over $[0, H]$.

Two cases may then arise:

1. $\psi'(H_A^*) \leq 0$. This implies that we must have $\psi'(H_A) \leq 0$ for all values of H_A in $[0, H]$. This in turn implies that ψ is decreasing over the interval $(0, H)$. Because $\psi(0) = \infty$ and $\psi(H) = -\infty$, there is a unique equilibrium given by $H/2$, and this equilibrium is globally stable because $\psi'(H_A) < 0$ in a neighborhood of H_A^* (see Figure 8.1).
2. $\psi'(H_A^*) > 0$. Because $\psi'(H_A)$ is concave, two values H_A^- and H_A^+ exist for which $\psi'(H_A) = 0$. As a result, $\psi(H_A)$ is strictly decreasing over $[0, H_A^-)$, strictly increasing over (H_A^-, H_A^+), and strictly decreasing over $(H_A^+, H]$. Consequently, three equilibria exist if and only if $\psi(H_A^-) < \psi(H/2) < \psi(H_A^+)$ (see Figure 8.2).

Because regions have identical endowments, (A.1) becomes

$$\psi'(H_A^*) = \psi'(H/2) = 2(\varepsilon + a) - 4(1-\alpha)/H - 2b(H+2)/S^2,$$

and thus the condition for the existence of three equilibria $\psi'(H_A^*) > 0$ is equivalent to

$$\varepsilon + a - b(H+2)/S^2 > 2(1-\alpha)/H,$$

which is exactly (8.13). When this inequality is reversed, there is a single equilibrium because $\psi'(H_A^*) \leq 0$.

B. In the case of two varieties, we know that the quadratic utility is given by (8.14):

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

In the case of $n > 2$ varieties, (8.14) may be extended as follows:

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

Letting $n \to \infty$ and $q_i \to 0$, we then obtain (8.25) in which M stands for the mass of varieties.

NOTES

1. Not surprisingly, Marshallian externalities are the engine of economic development in the new theories of growth (Romer 1986; Lucas 1988).
2. Remember that Chapter 6 focused on the informational spillover interpretation only.
3. For recent contributions, see Glaeser et al. (1992); Henderson, Kuncoro, and Turner (1995); Ciccone and Hall (1996); Hanson (1996); Feldman and Audretsch (1999); and Smith (1999). Eberts and McMillen (1999) provide a regional science-oriented survey. In the same vein, Lucas (2001) has argued that local interactions are a basic component of the engine of growth.
4. According to Prescott (1998), each place is characterized by a "social capital" affecting the overall productivity; this capital will vary from place to place.
5. The content of that section is also related to the brain drain problem (see, e.g., Miyagiwa 1991), but we do not address this issue here.
6. Note also that Black (2000) has argued in a convincing way that, in the presence of mobility costs, skilled workers have incentives to move before unskilled workers do.
7. The neoclassical model based on wage differentials has been revisited by Rosen (1979) to account for the importance of regional amenities in the decision to migrate. Subsequently, his analysis was extended by Robak (1982), who added firms to the model to permit individuals to be both consumers and workers.

8. The role of these two types of externalities has been emphasized in the new growth theories and is at the heart of urban economics.
9. Assuming that both 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 (8.15) 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 keeping the other structural parameters a and b of the demand system unchanged.
10. Given (8.19), arbitrage is never profitable.
11. If $\theta_A = \theta_B$, the second case never arises, and there are only two equilibria.
12. The results presented in this section have been extended to the case of several firms and districts by Soubeyran and Weber (2002), considering a Cournot framework and an integrated market.
13. This agrees with the well-known idea that firms may be locked into an inferior technology (Arthur 1994, chap. 2).
14. A similar conflict appears in a different but related context in which the relocation of firms is triggered by a wage differential (Cordella and Grilo 2001).
15. See Part B of the chapter appendix for a justification of the coefficients used in that utility function.
16. The intuition behind this interpretation is very similar to the one that stands behind the Herfindahl index used to measure industrial concentration. If we control for the total amount of the differentiated good consumption, the absolute value of the quadratic term in (8.25) increases with the concentration of consumption on a few varieties, thus decreasing utility.
17. Compare (8.15) and (8.27).
18. Similarly, we have $\alpha = a/b$, $\beta = 1/(b + d)$, and $\delta = d/b(b + d)$. Hence, there is a one-to-one correspondence between the two sets of parameters.
19. This agrees with the analysis of the interplay between agglomeration and growth developed by Martin and Ottaviano (1999).