The spatial selection of heterogeneous firms*

Toshihiro Okubo †   Pierre M. Picard ‡   Jacques-François Thisse§

Abstract

The aim of this paper is to study the spatial selection of firms once it is recognized that heterogeneous firms typically choose different locations in respond to market integration of regions having different sizes. Specifically, we show that decreasing trade costs leads to the gradual agglomeration of efficient firms in the large region because these firms are able to survive in a more competitive environment. In contrast, high-cost firms seek protection against competition from the efficient firms by establishing themselves in the small region. However, when the spatial separation of markets ceases to be a sufficient protection against competition from the low-cost firms, high-cost firms also choose to set up in the larger market where they have access to a bigger pool of consumers. This leads to the following prediction: the relationship between economic integration and interregional productivity differences first increases and then decreases with market integration.

Keywords: firm heterogeneity; spatial selection; trade liberalization

JEL Classification: F12; H22; H87; R12

---

*We thank Y. Kanemoto, D. Sasaki and T. Tabuchi for helpful comments and suggestions.

†RIEB, Kobe University (Japan). e-mail: okubo@rieb.kobe-u.ac.jp

‡University of Manchester (United Kingdom) and CORE, Université catholique de Louvain (Belgium).

§CORE, Université catholique de Louvain (Belgium), PSE (France) and CEPR. e-mail: jacques.thisse@uclouvain.be

1 Introduction

Firms are heterogeneous in many respects and, therefore, choose different strategies.[1] For instance, it is now well documented that more productive firms tend to export in several countries while less efficient firms tend to focus on their local markets only.[2] As market integration gets deeper, this discrepancy in export strategies amplifies because lower trade costs intensify product market competition and trigger a *selection effect* of firms located in each country (Melitz, 2003; Helpman et al., 2004; Melitz and Ottaviano, 2008; Okubo, 2008). However, firms' selection may also take place across space. Because lower trade costs weaken the traditional protection of geographical separation, one expects that more efficient firms tend to set up in larger markets where they can better exploit scale economies, whereas less efficient firms would relax competition by seeking protection in smaller markets. In other words, market integration may not only select the best firms within countries but it may also entice them to sort out across countries. Consequently, *the observed productivity differences between large and small markets could stem from the spatial selection of heterogeneous firms rather than from their selection within countries.* And indeed, some recent empirical studies devoted to firm heterogeneity and location choice point to that direction. Using US concrete industry data, Syverson (2004) observes that inefficient firms barely survive in large competitive markets and tend to leave them. This result is confirmed by the emerging literature that follows Syverson (Asplund and Nocke, 2006; Del Gatto et al., 2006; Syverson, 2007; Del Gatto et al., 2008; Foster et al.,

[1] Business analysts and industrial economists have put forth various explanations for the existence of persistent cost differences among competing firms (Porter, 1982; Röller and Sinclair-Desagné, 1998). More generally, Tirole's (1988) *Principle of Differentiation* suggests that firms have strategic incentives to be different. Hence, the existence of heterogenous firms should not come as a surprise.

[2] The empirical evidence is mounting (Bernard and Jensen, 1995, 1999, 2004; Bernard et al., 2003; Aw et al., 2000, Eaton et al., 2004; Tybout and Westbrook, 1995; Pavcnik, 2002; Ederington and McCalman, 2008).

2008).[3]

The aim of this paper is to study the spatial selection of firms once it is recognized that heterogeneous firms typically choose different locations in respond to market integration. In other words, we focus on the specific location choices made by heterogeneous firms, the reason being that trade liberalization has a strong impact on the intensity of competition across spatially separated markets, hence on the ability of firms to operate in large or small, close or distant, markets. One of the main features of our approach is that it allows us to fully capture the fact that, everything else being equal, high-cost firms tend to move away from low-cost firms to soften competition. Our emphasis is thus on sorting across locations. In the trade literature with heterogeneous firms, the selection of firms takes place through their entry and exit in the local markets but not through their relocation between markets. The focus is on firms' export strategy (local producer or exporter) and on the impact of trade liberalization on the average productivity of each trade partner. In contrast, our model as well as Baldwin and Okubo (2006) are more in line with economic geography, which highlights the impact of trade liberalization on firms' location choices and the emergence of endogeneous comparative advantage that determine the productivity of countries. By starting from a perspective different from that of Melitz (2003), Helpman et al. (2004) and others, our paper sheds new light on the firm selection problem, while contributing to the economic geography literature.

To achieve our goal, we build on Martin and Rogers (1995) who investigate how trade costs affect the spatial distribution of firms which are otherwise homogeneous. However, in order to better account for the impact of competition on firms' mark-ups and profits, we use the linear model of monopolistic competition developed by Ottaviano et al. (2002) and introduce cost heterogeneity across firms. Unlike the Dixit-Stiglitz model of monopolistic

[3] It is worth noting that another strand of literature, rooted in economic geography, confirms the fact that firms tend to be more productive in large markets (Head and Mayer, 2004; Redding and Venables, 2004; Redding and Sturm, 2008).

competition, the linear model captures the pro-competitive effect generated by decreasing mark-ups in larger markets and allows us to account for the impact that price competition has on the way heterogeneous firms distribute themselves across space. Our setting also allows for a full analytical description of the equilibrium firm distributions as well as a precise and detailed analysis of the impact of all structural parameters on firms' behavior. In particular, we show how *location choices act as a selection device across heterogeneous firms by determining the number and types of firms in each country.*

First of all, we show that (in)efficient firms cluster whereas the other group of firms choose distinct locations, depending on the heterogeneity parameters of the industry as well as on the level of trade costs. In this respect, it is worth stressing that Duranton and Overman (2005) have developed a new empirical method to study the main features of the spatial distribution of manufacturing firms, which accounts explicity for inter-firm distances. One of their main findings is that industries are highly heterogeneous in their location behavior: in some industries, larger plants tend to be clustered and smaller ones dispersed, whereas the opposite holds in other industries. Our analysis provides a theoretical foundation to this observation since, everything else being equal, the within-industry heterogeneity yields either the spatial concentration of large (efficient) firms together with the dispersion of small (inefficient) firms or the opposite pattern.

Specifically, we prove that decreasing trade costs lead to the gradual agglomeration of low-cost firms in the larger market because these firms are able to survive in a more competitive environment. In contrast, high-cost firms seek protection against competition from the low-cost firms by establishing themselves in the smaller country.[4] For some domain of trade cost values, perfect selection even arises with all low-cost firms being established in the large country and all high-cost firms in the small country. This two-way process of relocation triggers productivity differences between regions and countries.

[4] It is worth noting here that Vogel (2008) shows, in a spatial competition setting with heterogeneous firms, that efficient firms repel inefficient firms because they are able to build a larger market basis.

Note that such disparities are also exacerbated by growing asymmetries between market sizes. This is not the end of the story, however. As the global economy gets more and more integrated, the selection effect is turned up-side down, the market access effect stressed by economic geography becoming the dominant force. More precisely, as the spatial separation of markets ceases to be a sufficient protection against competition from the low-cost firms, high-cost firms also choose to set up in the larger market where they have access to a bigger pool of consumers. Our setting thus leads to the following prediction: *international productivity differences first increases and then decreases as market integration deepens*. This result is to be contrasted with Baldwin and Okubo (2006), who combine Martin and Rogers (1995) with Melitz (2003), and show that firms' response to decreasing trade costs is a one-way process in which only the low-cost firms established in the small country move to the large one.

Furthermore, contrary to general beliefs, the impact of heterogeneity on the number and types of firms in each country is not governed by a simple monotone relationship. In particular, firm heterogeneity may act as an agglomeration force. Specifically, when the share of inefficient firms is small, a higher degree of heterogeneity strengthens the agglomeration of both types of firms in the larger market. Indeed, this market is less competitive because it hosts a smaller number of efficient firms. On the other hand, when there is a large share of inefficient firms, increasing their number fosters their dispersion, the reason being that they now aim to relax competition within their own group. Finally, it is well known that only a handful of firms export in the real world (Mayer and Ottaviano, 2008). Therefore, it is interesting to isolate the effect of trade from the effect of firm heterogeneity on spatial sorting. To this end, we discuss the case in which all varieties are non-tradeable (Nocke, 2006). By focussing on this polar case, we avoid the technical difficulties associated with the existence of asymmetries in international market access (Behrens, 2005). We find that spatial equilibria have a structure very similar to the one with tradeable goods. However, a higher heterogeneity between firms yields different implications about firm location.

The remainder of the paper is organized as follows. In the next section, we present the linear footloose capital model. Section 3 thus investigates the equilibrium distribution of firms in the case of tradeable goods for different levels of trade costs and various measures of heterogeneity. With the aim of testing the robustness of our results, we study the case of non-tradeable goods in Section 4. The last section presents our conclusions and discusses some policy implications.

2 The model

Consider a world with two countries, labeled 1 and 2, and a unit total mass of consumers.[5] Each consumer is endowed with one unit of labor and one unit of capital, both of which are supplied inelastically.[6] Let $\lambda \in [1/2, 1)$ denote the share (and mass) of consumers in country 1, which implies that λ also measures that country's shares (and masses) of labor and capital. Consumers are immobile and can supply labor only in the country where they reside. In contrast, they are free to supply capital wherever they want. The spatial distribution of capital is endogenous and will be determined as an equilibrium outcome. To disentangle the various channels through which this happens, we will distinguish between a short-run equilibrium when firms are immobile, and a long-run equilibrium when they are mobile.

2.1 Technologies and trade costs

The economy involves two tradeable goods. The first good Z is produced under constant returns to scale and perfect competition by using labor as the only input. Without

[5]As long as capital is mobile across space, our analysis equally applies to the interregional and international levels .

[6]Because income effects are absent in our model, the distribution of capital ownership does not matter for our results.

loss of generality, we normalize the unit input requirement to 1. We assume that the homogeneous good is traded costlessly between the two countries.[7] Profit maximization in the homogeneous good sector then implies that $p_i^Z = w_i \equiv p^Z$ in both countries $i = 1, 2$. Note that such factor price equalization always holds when each country has enough labor to support some production of Z for any international allocation of capital. In our setting, we follow the literature and assume that both countries produce the two goods. We may then choose good Z as the numéraire, which implies $w_i = p^Z = 1$.

The second good is made available as a continuum of horizontally differentiated varieties, indexed by v. Each variety is produced by a single firm under increasing returns and monopolistic competition. A firm is run by an entrepreneur who rents one unit of capital (after normalization) and hires a quantity of labor proportional to its output. Whereas capital is homogeneous (say cash), *entrepreneurship is heterogeneous.*[8] Specifically, there are two types of entrepreneurs or firms:[9] *high-cost* entrepreneurs (H) requires $m > 0$ units of labor to produce one unit of the differentiated good, while *low-cost* entrepreneurs (L) require a lower amount of labor units, which we normalize to zero without loss of generality because firms' demands are linear (see Ottaviano *et al.* 2002). Thus, if r denotes the rental rate of capital and q the volume of production, high-cost and low-cost entrepreneurs incur the following cost functions:

$$C_H(q) = r + mq \qquad C_L(q) = r.$$

Let the share of high-cost and low-cost entrepreneurs in the economy be denoted by $(s, 1-s)$ with $0 < s < 1$. In our setting, firm heterogeneity is thus captured through two different parameters, m and s. As will be seen, they may have different impacts on the international distribution of firms.

[7] This assumption is standard in most trade and economic geography models. Introducing positive trade costs for the homogeneous good makes the analysis much more involved (Picard and Zeng, 2005).

[8] Note that the above heterogeneous managerial input can be replaced by any factor that reflects firms' specific technological advantage.

[9] The impact of the assumption is discussed in the concluding section.

Capital is inelastically supplied in an integrated financial market that encompasses both countries. Denote by N_i the share of capital invested in country i with $N_1 + N_2 = 1$. In the short run, (i) the international distribution of capital (N_1, N_2) and (ii) the shares of high-cost and low-cost firms in each country are all fixed. The share of θ-firm in country 1 is denoted by $0 \leq n_\theta \leq 1$ for $\theta = H, L$. Hence, for a given distribution of firms across countries and types, the global number of firms N_i in country $i = 1, 2$ is given by

$$N_1 = sn_H + (1-s)n_L \qquad N_2 = s(1-n_H) + (1-s)(1-n_L). \tag{1}$$

It is readily verified that $N_1 + N_2 = 1$.

Finally, international shipments of any variety of the differentiated good incur a trade cost of $t > 0$ units of the numéraire per unit of variety shipped.

2.2 Preferences and demands

All consumers share the same preferences over the homogeneous good and the horizontally differentiated good. Preferences are given by a quasi-linear utility function with a quadratic subutility. More precisely, a consumer of country $i = 1, 2$ solves the following problem:

$$\max_{q_{ii}(v),\, Z_i} \alpha \int_0^{N_i} q_i(v)dv - \frac{\beta - \gamma}{2} \int_0^{N_i} [q_i(v)]^2 dv \\ - \frac{\gamma}{2} \left[\int_0^{N_i} q_i(v)dv \right]^2 + Z_i$$

subject to the budget constraint

$$\int_0^{N_i} p_i(v)q_i(v)dv + Z_i = r + w_i + \overline{Z}_0$$

where N_i is the number of varieties *available* in country i, $q_i(v)$ and $p_i(v)$ are the consumption and the consumer price of variety v, Z_i is the consumption of the homogeneous good, r is the rental rate of capital, and $w_i = 1$. Note that consumers receive a rental rate that varies with the country and the firm-type in which they invest. However, because quasi-linear preferences do not capture income effects, the spatial distribution of rental rates does not affect the demands for differentiated varieties; the spatial distribution of rental rates affect neither firms' profits nor their location choice. Finally, $\overline{Z}_0 > 0$ is the initial endowment of the homogeneous good, which we assume to be sufficiently large for consumption of this good to be strictly positive at the market outcome. This assumption is made to capture the idea that consumers like to consume the two goods. As to the parameters, $\alpha > 0$ expresses the intensity of preference for the differentiated product, $\gamma > 0$ measures the substitutability between varieties, whereas $\beta - \gamma > 0$ expresses the intensity of the preference for variety, i.e. the consumers' bias toward a dispersed consumption of varieties.

The individual demand is

$$q_i(v) = a_i - (b_i + c_i N_i) p_i(v) + c_i P_i$$

where a_i, b_i and c_i are positive variables given by

$$a_i \equiv \frac{\alpha}{\beta + (N_i - 1)\gamma} \qquad b_i \equiv \frac{1}{\beta + (N_i - 1)\gamma} \qquad c_i \equiv \frac{\gamma}{(\beta - \gamma)\,(\beta + (N_i - 1)\gamma)}. \tag{2}$$

Setting

$$a = \frac{\alpha}{\beta} \qquad b = \frac{1}{\beta} \qquad c = \frac{\gamma}{(\beta - \gamma)\beta}$$

and inverting these expressions yields

$$\alpha = \frac{a}{b} \qquad \beta = \frac{1}{b} \qquad \gamma = \frac{c}{b\,(b + c)}$$

we have

$$a_i \equiv a\frac{b + c}{b + cN_i} \qquad b_i \equiv b\frac{b + c}{b + cN_i} \qquad c_i \equiv c\frac{b + c}{b + cN_i}.$$

Assume that trade costs are low enough for the differentiated good to be tradeable. In this case, we have $N_i = 1$ since all varieties are available everywhere. This implies $a_i = a$, $b_i = b$ and $c_i = c$. This assumption will be relaxed in the model with non-tradeable goods discussed in Section 4. As suggested by several empirical studies, product markets are segmented (Engel and Rogers, 1996; Wolf, 2000; Haskel and Wolf, 2001). In other words, each firm is free to set a consumer price specific to the country in which it sells its variety. Let us then denote by $p_{ki}^{\theta}(v)$ the consumer price of variety v set in country i by a θ-firm located in country $k = 1, 2$, and by $q_{ki}^{\theta}(v)$ the corresponding individual demand. More precisely, the individual demands in countries 1 and 2 for a variety produced in country 1 become:

$$\begin{aligned} q_{11}^{\theta}(v) &= a - (b + c)p_{11}^{\theta}(v) + cP_1 \\ q_{12}^{\theta}(v) &= a - (b + c)p_{12}^{\theta}(v) + cP_2 \end{aligned} \tag{3}$$

where

$$P_1 \equiv sn_H p_{11}^H + (1 - s)n_L p_{11}^L + s(1 - n_H)p_{21}^H + (1 - s)(1 - n_L)p_{21}^L \tag{4}$$

is the *consumer price index* of the differentiated good in country 1. Mirror expressions hold for country 2. The demands (3) reveal a common feature of monopolistically competitive models: each firms treats parametrically the price index P_i which represents an inverse measure of the intensity of competition in country i.

2.3 Prices and profits

In the monopolistically competitive sector, firms play a noncooperative game with a continuum of players, in which prices are the firms' strategies. In what follows, we provide all expressions for firms established in country 1. Since θ-firms located in the same country are symmetric, they charge the same prices and maximize their profits given by

$$\begin{aligned}\Pi_1^H &\equiv \pi_1^H - w_1^H - r_1^H \\ &= \lambda(p_{11}^H - m)q_{11}^H + (1-\lambda)(p_{12}^H - m - t)q_{12}^H - w_1^H - r_1^H \end{aligned} \tag{5}$$

and

$$\begin{aligned}\Pi_1^L &\equiv \pi_1^L - w_1^L - r_1^L \\ &= \lambda p_{11}^L q_{11}^H + (1-\lambda)(p_{12}^L - t)q_{12}^L - w_1^L - r_1^L \end{aligned} \tag{6}$$

where r_1^θ and w_1^θ are, respectively, the rental rate of capital and the entrepreneur's compensation paid by a θ-firm located in country 1, while π_1^θ denotes the operating profits of a θ-firm located in country 1.

In the short-run equilibrium, consumers maximize utility, firms maximize profits, and product and factor markets clear. A low-cost firm maximizes its profits by maximizing each term in (5) independently. Furthermore, because there is a continuum of firms, each firm is negligible and takes the consumer price index P_i, $i = 1, 2$, as given when choosing its optimal prices. Hence, maximizing each profit function (5) and (6) yields the following profit-maximizing prices:

$$p_{11}^L(P_1) = \frac{a + cP_1}{2(b+c)} \qquad p_{21}^L(P_1) = p_{11}^L(P_1) + \frac{t}{2} \tag{7}$$

$$p_{11}^H(P_1) = p_{11}^L(P_1) + \frac{m}{2} \qquad p_{21}^H(P_1) = p_{11}^L(P_1) + \frac{m+t}{2} \tag{8}$$

Those prices can be viewed as a high- or low-cost firm's reaction function to its local market conditions, given by the price index (4). The equilibrium value of the price index must be consistent with firms' pricing decisions. Substituting (8) and (7) into (4) and using (4) yields the following values:

$$\begin{aligned}P_1 &= \frac{a + (b+c)(sm + tN_2)}{2b+c} \\ P_2 &= \frac{a + (b+c)(sm + tN_1)}{2b+c}.\end{aligned}$$

Plugging back these expressions into (8) and (7) yields the (Nash) equilibrium *national* and *international* prices:

$$p_{11}^L = \frac{a}{2b+c} + \frac{c(ms+tN_2)}{2(2b+c)} \qquad p_{21}^L = p_{11}^L + \frac{t}{2}$$

and

$$p_{11}^H = p_{11}^L + \frac{m}{2} \qquad p_{21}^H = p_{11}^L + \frac{m+t}{2}. \tag{9}$$

The following comments are in order. First, *a larger share of low-cost firms leads to lower prices in the two countries because the two markets become more competitive.* Second, increasing trade costs makes the penetration of foreign varieties more difficult and leads to higher prices in both markets. In other words, trade costs act as competition barriers. Hence, the price charged by the domestic low-cost firms is the lowest one (p_{11}^L), whereas the highest one is the price set by the foreign high-cost firms (p_{21}^H). Third, any firm having to pay a higher cost to supply a market absorbs half of the corresponding cost wedge, m or t.

In the foregoing, we have implicitly assumed that both the marginal cost m and trade cost t are sufficiently low for all firms to export regardless of their distributions across countries and types. In what follows, we focus on the situation where trade always takes place. That is, a high-cost firm finds it profitable to export to the large market when all the other firms are low-cost and located in country 1 ($s = 0, N_1 = 1$). This yields the following *trade feasibility condition:*

$$t + m < t_{trade} \equiv \frac{2a}{2b+c}. \tag{10}$$

Finally, we assume perfect competition in the capital market. Free entry and exit implies that potential entrepreneurs compete for capital until rental rates exactly absorb firms' operating profits. In other words, the equilibrium rental rate corresponding to (1) is determined for θ-firms in country i by a bidding process for capital, which ends when no θ-firm can earn a strictly positive profit at the equilibrium market prices. Operating profits for each type of firm in country 1 are given by

$$\pi_1^H = (b+c)\left[\lambda(p_{11}^H - m)^2 + (1-\lambda)\left(p_{22}^H - \frac{t}{2} - m\right)^2\right]$$

$$\pi_1^L = (b+c)\left[\lambda(p_{11}^L)^2 + (1-\lambda)\left(p_{22}^H - \frac{t}{2}\right)^2\right]$$

where equilibrium prices are given by (9). It is readily verified that, for any given distribution of firms, *operating profits of low-cost firms are higher than those of high-cost firms*: $\pi_i^L > \pi_i^H$.[10]

3 Spatial selection as an equilibrium outcome

In the long run, capital and entrepreneurs are mobile, meaning that firms may change location, and enter or exit the market. The long-run equilibrium is thus determined by the capital-owners' and entrepreneurs' choices about the location of their investments and firms. We assume that the total supply of capital is perfectly inelastic and normalized to unity, and that the number of low-cost entrepreneurs is exogenously fixed to $1-s$; in contrast, the number of high-cost entrepreneurs is arbitrarily large. For the sake of simplicity, we assume that entrepreneurs have zero reservation wages and spend their income in the rest of the world.[11]

The amount of capital demanded by entrepreneurs depends on their location choice as well as on the compensation scheme they are offered. Let us fix the rental rate to r. Then, θ-entrepreneurs choose the location that maximize their compensation w^θ subject to their participation constraint $w^\theta \geq 0$. Clearly, entrepreneurs maximize their compensation

[10] Note that all the expressions derived above are identical to those obtained by Ottaviano *et al.* (2002) when $m = s = 0$.

[11] Though not unrealistic, this assumption is primarily made to highlight the effect of firms' heterogeneity while avoiding the complication due to the demand linkages generated by entrepreneurs. This issue has been studied by Picard *et al.* (2004) in the case of homogeneous firms.

by making their participation constraint binding ($w^\theta = \max_j \pi_j^\theta - r$) and by locating in the market that yields the highest operating profits ($i \in \max_j \pi_j^\theta$). Since $\pi_j^H < \pi_j^L$, we can readily characterize the demand for capital $D(r)$ in the following way. First, no entrepreneurs demand capital ($D(r) = 0$) if $r > \max_j \pi_j^L$; second, any subset of low-cost entrepreneurs demand capital ($D(r) \in [0, 1-s]$) if $r = \max_j \pi_j^L$; third, all low-cost entrepreneurs demand capital ($D(r) = 1-s$) if $\max_j \pi_j^H < r < \max_j \pi_j^L$; and, last, all low-cost entrepreneurs and any subset of high-cost entrepreneurs demand capital ($D(r) \in [1-s, \infty)$) if $r = \max_j \pi_j^H$.

The capital market equilibrium occurs when supply meets demand. Because the number of low-cost entrepreneurs is smaller than the supply of capital, the capital market clears at a rental rate equal to $r^* = \max_j \pi_j^H$. At this rate, there are $1-s$ low-cost firms and s high-cost firms. In the long-run equilibrium, high-cost entrepreneurs get zero compensation ($w^H = 0$), whereas low-cost entrepreneurs receive a compensation equal to $w^L = \max_j \pi_j^L - r^* > 0$. The distribution of firms is then readily determined as follows. High-cost firms do not establish themselves in a country that would yield a negative compensation to their entrepreneurs. Consequently, the share of high-cost firms in country 1 is given by

$$n_H \begin{cases} = 0 & \text{if} \quad \pi_1^H < \pi_2^H \\ \in [0,1] & \text{if} \quad \pi_1^H = \pi_2^H \\ = 1 & \text{if} \quad \pi_1^H > \pi_2^H. \end{cases}$$

Because they get a rent w^L that increases with their firms' profits, low-cost entrepreneurs choose to set up in the most profitable country. Hence, the share of low-cost firms in country 1 is given by

$$n_L \begin{cases} = 0 & \text{if} \quad \pi_1^L < \pi_2^L \\ \in [0,1] & \text{if} \quad \pi_1^L = \pi_2^L \\ = 1 & \text{if} \quad \pi_1^L > \pi_2^L \end{cases}$$

Therefore, the long-run equilibrium depends on profit differentials for each type of entrepreneurs/firms.

3.1 Perfect or partial selection

Since firms are heterogeneous, the same product market conditions have different impacts on firms' profits. We first show that low-cost firms are able to offer better rental rates in the large country. Indeed, the operating profit gap for high-cost firms between countries 1 and 2 is given by

$$\Delta\pi^H(n_H, n_L, \lambda) \equiv \pi_1^H - \pi_2^H = (b+c)t\left[\lambda\left(p_{11}^H - m - \frac{t}{4}\right) - (1-\lambda)\left(p_{22}^H - m - \frac{t}{4}\right)\right].$$

Plugging the equilibrium prices (9) into this expression, we obtain

$$\Delta\pi^H(n_H, n_L, \lambda) = \frac{t}{2}\left\{(2\lambda-1)\left[a-(b+c)\left(m+\frac{t}{2}\right)\right] + c[\lambda P_1 - (1-\lambda)P_2]\right\}. \quad (11)$$

Likewise, for the low-cost firms, we get

$$\Delta\pi^L(n_H, n_L, \lambda) = \frac{t}{2}\left\{(2\lambda-1)\left[a-(b+c)\frac{t}{2}\right] + c[\lambda P_1 - (1-\lambda)P_2]\right\}. \quad (12)$$

Clearly, the first term of $\Delta\pi^H$ is always strictly smaller than the first term of $\Delta\pi^L$, whereas the second is the same and given by

$$\lambda P_1 - (1-\lambda)P_2 = \frac{2\lambda-1}{2b+c}[a+(b+c)ms] + \frac{t(b+c)}{2b+c}\cdot I$$

where $I \equiv \lambda - [(1-s)n_L + sn_H]$. Consequently, it must be that

$$\Delta\pi^H(n_H, n_L, \lambda) < \Delta\pi^L(n_H, n_L, \lambda) \qquad \forall\lambda, n_H, n_L.$$

This implies that an interior configuration, i.e. $0 < n_H < 1$ and $0 < n_F < 1$, cannot be a long-run equilibrium. Stated differently, *at least one type of firms must be agglomerated at the long-run equilibrium.* So, we are left with five possible configurations: (i) co-agglomeration of all types of firms in the large market when $0 < \Delta\pi^H < \Delta\pi^L$ and thus $n_H^* = n_L^* = 1$; (ii) *partial selection* of high-cost firms in the large market when

$\Delta\pi^H = 0 < \Delta\pi^L$, so that $n_L^* = 1$ while n_H^* is an interior solution; (iii) *perfect selection* of low-cost and high-cost firms when $\Delta\pi^H < 0 < \Delta\pi^L$, so that $n_H^* = 0$ and $n_L^* = 1$; (iv) partial selection of low-cost firms in the small market when $\Delta\pi^H < 0 < \Delta\pi^L$, which implies that $n_H^* = 0$ and n_L^* is an interior solution; and finally (v) co-agglomeration in the small market when $\Delta\pi^H < \Delta\pi^L < 0$, meaning $n_H^* = n_L^* = 0$.

By studying the behavior of functions $\Delta\pi^H$ and $\Delta\pi^L$ with respect to n_H and n_L in $[0, 1]$, we can determine which configuration is a long-run equilibrium. Assume $s < 1/2$ and set

$$\mu_L \equiv \frac{c\,(1-2s)\,t}{4\,(2a-bt-cms)} < \mu_H \equiv \frac{c\,(1-2s)\,t}{4\,[2a-b\,(2m+t)-cm\,(1-s)]}$$
$$< \mu_A \equiv \frac{ct}{4\,[2a-b\,(2m+t)-cm\,(1-s)]} < \frac{1}{2}$$

where the dominators are positive and the last inequality holds because of the trade feasibility condition. When $s \geq 1/2$, we set $\mu_L = \mu_H = 0$.

It is easy to see that the (unique) solutions to the equations $\Delta\pi^H\,(n_H, 1, \lambda) = 0$ and $\Delta\pi^L\,(0, n_L, \lambda) = 0$ are respectively given by

$$\widetilde{n}_H = \frac{2s-1}{2s} + \frac{2a-bt-m[2b+c(1-s)]}{cst}(2\lambda-1) \tag{13}$$
$$\widetilde{n}_L = \frac{1}{2(1-s)} + \frac{2a-bt+cms}{ct(1-s)}(2\lambda-1) > 0. \tag{14}$$

The following result is then proved in Appendix A.

Proposition 1 *Under the trade feasibility condition, the equilibrium location of high- and low-cost firms is characterized by*

(i) co-agglomeration in the large country ($n_L^* = n_H^* = 1$) when $\lambda - 1/2 \geq \mu_A$;

(ii) partial selection of high-cost firms in the large country ($n_L^* = 1 > n_H^* = \widetilde{n}_H$) when $\mu_A > \lambda - 1/2 \geq \mu_H$;

(iii) perfect selection of low-cost and high-cost firms ($n_L^* = 1 > n_H^* = 0$) when $\mu_H > \lambda - 1/2 \geq \mu_L$; and

(iv) partial selection of low-cost firms in the small country ($n_L^* = \widetilde{n}_L > n_H^* = 0$) when $\mu_L > \lambda - 1/2$.

Firms faces a *proximity-competition trade-off* in that they benefit from a better proximity to the large market but face tougher competition in this market when more firms agglomerate there. Proposition 1, illustrated in Figure 1, tells us how market size asymmetry affects the selection of low-cost and high-cost firms. Specifically, when the two countries have very different sizes, all firms co-agglomerate in the large country because the proximity benefit outweighs the cost of competition. Regardless of the value of trade costs, country 1 is so large that the access to its market is always the dominant force. For the same reason, co-agglomeration in the small country is never an equilibrium. As the difference in market size becomes smaller, some high-cost firms re-locate in the small country, which offers them better protection against the competition of low-cost firms, whereas high-cost firms completely sort out. The long-run equilibrium then involves perfect selection, with all high-cost firms being located in the small country and all low-cost firms in the large one. Last, as market size asymmetries become very small, even some low-cost firms find it profitable to set up in the small country in order to weaken competition with the bulk of low-cost firms located in the large country.[12]

Insert Figure 1 about here

Clearly, for all the configurations described in Proposition 1 to be possible, s must be smaller than 1/2. When s exceeds 1/2, the configurations (iii) and (iv) never arise because μ_H and μ_L are equal to zero. In other words, *the large country always hosts all*

[12] In the special case of two symmetric countries, it is readily verified that partial selection of high-cost (resp., low-cost) firms prevails when s is larger (resp., smaller) than 1/2. When $s = 1/2$, any distribution of firms is an equilibrium.

the low-cost firms as well as some high-cost ones. This is because the number of low-cost firms is sufficiently small to make the intensity of competition in the large market weak enough for some high-cost firms to locate there. In contrast, when low-cost firms are many ($s > 1/2$), these may want to get dispersed between the two countries in order to relax competition within their own group. In this case, *more than half of the low-cost firms set up in the large market* (see (14)), *while all the high-cost firms agglomerate in the small one* in order to be located away from the larger number of low-cost firms.

One of the main results known in economic geography is the home market effect (henceforth, HME), which states that the large country attracts a more than proportionate share of firms (Helpman and Krugman, 1985; Ottaviano and Thisse, 2004). This result has been derived in the case of homogeneous firms. It is therefore natural to ask the question: how is the HME affected when firms are heterogeneous? The HME holds if the large country hosts a share of the industry that exceeds its share in global expenditure, i.e. $N_1^* > \lambda$. It turns out to be more convenient to check whether $N_1^* - 1/2$ is larger than $\lambda - 1/2$.

(i) In the co-agglomeration configuration, $N_1^* - 1/2$ is equal to $1/2$ while $\lambda - 1/2$ is slightly less than $1/2$. (ii) In the configuration with partial selection of high-cost firms in the large country, we have

$$N_1^* - \frac{1}{2} = [(1-s)\cdot 1 + s \cdot \widetilde{n}_H] - \frac{1}{2} = \frac{2a - bt - m[2b + c(1-s)]}{ct}(2\lambda - 1). \qquad (15)$$

Under the trade feasibility condition, it is readily verified that the term multiplying $(2\lambda - 1)$ is always larger than $1/2$. (iii) In the configuration with perfect selection, we have

$$N_1^* - \frac{1}{2} = [(1-s)\cdot 1 + s \cdot 0] - \frac{1}{2} = \frac{1}{2} - s.$$

Using the trade feasibility condition, this expression is larger than μ_H, which is itself larger than $\lambda - 1/2$. (iv) Finally, in the configuration with partial selection of low-cost firms in the small country, we obtain

$$N_1^* - \frac{1}{2} = [s \cdot 0 + (1-s)\widetilde{n}_L] - \frac{1}{2} = \frac{2a - bt + cms}{ct}(2\lambda - 1). \qquad (16)$$

The term multiplying $(2\lambda - 1)$ is larger than the corresponding term in (15), so that it must exceed $1/2$. To sum up, in all equilibrium configurations, we have

$$N_1^* > \lambda.$$

Hence, we have:

Proposition 2 *Under the trade feasibility condition, the home market effect holds regardless of the degree of firms' heterogeneity.*

In the homogeneous firm case, the intensity of the HME rises when trade costs decrease. Since the terms multiplying $\lambda - 1/2$ in configurations (ii) and (iv) include trade costs as well as firm heterogeneity parameters, we may expect a richer set of results. To see what happens, the next two sections thoroughly discuss the impact of trade costs and firm heterogeneity on firms' locations.

3.2 Trade costs

We are equipped to undertake the standard thought experiment of economic geography by determining the values of trade costs for which the above configurations are long-run equilibria. For zero trade costs ($t \to 0$), we get $(\mu_A, \mu_H, \mu_L) = (0, 0, 0)$, which implies that all types of firms agglomerate in the large country. Furthermore, it is readily verified that a rise in trade cost (weakly) decreases the number of firms in the large country. Indeed, all the thresholds (μ_A, μ_H, μ_L) increase with t, whereas both $\widetilde{n}_H$ and $\widetilde{n}_L$ decrease with t. Therefore, as t increases, the graph of n_H^* is shifted rightward (see Figure 1 for an illustration). This means that a rise in trade costs (weakly) decreases the number of high-cost firms in the large country. As the same argument applies to the graph of $\widetilde{n}_L$, a rise in trade costs also (weakly) decreases the number of low-cost firms in the large country. The long-run equilibrium can then be characterized as a function of trade costs as it follows. Let us fix the country size asymmetry λ and let t_A be the value of trade costs such that μ_A is equal to $\lambda - 1/2$. When $s < 1/2$, let t_H and t_L be the values of trade costs such that μ_H and μ_L are respectively equal to $\lambda - 1/2$. Because $\mu_A < 1/2$ under the trade feasibility condition, t_A is always positive and smaller than t_{trade}. Likewise, if $s < 1/2$, we have $t_A < t_H < t_L < t_{trade}$. The above discussion can be summarized in the following proposition:

Proposition 3 *Under the trade feasibility condition, the equilibrium location of high- and low-cost firms is characterized by*
(i) co-agglomeration in the large country if $t \in (0, t_A)$;
(ii) partial selection of high-cost firms in the large country if $s < 1/2$ *and* $t \in [t_A, t_H)$ *or if* $s \geq 1/2$ *and* $t \in [t_A, t_{trade})$;
(iii) perfect selection of high- and low-cost firms if $s < 1/2$ *and* $t \in [t_H, t_L)$; *and*
(iv) partial selection of low-cost firms in the small country if $s < 1/2$ *and* $t \in [t_L, t_{trade})$.

This proposition is illustrated in the left hand panel of Figure 2. Decreasing trade costs affects the location and selection of firms as increasing market size asymmetries does. In particular, as market integration gets deeper, i.e. as trade costs fall from t_{trade}, low-cost firms sort completely out of the small country before high-cost firms begin to sort out. In other words, lower trade costs first entice more low-cost firms to locate in the large country and then, after all of them have relocated, it entices high-cost firms to move away from the small market. Consequently, the share of low-cost firms in the large country always exceeds the share of high-cost firms, whereas the opposite holds in the small country.

Insert Figure 2 about here

Proposition 2 has another important implication about the relative competitiveness of different territories, namely *the large market is always more productive than the small* *market*, which agrees with the empirical studies mentioned in the introduction. However, market integration is does not always exacerbate international disparities. Indeed, *the relationship between* market integration and the productivity *gap is bell-shaped*. This is illustrated in the right hand panel of Figure 2 where the productivity gap is measured as the difference between the averages of marginal costs in the small and large country (this difference is always positive). As trade costs fall from very high values to t_L, the productivity gap widens because low-cost firms relocate in the large country. In contrast, average national productivities do not change in the trade cost interval $[t_H, t_L]$ because there is no relocation of firms. Finally, when trade costs fall further below t_H, high-cost firms now move to the large country, thus reducing the productivity of this country and, therefore, diminishing the productivity gap. Hence, in the presence of cost heterogeneity, the impact of trade costs on average productivity is not monotone. We summarize this result in the following proposition.

Proposition 4 *The large country is always more productive than the small one. However, as market integration gets deeper, the productivity gap between the large and small countries first increases and then decreases.*

Observe, in passing, that Proposition 1 implies that a similar result holds for asymmetries in population sizes as measured by λ.

3.3 Firm heterogeneity

In this section, we turn to another important issue, i.e., the impact of firms' heterogeneity on their location choices. There are two ways to tackle this issue since there are two parameters that capture firms' heterogeneity, i.e. m and s. We first consider the impact of an increase in the cost differential measured by the difference between high-cost and low-cost firms' marginal costs.

It is easy to see that $\widetilde{n}_L$ increase with m. In other words, when partial selection of low-cost firms in the small country prevails, increasing m leads more low-cost firms to set up in the large country. Indeed, a larger cost differential makes it harder for the high-cost firms to export in the large market, thus softening competition in that market. This in turn entices some low-cost firms to relocate there. In contrast, it is readily verified that $\widetilde{n}_H$ decreases with m : when partial selection of high-cost firms in the large country prevails, increasing m induces fewer high-cost firms to set up in the large country. This is because competition in that country becomes harsher for these firms, thereby inducing them to seek protection against competition from the low-cost ones by establishing themselves in the small country. Thus, we have:

Proposition 5 *Under partial selection, a larger cost differential leads more low-cost firms to locate in the large country or more high-cost firms to locate in the small country.*

We now investigate the impact of the *share* of high-cost firms in the global economy. First of all, note that s has no impact on the location pattern in configuration (i), which involves co-agglomeration of both types of firms in the large country. In configuration (ii), which involves partial selection of the high-cost firms, the global number of firms located in the large country is equal to $N_1^* = 1 - s + s\widetilde{n}_H$ so that, by (16), we get that

$$\frac{dN_1^*}{ds} = \frac{m}{t}(2\lambda - 1) > 0.$$

Hence, as the number of high-cost firms in the global economy grows, the global number of firms located in the large country increases. Furthermore, as the number of high-cost firms rises, their number in the large country grows at a faster pace since

$$\frac{d(s\widetilde{n}_H)}{ds} > 1.$$

This is because the lower number of low-cost firms in country 1 makes competition in this market less fierce, thus allowing a more than proportionate number of high-cost firms to locate in this country. Low-cost firms are, therefore, less effective in deterring high-cost firms to set up in the larger market. As s increases, the selection effect becomes weaker and yields a larger concentration of high-cost firms in the large country.

In configuration (iii) involving perfect selection of low-cost and high-cost firms, the number of firms in country 1 is given by $N_1^* = 1 - s$, thereby decreasing with the share of high-cost firms in the global economy. Although more high-cost firms are now established in the small country, all these firms prefer to stay in country 2 because of the selection effect. In other words, an increase in the number of high-cost firms is inversely related to the number of firms located in the large country.

Last, in configuration (iv) involving partial selection of low-cost firms, the global number of firms located in the large country is given by $N_1^* = (1 - s)\widetilde{n}_L$. Using (16), we can readily verified that

$$\frac{dN_1^*}{ds} = \frac{m}{t}(2\lambda - 1) > 0$$

which means that a growing number of high-cost firms in the global economy leads to a larger number of high-cost firms in the large country. Because

$$\frac{d\widetilde{n}_L}{ds} > 1$$

the share of low-cost firms in the large country ($\widetilde{n}_L$) increases faster than the global number of low-cost firms $(1 - s)$ decreases. This is because the lower number of low-cost firms in country 1 makes competition in this market less fierce, thus allowing a larger share of these firms to locate there. Figure 3 summarizes those results.

Insert Figure 3 about here

As shown by Figure 3, firms' agglomeration in the large market is a non monotone function of the number of high-cost firms. An increase in the global number of high-cost firms induces more agglomeration in the large country when this number is low or high enough. However, there is more dispersion when this number takes intermediate values. This allows us to make the following point. As also shown by this figure, for $s \in (0, s_L)$, firms agglomerate more intensively in a heterogeneous environment than in a homogeneous environment involving low-cost firms only (i.e. $N_1^*(s) > N_1^*(s = 0)$); conversely, when $s \in (s_H, 1)$, they agglomerate less intensively in a heterogeneous environment than in a homogeneous environment with high-cost firms only (i.e. $N_1^*(s) < N_1^*(s = 1)$). Hence, *heterogeneous environments can yield economic spaces that are either more or less agglomerated than their homogeneous counterparts.* Hence, an economy with low-cost and high-cost firms can be associated with a spatial pattern that is not the corresponding combination of the two homogeneous patterns. Among other things, this implies that there is a discrepancy between the impacts of "marginal" and "finite" changes in the number of high-cost firms on firms' agglomeration. Indeed, whereas a marginal rise fosters agglomeration in configurations (ii) and (iv), an increase shifting the economy from configuration (ii) to configuration (iv) leads to a different selection pattern that results in more dispersion.

This leads to the following results, the proof of which is contained in Appendix B.

Proposition 6 *The size of the manufacturing sector established in the large country is a non monotone function of the global number of high-cost firms: it first increases, then decreases and last increases again to become constant. Furthermore, there exist thresholds* (s_L, s_H), *with* $0 \leq s_L \leq s_H \leq 1/2$, *such that for* $s \in (0, s_L)$, *the heterogenous economy results in more agglomeration than the economy with only low-cost firms, and such that for* $s \in (s_H, 1)$, *the heterogenous economy yields less agglomeration than the economy with only high-cost firms.*

It is also worth comparing two economies involving low-cost or high-cost firms only. They do not display the same pattern of firms because $N_1^*(s = 0) > N_1^*(s = 1)$ once full agglomeration does not prevail. In other words, there are more firms in the large country when firms are low-cost than when firms are high-cost. This is because the share of trade cost in marginal cost is larger for low-cost firms than for high-cost firms. As a result, the opportunity cost to relocate to the small country is a larger for low-cost firms. Low-cost firms are therefore more reluctant to locate in the small country, meaning that more firms agglomerate in the large country when the economy involves only low-cost firms. To put it in another way, once all firms are homogeneous, *the HME is stronger when firms are low-cost*.

Observe that a unilateral change in m or s gives rise a corresponding change in the global average marginal cost ms. Consequently, the above comparative statics implies the variation of this cost, which affects indirectly the equilibrium outcome. It is therefore worth comparing the locus depicted in Figure 3 with the distribution of firms obtained when all firms have the same marginal cost equal to ms. In this benchmark case, we know from Ottaviano and Thisse (2004) that the equilibrium distribution of firms is given by

$$N_1^* = \frac{1}{2} + \frac{2(2a - 2bms - bt)}{ct}\left(\lambda - \frac{1}{2}\right) \tag{17}$$

which is a linear and downward sloping function of s. Since firms are homogeneous both when $s = 0$ and $s = 1$ with, respectively, $N_1^*(s = 0) > N_1^*(s = 1)$ firms located in country 1, the line (17) intersects the locus of Figure 3 at a single point $\widehat{s}$ that corresponds to configuration (iii). Hence, compared to the homogeneous benchmark case, there is more agglomeration in the heterogeneous case when $s < \widehat{s}$ and less agglomeration when $s > \widehat{s}$. As seen in the foregoing, the presence of high-cost firms softens competition and facilitates the agglomeration of low-costs firms in the large market. When s is not too large, this leads to a larger agglomeration of firms of both types in the big market. On the other hand, when s is large, a larger share of high-cost firms relax competition makes the overall distribution of firms less uneven. This is because the economy is now dominates by inefficient firms which try to alleviate competition among themselves.

4 The selection of local firms

In this section, we assume that trade cost are too high to allow firms to export. This implies that N_i is now given by the number of varieties produced in country i. This number is no longer constant since it varies with the number of firms that choose to locate in country i. The individual demand is thus given by

$$q_i(v) = a_i - (b_i + c_i N_i)p_i(v) + c_i P_i.$$

where the coefficients are given in (2).

A θ-type firm ($\theta = H, L$) located in country 1 (mirror expressions hold for country 2) maximizes its local profit given by

$$\Pi_1^\theta = \lambda \left(p_1^\theta - m_\theta\right) q_1^\theta$$

where $m_\theta = 0$ if $\theta = L$ and $m_\theta = m$ if $\theta = H$. Applying the first-order conditions yields

$$p_1^\theta(P_1) = \frac{1}{2}\frac{a + cP_1}{b + cN_1} + \frac{m_\theta}{2}$$

where the consumer price index is now given by

$$P_1 = sn_1^H p_1^H(P_1) + (1-s)\, n_1^L p_1^L(P_1).$$

It is readily verified that the equilibrium price set by a θ-firm located in country 1 is given by

$$p_1^\theta = \frac{1}{2}\frac{2a + cmsn_1^H}{2b + cN_1} + \frac{m_\theta}{2}$$

which decreases with N_1. The corresponding demand writes as follows:

$$q_1^\theta = a\frac{b+c}{2b + cN_1} - \frac{1}{2}(b+c)\left(m_\theta - m\frac{csn_1^H}{2b + cN_1}\right).$$

For a θ-type firm to operate in country 1, we must have $q_1^\theta > 0$ for all (n_1^H, n_1^L), with $\theta \in \{H, L\}$. This is equivalent to

$$a_1 > \frac{1}{2}(2b_1 + c_1)\, m.$$

Rewriting this condition in terms of the parameters (a, b, c), we get the *production feasibility condition*

$$a > \frac{1}{2}(2b + c)\, m \tag{18}$$

which we assume to hold from now. The equilibrium profits are

$$\Pi_1^\theta = \frac{1}{4}\lambda\,(b + c)\left[\frac{a}{b} - F(n_1^H, n_1^L) - m_\theta\right]^2 \tag{19}$$

where

$$F(n_1^H, n_1^L) \equiv \frac{c}{b}\frac{a\,(1 - s)\, n_1^L + (a - bm)\, sn_1^H}{2b + cn_1^L\,(1 - s) + csn_1^H}.$$

It can be shown that F increases in n_1^L and n_1^H under the production feasibility condition. Thus, profits are lower in a country that hosts more high- and low-cost firms.

Using (19), the profit differential is now equal to

$$\begin{aligned}
\Delta\pi^\theta\,(n_H, n_L, \lambda) &= \Pi_1^\theta - \Pi_2^\theta \\
&= \frac{1}{4}\,(b + c)\left[\begin{array}{c} \lambda\left(\frac{a}{b} - F\,(n_H, n_L) - m_\theta\right)^2 \\ -\,(1 - \lambda)\left(\frac{a}{b} - F\,(1 - n_H, 1 - n_L) - m_\theta\right)^2 \end{array}\right]
\end{aligned}$$

It follows from this expression that $\Delta_\theta \geq 0$ iff $\sqrt{\lambda}\ (a/b - F\,(n_H, n_L) - m_\theta) > \sqrt{1 - \lambda}$ $(a/b - F\,(1 - n_H, 1 - n_L) - m_\theta)$. Because $\lambda > 1/2$, this inequality is more constraining for larger marginal cost m_θ. This in turn implies that $\Delta\pi^L > \Delta\pi^H$ for all (n_H, n_L). Therefore, as in the case of tradeable goods, we are left with five possible configurations: (i) co-agglomeration of all types of firms in the large country; (ii) partial selection of high-cost firms in the large country; (iii) perfect selection of low-cost and high-cost firms; (iv) partial selection of low-cost firms in the small country; and finally (v) co-agglomeration in the smaller country. Here too, the last configuration never turns out to be an equilibrium outcome.

As in the previous section, by studying the behavior of functions $\Delta\pi^H$ and $\Delta\pi^L$ with respect to n_H and n_L in $[0, 1]$, we can determine which configuration is a long-run equilibrium. The unique solution to the equations $\Delta\pi^H\,(n_H, 1, \lambda) = 0$ and $\Delta\pi^L\,(0, n_L, \lambda) = 0$

is given by

$$\widehat{n}_L \equiv \frac{2}{c(1-s)} \frac{a(2b+c)\sqrt{\frac{\lambda}{1-\lambda}} - b(2a+cms)}{2a\sqrt{\frac{\lambda}{1-\lambda}} + 2a + cms}$$

$$\widehat{n}_H \equiv \frac{1}{cs} \frac{(2b+cs)[2(a-bm) - cm(1-s)]\sqrt{\frac{\lambda}{1-\lambda}} - 2(a-bm)[2b+c(1-s)]}{[2(a-bm) - cm(1-s)]\sqrt{\frac{\lambda}{1-\lambda}} + 2(a-bm)}$$

Defining the thresholds

$$\eta_A \equiv \frac{1}{b}\frac{(a-bm)(2b+c)}{2(a-bm) - cm(1-s)}$$

$$\eta_H \equiv 2\frac{(a-bm)[2b+c(1-s)]}{(2b+cs)[2(a-bm) - cm(1-s)]}$$

$$\eta_L \equiv \frac{(2a+cms)[2b+c(1-s)]}{2a(2b+cs)}.$$

we have the following result (the proof is omitted for it is similar to that of Proposition 1).

Proposition 7 *Under the production feasibility condition, the equilibrium location of high- and low-cost firms is characterized by*

(i) co-agglomeration in the large country ($n_L^* = n_H^* = 1$) *when* $\sqrt{\frac{\lambda}{1-\lambda}} \geq \eta_A$;

(ii) partial selection of high-cost firms in the large country ($n_L^* = 1 > n_H^* = \widehat{n}_H$) *when* $\eta_A > \sqrt{\frac{\lambda}{1-\lambda}} \geq \eta_H$;

(iii) perfect selection of low-cost and high-cost firms ($n_L^* = 1 > n_H^* = 0$) *when* $\eta_H > \sqrt{\frac{\lambda}{1-\lambda}} \geq \eta_L$; *and*

(iv) partial selection of low-cost firms in the small country ($n_L^* = \widehat{n}_L > n_H^* = 0$) *when* $\eta_L > \sqrt{\frac{\lambda}{1-\lambda}}$.

Hence, we may conclude that *the tradeable and non-tradable goods cases obey the same general principles*, thus suggesting that our assumption of trade feasibility is not restrictive. There are some differences between the two situations, however. For example, it is readily verified that both $\widehat{n}_L$ now decreases with m. Indeed, when the cost differential m rises, the competitiveness of the low-cost firms is unaffected in the large market since no high-cost firms are located there. In contrast, the competitiveness of the low-cost firms increases in the small market, thus making it more attractive to them.

5 Concluding remarks

Our main results may be summarized as follows. First, spatial selection arises in that efficient firms tend to locate in the large country and inefficient firms in the small one. Second, the presence of a not too large number of inefficient firms softens market competition and makes it more likely to observe a more agglomerated pattern of firms. Third, decreasing trade costs leads more efficient firms to locate in the large market. Last, when trade costs are sufficiently low, both types of firms are agglomerated in the large country, regardless of the cost difference across firms. These results lead us to conjecture that, for intermediate levels of market integration, when firms are free to choose to become more efficient through R&D expenditures, most of the firms which invest in R&D will locate in the large country, whereas most of the firms that refrain from making this effort will set up in the small country.

Our model has some implications regarding the current North-South trade liberalization debate. We find that a trade cost reduction via bilateral trade liberalization first leads efficient firms to congregate in the larger market and inefficient firms to set up in the smaller one. Further bilateral trade liberalization may even end up with full agglomeration in the North. Hence, it seems fair to say that bilateral trade liberalization, mainly promoted by developed countries, fosters the agglomeration highly productive firms in the North. Hence, the South has an incentive to deviate from this kind of liberalization toward Northern unilateral trade liberalization in which the South seeks protection by setting higher trade costs. This is because higher trade costs allow the South to attract firms from the North and foster relocation in the South, a situation that can be viewed as a special type of offshoring. As a result, the South might intend to deter the current trade liberalization process to promote its industrialization. And indeed, the WTO, UNCTAD and many developed countries allow for Special and Differential Provision as well as for Generalized System of Preferences toward developing countries.

Admittedly, our framework suffers from several drawbacks. First, though fairly standard in the economics literature, the assumption of two firm types is restrictive and one may wonder how robust are our results when several types of firms are considered. It is well known from incentive theory that dealing with an arbitrary finite number of types gives rise to several technical difficulties, but they do not deeply affect the main results (Laffont and Martimort, 2002). Extending our setting to the case of a continuum of types, it can be shown that, apart from the case of agglomeration, perfect sorting arises: the larger region accommodates the more efficient firms while the less efficient firms locate in the smaller region. This result together with Proposition 1 suggest that, in the case of $n > 2$ types of firms, the spatial selection of firms across markets obeys a cascade-like process based on their respective productivity. Second, the assumption of homogeneous labor and equal wage across countries is fairly restrictive. In particular, one expects the more efficient firms to attract the more efficient workers in the larger country, thus generating a double selection effect. Last, our setting should be extended to the case of several countries in order to study the impact of economic integration through full or regional trade agreements. Such an extension should rank high on the research agenda.

References

[1] Asplund, M and V. Nocke (2006) Firm turnover in imperfectly competitive markets. *Review of Economic Studies* 73, 295-327.

[2] Aw, B., S. Chung and M. Roberts (2000) Productivity and turnover in the export market: micro evidence from Taiwan and South Korea. *World Bank Economic Review* 14, 65-90.

[3] Baldwin, R. and T. Okubo (2006) Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography* 6, 323-346.

[4] Behrens, K. (2005) How endogenous asymmetries in interregional market access trigger regional divergence. *Regional Science and Urban Economics* 35, 471-492.

[5] Bernard, A.B., J. Eaton, J.B. Jensen, and S. Kortum (2003) Plants and productivity in international trade. *American Economic Review* 93, 1268-1290.

[6] Bernard, A. B. and J. B. Jensen (1995) Exporters, jobs, and wages in US manufacturing, 1976-1987. *Brookings Papers on Economic Activity, Microeconomics*. Washington DC.

[7] Bernard, A.B. and J.B. Jensen (1999) Exceptional exporter performance: cause, effect, or both? *Journal of International Economics* 47, 1-26.

[8] Bernard A.B. and J.B. Jensen (2004) Why some firms export. *Review of Economics and Statistics* 86, 561-569.

[9] Del Gatto, M., G. Mion and G.I.P. Ottaviano (2006) Trade integration,firm selection and the costs of non-Europe. Discussion Paper 5730, Centre for Economic Policy Research.

[10] Del Gatto, M., G.I.P. Ottaviano and M. Pagnini (2008) Openness to trade and industry cost dispersion: evidence from a panel of Italian firms. *Journal of Regional Science* 48, 97-129.

[11] Duranton, G. and H.G. Overman (2005) Testing for location using micro-geographic data. *Review of Economic Studies* 72, 1077-1106.

[12] Ederington, J. and P. McCalman (2008) Endogenous firm heterogeneity and the dynamics of trade liberalization. *Journal of International Economics* 74, 422–440.

[13] Engel, C. and J.H. Rogers (1996) How wide is the border? *American Economic Review* 86, 1112-1125.

[14] Foster, L, J. Haltiwanger and C. Syverson (2008) Reallocation, firm turnover, and efficiency: selection on productivity or profitability? *American Economic Review,* forthcoming.

[15] Haskel, J. and H. Wolf (2001) The law of one price - A case study. *Scandinavian Journal of Economics* 103, 545-558.

[16] Head, K. and T. Mayer (2004) The empirics of agglomeration and trade. In *Handbook of Regional and Urban Economics. Volume 4.* In J.V. Henderson and J.-F. Thisse (eds.) Amsterdam, North-Holland, 2609-2669.

[17] Helpman, E and P. Krugman (1985) *Market Structure and Foreign Trade.* The MIT Press.

[18] Helpman, E, M. Melitz and S. Yeaple (2004) Export versus FDI with heterogeneous firms. *American Economic Review* 94, 300-316.

[19] Laffont, J.-J. and D. Martimort (2002) *The Theory of Incentives. The Principal-agent Model.* Princeton University Press.

[20] Martin, P and C.A. Rogers (1995) Industrial location and public infrastructure. *Journal of International Economics* 39, 335-351.

[21] Mayer, T. and G.I.P. Ottaviano (2008) *The Happy Few: The internationalisation of European firms.* Brussels: BRUEGEL Blueprint Series III.

[22] Melitz, M. (2003) The impact of trade on intraindustry reallocations and aggregate industry productivity. *Econometrica* 71, 1695-1725.

[23] Melitz, M., and G.I.P. Ottaviano (2008) Market size, trade, and productivity. *Review of Economic Studies* 75, 295-316.

[24] Nocke, V. (2006) A gap for me: entrepreneurs and entry. *Journal of the European Economic Association* 4, 929-955.

[25] Okubo, T. (2008) Firm heterogeneity and Ricardian comparative advantage within and across sectors. *Economic Theory*, forthcoming.

[26] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited. *International Economic Review* 43, 409-436.

[27] Ottaviano, G.I.P. and J.-F. Thisse (2004) Agglomeration and economic geography. In *Handbook of Regional and Urban Economics, Volume 4.* In J.V. Henderson and J.-F. Thisse (eds.) Amsterdam, North-Holland, 2563-2608.

[28] Pavcnik, N. (2002) Trade liberalization, exit, and productivity improvement: evidence from Chilean plants. *Review of Economic Studies* 69, 245-276.

[29] Picard P.M., J.-F. Thisse and E. Toulemonde (2004) Economic geography and the role of profits. *Journal of Urban Economics* 56, 144-167.

[30] Picard, P. and D.-Z. Zeng (2005) Agricultural sector and industrial agglomeration. *Journal of Development Economics* 77, 75-106.

[31] Porter, M.E. (1982) *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York, Free Press.

[32] Redding, S and D. Sturm (2008) The cost of remoteness: evidence from Germany division and reunification. *American Economic Review*, forthcoming.

[33] Redding, S and A. Venables (2004) Economic geography and international inequality. *Journal of International Economics* 62, 53-82.

[34] Röller, L.-H. and B. Sinclair-Desgagné (1996) On the heterogeneity of firms. *European Economic Review* 40, 531-539.

[35] Syverson, C. (2004) Market structure and productivity: a concrete example. *Journal of Political Economy* 112, 1181-1222.

[36] Syverson, C. (2007) Prices, spatial competition, heterogeneous producers: an empirical test. *Journal of Industrial Economics* 55, 197-222.

[37] Tirole, J. (1988) *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

[38] Tybout, J. and D. Westbrook (1995) Trade liberalization and dimensions of efficiency change in Mexican manufacturing industries. *Journal of International Economics* 39, 53-78.

[39] Vogel, J. (2008) Spatial competition with heterogeneous firms. *Journal of Political Economy* 116, 423-466.

[40] Wolf, H. (2000) Intranational home bias in trade. *Review of Economics and Statistics* 82, 555-563.

6 Appendix A

It is readily verified that $\Delta\pi^L(1,1,\lambda) > \Delta\pi^H(1,1,\lambda) > 0$ iff $\lambda - 1/2 > \mu_A$, that $\Delta\pi^L(0,1,\lambda) > 0 > \Delta\pi^H(0,1,\lambda)$ iff $\mu_H > \lambda - 1/2 \geq \mu_L$, and that $0 > \Delta\pi^L(0,0,\lambda) > \Delta\pi^H(0,0,\lambda)$ iff $\lambda - 1/2 < \mu_0 \equiv -ct/[4(2a - bt - cms)] < 0$. Since $\mu_0 < 0$ when $\Delta\pi^L(0,0,\lambda) < 0$, the inequality $\lambda - 1/2 < \mu_0$ never holds, which implies that $(0,0)$ is not a long-run equilibrium. For $\mu_A > \lambda - 1/2 \geq \mu_H$, the function $\Delta\pi^H(n_H, 1, \lambda)$ decreases in n_H and has a single zero at $n_H = \widetilde{n}_H$, which characterizes the equilibrium. Indeed, when $n_H < \widetilde{n}_H$, all firms are enticed to move in country 1, whereas all firms are enticed to move into country 2 when $n_H > \widetilde{n}_H$. A similar argument applies to $\Delta\pi^L(0, n_L, \lambda)$.

7 Appendix B

For clarity, denote the number of firms in country 1 by $N_1^*(\lambda, s)$ and the thresholds μ_L, μ_H and μ_A by $\mu_L(s)$, $\mu_H(s)$ and $\mu_A(s)$. First, we show that $N_1^*(\lambda, 0) > N_1^*(\lambda, 1)$ if $N_1^*(\lambda, 1) < 1$ while $N_1^*(\lambda, 0) = N_1^*(\lambda, 1) = 1$ if $N_1^*(\lambda, 1) = 1$. To this end, consider the following functions of λ:

$$N^*(\lambda, 0) = \begin{cases} 1 & \text{if } \mu_L(0) \leq \lambda - 1/2 \\ \frac{1}{2} + \frac{2a-bt+cm}{ct}(2\lambda - 1) & \text{if } \mu_L(0) > \lambda - 1/2 \end{cases}$$

$$N^*(\lambda, 1) = \begin{cases} 1 & \text{if } \mu_A(1) \leq \lambda - 1/2 \\ \frac{1}{2} + \frac{2a-bt-2mb}{ct}(2\lambda - 1) & \text{if } \mu_A(1) > \lambda - 1/2 \end{cases}$$

where $\mu_L(0) = ct/\left[4\left(2a - bt\right)\right] < \mu_A(1) = ct/\left\{4\left[2a - b\left(2m + t\right)\right]\right\}$. The function $N^*(\lambda, 0)$ increases with λ at a higher rate than the function $N^*(\lambda, 1)$. We can therefore conclude that $N_1^*(\lambda, 0) > N_1^*(\lambda, 1)$ if $\mu_A(1) \leq \lambda - 1/2$, while $N_1^*(\lambda, 0) = N_1^*(\lambda, 1) = 1$ if $\mu_A(1) > \lambda - 1/2$. Noting that $\mu_A(1) \leq \lambda - 1/2$ implies that $N_1^*(\lambda, 1) = 1$, we have proved our claim.

Second, we prove that there exist two thresholds $0 \leq s_L \leq s_H \leq 1/2$ such that $N_1^*(\lambda, s) > N_1^*(\lambda, 0)$ if $s \in (0, s_L)$ and $N_1^*(\lambda, s) > N_1^*(\lambda, 1)$ if $s \in (s_H, 1)$. Since N_1^* increases in configuration (ii) and (iv), those thresholds are determined in configuration (iii) by the equalities $N_1^*(\lambda, 0) = 1 - s_L$ and $N_1^*(\lambda, 1) = 1 - s_H$ (see Figure 2). So, we get

$$s_L = \begin{cases} 0 & \text{if } \mu_L(0) \leq \lambda - 1/2 \\ \frac{1}{2} - \frac{2a-bt+cm}{ct}(2\lambda - 1) & \text{if } \mu_L(0) > \lambda - 1/2 \end{cases}$$

$$s_H = \begin{cases} 0 & \text{if } \mu_A(1) \leq \lambda - 1/2 \\ \frac{1}{2} - \frac{2a-bt-2mb}{ct}(2\lambda - 1) & \text{if } \mu_A(1) > \lambda - 1/2. \end{cases}$$

This allows us to infer that, if they exist, those thresholds respect the condition $s_L \leq s_H \leq 1/2$. It remains to show that the thresholds (s_L, s_H) are compatible with configuration (iii).

On the one hand, the threshold s_L is compatible with configuration (iii) if condition (a)

$$\mu_L(s_L) < \lambda - 1/2 < \mu_H(s_L)$$

evaluated at $s_L = 0$ holds when $\mu_L(0) \leq \lambda - 1/2$, or if condition (b)

$$\mu_L(s_L) < \lambda - 1/2 < \mu_H(s_L)$$

evaluated at

$$s_L = \frac{1}{2} - \frac{2a - bt + cm}{ct}(2\lambda - 1)$$

holds when $\mu_L(0) > \lambda - 1/2$. Condition (a) is never fulfilled because $\mu_H(0) < 0$. Condition (b) is equivalent to

$$\frac{ct}{4\,(2a - bt)} < \lambda - 1/2 < \frac{(c - 4b)\,t}{4\,(2a - bt + cm)}$$

which defines a non-empty set if $c^2 m + 8ab < 4b^2 t$.

On the other hand, the threshold s_H is compatible with configuration (iii) if condition (c)

$$\mu_L(s_H) < \lambda - 1/2 < \mu_H(s_H)$$

evaluated at $s_H = 0$ when $\mu_A(1) \leq \lambda - 1/2$ or if condition (d)

$$\mu_L(s_H) < \lambda - 1/2 < \mu_H(s_H)$$

evaluated at

$$s_H = \frac{1}{2} - \frac{2a - bt - 2mb}{ct}(2\lambda - 1)$$

when $\mu_A(1) > \lambda - 1/2$. It is readily verified that condition (c) is never fulfilled, whereas condition (d) is satisfied iff

$$\frac{(c - 4b)\,t}{4\,(2a - bt - 2mb)} < (\lambda - 1/2) < \frac{ct}{4\,(2a - bt - 2mb)}$$

which defines a non-empty set. Consequently, there exists an non-empty set of parameters for which (s_L, s_H) is included in the interval $(0, 1/2)$.

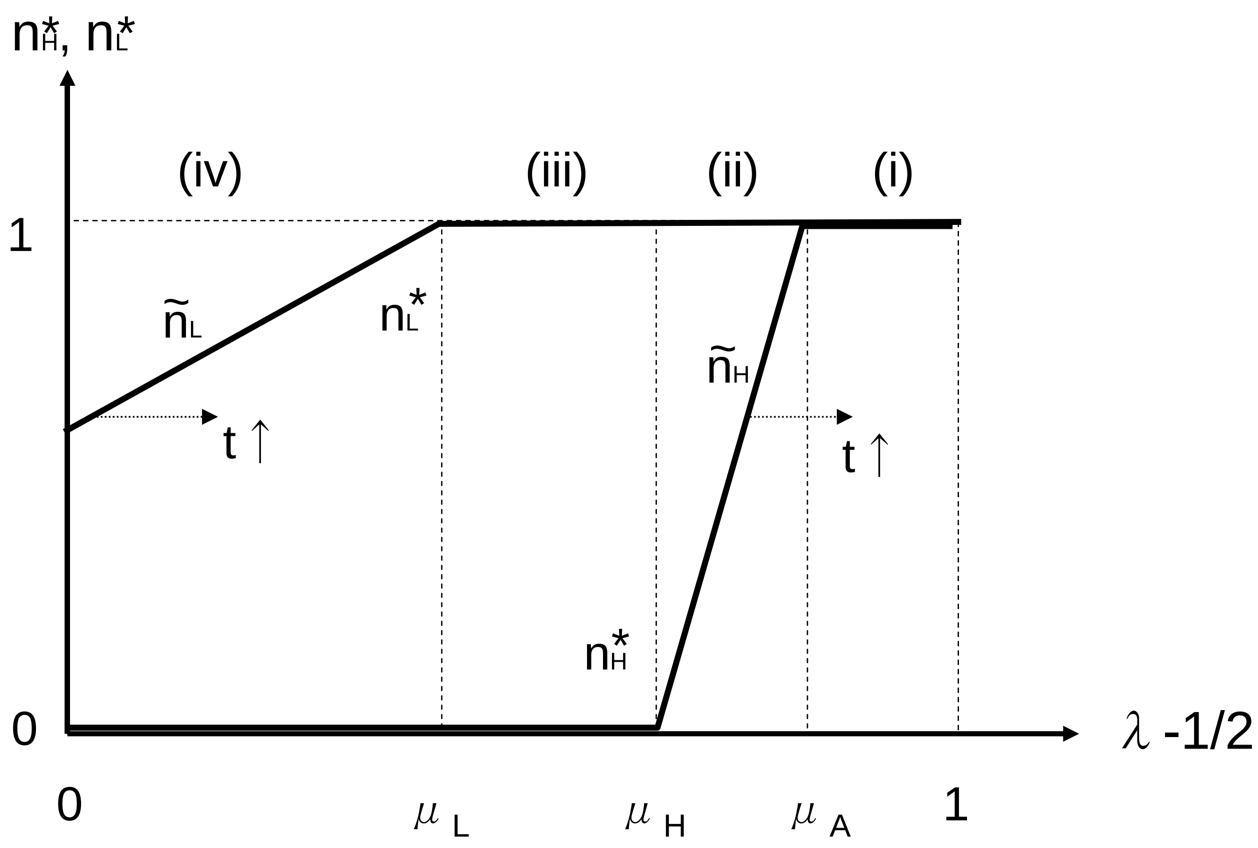

Figure 1: Spatial equilibrium and selection of high- and low-cost firms

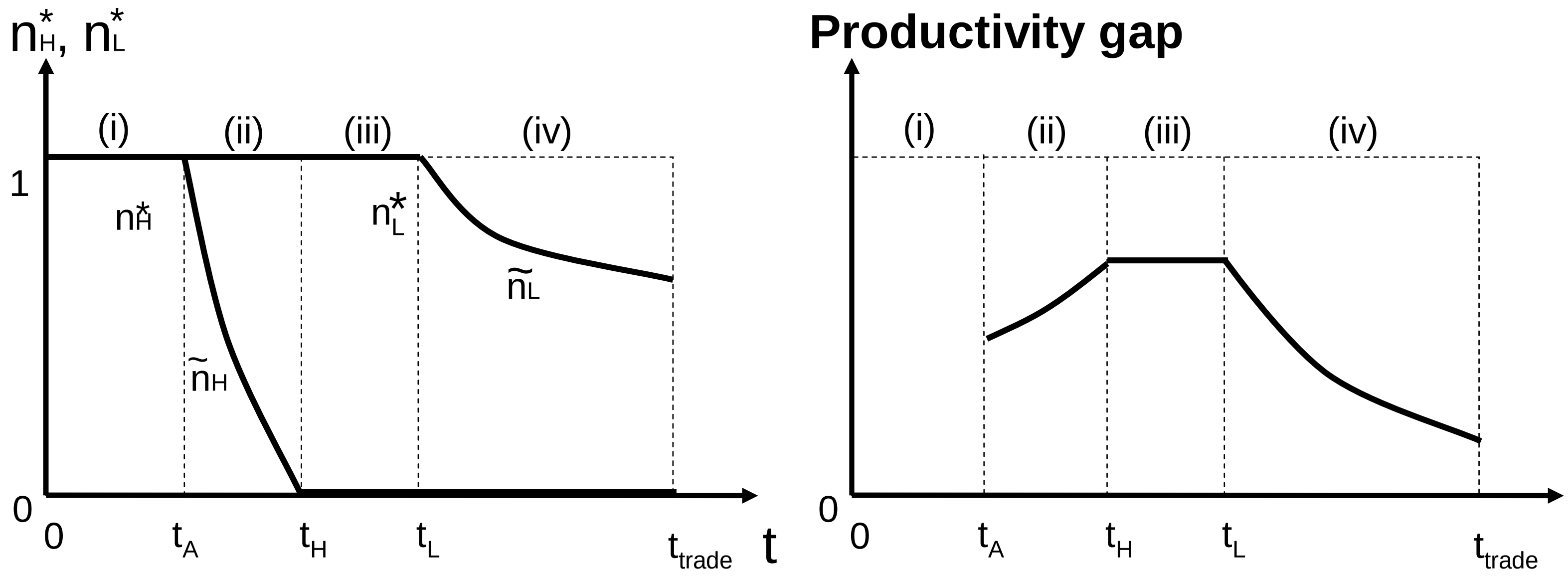

Figure 2: Spatial equilibrium and productivity gap between regions as a function of trade cost

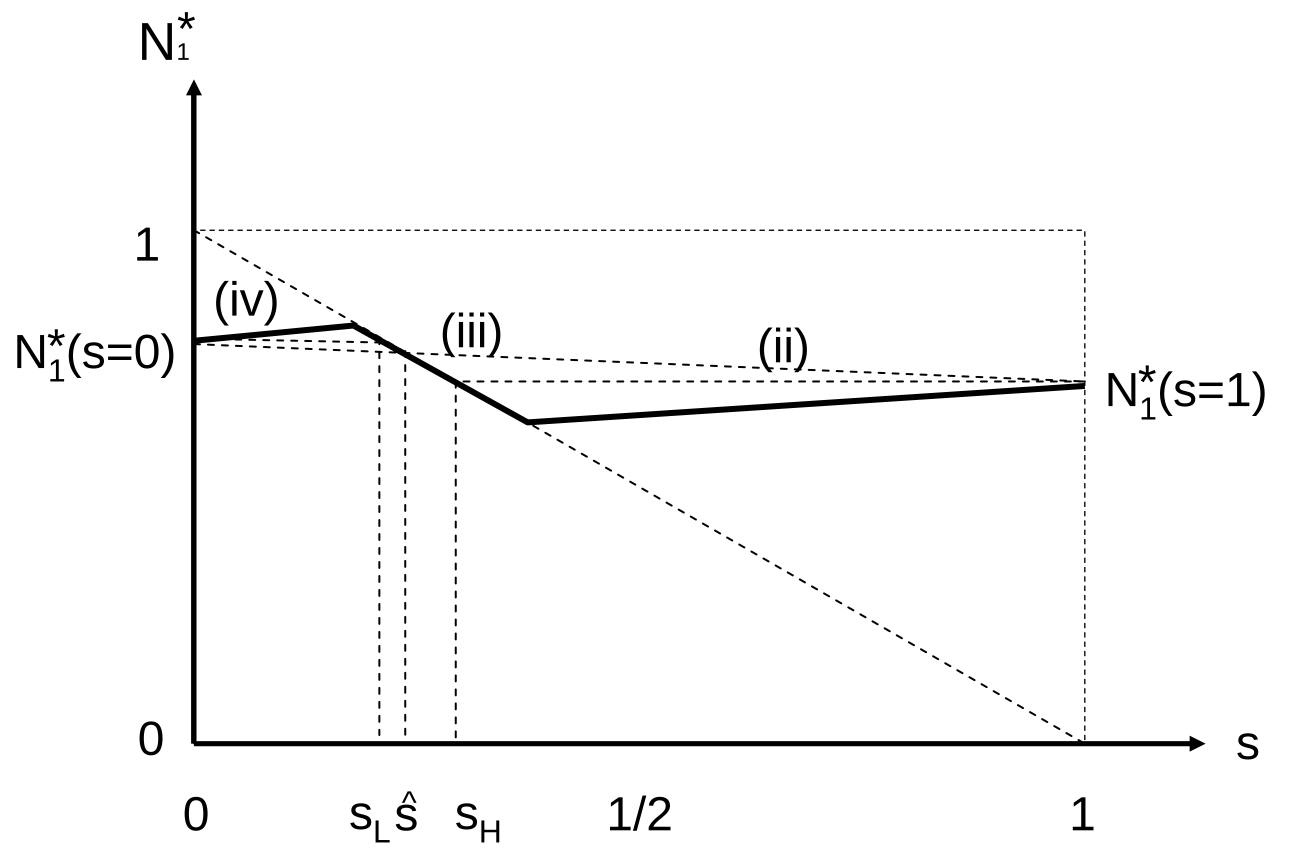

Figure 3: Agglomeration a function of the mass of inefficient firms