The Spatial Pattern of FDI: Some Testable Hypotheses[*]

Pamela Bombarda[†]

Graduate Institute of International Studies, Av. de la Paix 11A, CH-1202 Geneva, Switzerland

University of Bologna, Italy

This version: March 2008

Abstract

This paper presents a simple extension of the standard FDI model of Markusen and Horstmann (1992), which allows for a richer set of export versus FDI predictions when there are multiple nations. We propose a model with heterogeneous firms where the spatial pattern of FDI depends upon distance-linked communications costs as well as trade costs, so that some firms may service near markets with FDI and far markets with exports. The results line up with existing empirics on the aggregate knowledge-capital model and the firm-level evidence of Helpman-Melitz-Yeaple (2004).

JEL classification: F12; F23;

Keywords: intermediate goods, multinational firms, communication costs.

[*]Acknowledgements: I would like to thank R. Baldwin, my thesis supervisor, for very helpful comments and suggestions. I also wish to thank G. Calzolari, and G.I.P. Ottaviano for very useful discussions. Remarks from participants at the 2007 Zeuthen Workshop in Copenhagen, at the 2007 IAW Workshop in Tuebingen, at the 2008 VIIIth RIEF Doctoral Meeting (Barcelona) and seminar participants at the Graduate Institute and at the Geneva University also contributed to improve this paper. All remaining errors are mine.

[†]Corresponding author. Tel.: +41 22 908 59 21; fax: +41 22 733 30 49; E-mail address: bombard2@hei.unige.ch.

1 Introduction

Markusen and Horstmann (1992) developed a model in which market structure is determined endogenously as the outcome of firms' plant location decisions. They incorporated multinational firms (MNFs) into a general equilibrium trade model where firms benefit from internalization due to increasing returns at the firm level. Brainard (1993) followed a similar line of research by focusing on the location decisions proposing the so-called proximity versus concentration hypothesis, or scale versus proximity[1]. This hypothesis highlights the trade-off between reducing trade costs by locating near customers and concentrating production in only one location (which gives rise to scale economies at the plant level). In these models, firms are more likely to be engaged in foreign direct investment (FDI) activities when trade costs are high. Thus, foreign subsidiary's sales increase with distance. For the same reason, horizontal FDI is not encouraged by a reduction in trade costs. On the contrary, when trade costs fall, scale economies can outweigh the benefit from locating near customers. In this case, export activities are more profitable. Hence, the proximity versus concentration hypothesis predicts that the fall in trade costs should reduce FDI and encourage exports.

Comparing this theory with the empirical evidence on FDI, we discover that the spatial distribution of affiliates is much richer than the scale-versus-proximity would predict. In fact, despite the reduction in transport costs across countries, there has been a consistent growth of FDI inflows. The data shows that multinational enterprises account for a very significant fraction of world trade flows; and trade in intermediate inputs between divisions of the same firm constitutes an important portion of these flows (Hanson, Mataloni and Slaughter, 2001). Alfaro and Charlton (2007), using a new firm level data set, establish that vertical FDI (subsidiaries which provide inputs to their parent firms) is larger than commonly thought, even within developed countries[2]. This result is in line with Bernard, Jensen and Schott (2007), as they find that the proportion of intra-firm trade is higher between rich countries than between rich and poor countries.

Moreover, Buch, Kleinert, Lipponer and Toubal (2005), Carr, Markusen, and Maskus (2001) and Mayer and Head (2004) show that including distance as a proxy for trade costs negatively affects affiliate sales: a reduction in trade costs coincide with FDI growth[3]. Other empirical studies depart from the classical bilateral FDI assumption and introduce an element which takes into account spatial

[1] The proximity versus concentration hypothesis predicts that "firms are more likely to expand their production horizontally across borders the higher are the transport costs and trade barriers and the lower are investment barriers and the size of scale economies at the level at the plant level relative to the firm level" (Brainard, 1997).

[2] This result depends on the level of disaggregation they considered.

[3] This seems to be confirmed in the EU, where under the single market situation a reduction in trade costs has been achieved.

interdependence in foreign direct investment. Blonigen, Davies, Waddell and Naughton (2007) and Baltagi, Egger and Pfaffermayr (2006) attempt to test empirically the complex integration strategies of multinationals by considering the role of third countries as a determinant of FDI. They all find significative and positive third country effects.

This data shows a broad range of strategies that multinational enterprises can undertake, highlighting the fact that the classical distinction between horizontal and vertical FDI is not accounting for all the facts. Trade and taxes are important policies which can affect the mix of affiliate strategies. Also distance, not only geographical but also cultural, plays an important role in the strategy choices of the multinational enterprises.

In what follow we try to reconcile the MNFs theory with these recent empirical findings to explain the pattern of supply mode decisions. In order to do so, we introduce variations by firms by markets as a new element with respect to the existing literature where variations are by market (homogeneous firm) or by firm in a single market (Helpman, Melitz and Yeaple, 2004). In this model we will have firms switching type from FDI to export (or vice versa) in relation to distance: there will not be a one to one correspondence between firm type and characteristics. This model changes the usual MNF pattern by introducing non-monotonic relationships between affiliate sales and distance. As far as we know there are no other papers that look into the spatial implication of the location of FDI.

Our model is closely related to Melitz (2003) and Helpman, Melitz and Yeaple (2004), hereafter HMY. Melitz (2003) added two crucial elements to the new trade theory. The first is the fixed market entry costs that a potential entrant has to pay. The second is heterogeneity in firms' productivity. By introducing firm heterogeneity in the Krugman (1980) model, he observed how an increase in trade exposure leads to a reallocation of firms toward the more efficient, without necessarily inducing an increase in the productive efficiency of individual firms[4]. In line with Melitz (2003), HMY built a multi-country, multi-sectoral general equilibrium model with the intent to analyze the decision of heterogeneous firms to serve foreign markets either through exports or local sales (FDI). Similar to Melitz (2003), they work with heterogeneous firms, identical nations, a single factor, but with more sectors. They find that at the (sectoral) aggregate level, the ratio of FDI to export sales should be higher in industries with higher productivity dispersion. Their results rely on the assumption of perfectly symmetric countries and on the absence of asymmetries in trade costs or in fixed costs. As a consequence,, a firm that exports to one single country will also export to every other country. This

[4] This result is partially contradicted by Baldwin and Nicoud (2007), where they pointed out that "although freer trade improves industry productivity in a level sense, it harms it in a growth sense".

could limit our comprehension of reality, where usually a firm chooses a mixture of supply modes.

Building on Melitz (2003) and HMY, this paper considers the role of distance on the decision of whether and how a firm chooses to serve a foreign market. Using within-sector heterogeneity and identical countries we assume that trade costs (which depends on both "trade openness" and distance) apply to both exports and FDI because both involve transportation (the first of a finished good, the second of an intermediate good). To enrich the spatial implications of the Markusen-Venables model, it is necessary to add a few new elements to this standard model. First, we assume that the cost of coordination rises with distance, the fixed cost of supply a market by local production rises with the market's distance. Second, in keeping with the strong positive correlation between trade and FDI, we assume that an essential intermediate good must be produced at home due, say, to issues of intellectual property protection, the need for highly specialised employees, or even overwhelming scale economies that makes production of the intermediate in a single plant the optimised outcome. In the model, we take the home production of the intermediate as a given. These two elements allow for a richer range of possible mode-of-supply outcomes. For instance, if the trade costs rose slowly with distance but coordination costs rose with distance in an S-shape, we might see firms supply nearby markets with FDI and more distant market via exports.

We also introduce firm heterogeneity to allow for the possibility that different firms in the same country and industry may choose different modes of supply for a given foreign market. In this way, a single firm may choose to service some markets via exports and other by horizontal FDI, while another firm in the same industry may choose to service all through FDI. And some firms may service the nearby with exports and far away with FDI. It all turns on the interplay between a distant-related fixed costs and variable distant-related trade cost, with, of course, firm-level productivity heterogeneity.

We chose a model which uses the heterogeneous firms approach because it allows us to model some aspects of international modes of supply that until now have only been studied empirically. More precisely, the heterogeneous firms approach serves three purposes: to explain patterns of productivity differences between multinationals, exporters and national firms, to dull the knife-edge result in which homogenous firms all choose the same (or are indifferent between) supply modes, and more importantly to introduce different supply modes in different destination markets for each firm. Despite the gain in terms of a more succinct and transparent framework, the homogenous firms approach would not have allowed to observe the switch from one mode of supply to another in the same static equilibrium.

Likewise, the model can also be viewed as an enrichment of the cornerstone models with firm heterogeneity. Firstly, we give a role to distance by introducing trade and coordination costs in the MNF's activity. Secondly, we consider N symmetric countries located evenly around a circular trade route. This allows for various differences in distance between markets, while controlling for market size effects and thus permitting analytic solutions. The higher level of heterogeneity, not only among firms but also among countries, keeps the analysis closer to reality. Indeed, the introduction of firm spatial distribution gives a role to distance in determining the mode of supply decisions. This result is more comprehensive than HMY's result, where the symmetric assumption yields equilibrium where a firm that can engage in foreign market activities will be active in every foreign market independently of distance[5]. Our result is in line with recent empirical findings which confirm that the number of multinational firms is decreasing with distance. Through this work we will show that spatial distribution of affiliates is much richer than what the scale-versus-proximity model predicts.

Motivation and Intuitions. In what follows we propose some justifications of the main assumptions of the model as well as a brief description of the main results. In our model, while exporting involves a trading cost which is increasing in distance, multinationality involves a fixed cost which is also increasing in distance. Following Antras, Garicano and Rossi-Hansberg (2006), there is evidence of a travelling cost for workers from home to the host country in order to coordinate production. In addition, since multinationals need to import an intermediate good, multinationality also involves a trading cost which is increasing in distance. However, since only one of the two intermediate goods used is imported by the multinationals, the marginal cost of affiliates is always less than the marginal cost of exporting. The interaction between these two assumptions changes the usual MNF pattern by introducing non-monotonic relationships between the supply mode via FDI and distance: affiliate sales may be decreasing in distance.

The presence of firm heterogeneity in productivity raises the possibility that a more productive firm from country-i will have affiliates sell in country-j, but a less productive firm in country-i will export to country-j. More interestingly, since in this model distance becomes crucial, the same firm may have an affiliate in a near market while exporting to a farther destination. Hence in this model moving away from the parent generates the following pattern: a firm exports, then does FDI, then exports again[6].

This paper is organized as follows. Section 2 presents the model. Section 3 characterizes its

[5] In relation to its own productivity, it will be active as an exporter or as a subsidiary.

[6] Even though it seems a natural implication of this result, in this model we do not allow for the possibility to export from a foreign affiliate.

equilibrium and investigates the role of distance and the effects of trade liberalization. Section 4 considers the effects of distance on sales. The last section concludes.

2 Theoretical Framework

We study the supply mode decision between FDI and export in a multi-country framework. For this purpose, we merge the model by Helpman, Melitz and Yeaple (2004) with the one by Antras and Helpman (2004) adding few additional elements.

2.1 Preferences

Consumers in each country share the same preferences over two final goods: a homogeneous good, z, and a differentiated good, x. We assume a two-tier preference with Cobb-Douglas in the upper tier and CES in the lower tier. A fraction of income, β, is spent on the differentiated good, $c\,(v)$, and the rest $(1-\beta)$ is spent on the homogeneous good, z. The utility function is

$$U = z^{(1-\beta)} \left[\int_{v \in V} c(v)^{(\sigma-1)/\sigma} dv \right]^{\frac{\sigma\beta}{\sigma-1}}$$

where $\sigma > 1$ represents the elasticity of substitution between any two products within the group and V is the set of available varieties.

2.2 Supply

There are N identical countries located evenly around a circular trade route. We assume N to be odd, so that starting from an origin country, two destination countries are located at the same distance (each destination has a clone). We denote the distance between adjacent markets with "d". This implies that distance between markets rises by steps of d. In this framework we have two final goods, two intermediate goods and one factor. Each country is endowed with labour, L, which is supplied inelastically.

There are two sectors, one homogeneous and one differentiated. The homogeneous sector produces a homogeneous good, z, with constant returns to scale and perfect competition. In this sector the technology is simple. We choose units of z such that one unit of labour is required per unit of output. Thus, the unit cost function is w, where w is the wage rate for labour. This unit cost function represents marginal and average costs. In the homogeneous sector, competition ensures price equal marginal costs, $p_z = w$. It is convenient to choose good z as the numeraire, so that $p_z = 1$; hence, the pricing condition will become: $1 = w$. Assuming the nations are large enough, it is easy to show that homogeneous good z is produced in every country. Since it is freely traded on international markets, the cost of producing it is equal in every country, so wages are equalized.

The differentiated sector produces a continuum of horizontally differentiated varieties, $x(v)$, from two intermediate goods (or tasks), y_1 and y_2. Both y_1 and y_2 are produced with one unit of labour, but y_1 can only be made at home, due to technological appropriability issues. Each variety is supplied by a Dixit-Stiglitz monopolistically competitive firm which produces under increasing returns to scale which arise from a fixed cost.

We consider three modes of supply in the x-sector; firms which sell only domestically (D-mode); firms who export (X-mode), and firms who supply the foreign market via FDI (M-mode). Hence, when a firm decides to serve the foreign market, it chooses whether to export domestically produced goods or to produce in foreign via affiliate production.

As in Helpman, Melitz and Yeaple (2004), this choice is affected by the classical scale versus proximity trade-off. Nevertheless, in our model, geographical distance between countries matters for two different reasons, namely trade and coordination costs. The fact that y_1 can only be made at home plays an important role. If a firm chooses to supply the foreign market via local sales of its affiliates, the affiliate must import the intermediate good y_1 from the home nation. This implies that the M-mode does not entirely avoid trade costs. We also assume that workers from home must periodically travel to host country to coordinate final production. This implies that there is a second distance-related cost that we call coordination cost .

Following Helpman, Melitz and Yeaple (2004), entering the x sector involves a fixed variety-development cost f_I[7]. Subsequently, each entrant draws a labour per unit output coefficient (called a) from a cumulative density function $G(a)$ that is common to every country. The support of the continuous random variable a is $0 \leq a \leq a_0$. Upon drawing its own parameter a, each firm decides to exit (this happens if it has a low productivity draw), or to produce. In this case, the firm must face additional fixed costs linked to the mode of supply chosen. If it chooses to produce for its own domestic market, it pays the additional fixed market entry cost, f_D. If the firm chooses to export, it bears the additional costs f_X of meeting different market specific standards (for example, the cost of

[7] The subscript I stands for innovation.

creating a distribution network in a new country). Finally, if the firm chooses to serve foreign markets through FDI, there would be two types of fixed costs[8]. Firstly, there is a fixed cost of creating a distribution network as well as building up new capacities in the foreign country, f_M. In addition to this fixed cost, we assume there is a distance-related cost for workers travelling from home to the host country to coordinate the establishment of the foreign affiliate. We call this f and note that it rises with distance.

As mentioned, the homogeneous sector is not subject to trade costs, but the x-sector is subject to iceberg costs proportional to round-the-circle distance[9]. More precisely, in the X-mode case, the entire final good is subject to iceberg costs, while with M-mode only the intermediate good y_1 is subject to iceberg costs. Selling one unit in the export market located d-steps away, would require shipment from the origin country of $d\tau \geq 1$ units for the exporting sector and $(d\tau)^\eta$ for the FDI sector, where τ represents the iceberg trade cost and η the share of intermediate good used in final production. Since FDI is affected by trade costs, its marginal cost increases with distance, as well as the distance-related coordination cost.

2.3 Intermediate Results

2.3.1 Demand

Given preferences across varieties have the standard CES form, the demand function is,

$$x_i(v) = A_i p_i(v)^{-\sigma} \quad \text{where } A \equiv \frac{\beta E_i}{P^{1-\sigma}}$$

where i indicates the country, $x_i(v)$ represents the consumption of typical variety v, A_i is the demand shifter and finally $p_i(v)$ is the consumer price index of variety v. A_i is exogenous from the perspective of the firm and composed by the aggregate level of spending on the differentiated good in country i, βE^i divided by the CES price index, $P^{1-\sigma}$. Country symmetry allows us to drop the country subscript. The inverse demand function is given by

$$p(v) = A^{\frac{1}{\sigma}} x(v)^{-\frac{1}{\sigma}} \tag{1}$$

[8] In our model when a firm chooses to serve foreign markets via FDI it means local production of the intermediate good, y_2, only.

[9] Trade costs are broadly defined, so as to include different kind of impediments: trade barriers, cultural differences etc.

2.3.2 Organization and Product Variety

Given that f_I has been paid, the output of every variety is described by a Cobb-Douglas function of the intermediate goods,

$$x(v) = \frac{1}{a(v)}\left(\frac{y_1}{\eta}\right)^{\eta}\left(\frac{y_2}{1-\eta}\right)^{1-\eta}, \ 0 < \eta < 1 \tag{2}$$

where $1/a(v)$ represents the firm specific productivity parameter and η is the Cobb-Douglas cost share of y_1, common across all nations. When trade is possible, firms that produce decide whether to supply a particular market and how, i.e. via export or FDI strategies. This will depend on their own productivity and on the distance between the orgin and the destination country. As mentioned before, the marginal costs in the exporting sector will be higher than the one in the FDI sector. Hence, despite the existence of N symmetric countries, the fact that they are located evenly around a circular trade route makes distance playing a role in the consumer price[10].The trade cost is constructed in the following way: between the origin country and the nearest destination country the trade cost is τ; between the origin country and the second nearest destination country, the trade cost is defined as 2τ; between the origin country and the third nearest destination country 3τ and so on until the most distant country $(N-1)/2$ is reached, here the trade costs is $[(N-1)/2]\,\tau$. The condition in which we are interested in goes from the destination country 1-step away from the origin to the destination country $(N-1)/2$-steps away from the origin. We exploit the mirror image nature of the circle: countries $(N-1)/2$ to $N-1$ steps away are just a mirror image.

Since y_1 and y_2 are produced with L whose wage is unity, the marginal costs, mc_{Di} , for local production in every origin country is the following,

$$mc_D = a(v)$$

where country symmetry allows us to drop the country subscript[11]. The marginal cost for exporting to a market that is d-steps away is linear in $d\tau$,

$$mc_{X,d} = a(v)\,d\tau$$

[10] Since countries are symmetric and located around a circular trade route, destination countries are labelled: 1, 2, 3... N-1.

[11] See appendix A1 for details of the cost minimization problem.

where d and τ represent distance and trade cost respectively. Finally, the marginal cost for supplying the d market via local sales of foreign affiliates is concave in $d\tau$,

$$mc_{M,d} = a(v)(d\tau)^{\eta}$$

Note that in this last marginal cost distance matters but only in relation to cost share, η, of the intermediate good y_1 used in the production of the final good. Using the mark up, $\sigma/(\sigma-1)$, we can easily derive the price for each particular mode of supply decisions.

2.3.3 Mode of Supply Decisions

The mode of supply decision choice will involve the comparison of profit levels taking into account the various fixed and variable trade costs. A firm can decide to: (*i*) not supply a market, (*ii*) supply it via exports, or (*iii*) supply it via local sales of foreign affiliates. Of course, the local market is supplied by local sales, if the firm is active (*iv*).

The optimal mode of supply depends, as in HMY, on a firm's productivity. As described above, four cases are relevant.

Case (*i*). If the firm decides not supply a market and exits, the operating profits are zero.

Case (*ii*). If the firm decides to supply a market via exports, the profits from exporting to a market d-steps are linearly decreasing in d and τ,

$$\pi_{X,d} = [p_X(v) - a(v)\,d\tau]\,x(v)_X - f_X \tag{3}$$

where $x(v)_X$ represents the quantity exported.

Case (*iii*). If the firm decides to supply a market via FDI, the profits realized by a subsidiary located in the d-steps away market depend on the interaction between d and τ,

$$\pi_{M,d} = [p_M(v) - a(v)(d\tau)^{\eta}]\,x(v)_M - f(d) - f_M \tag{4}$$

where $f(d)$ represents the fixed coordination costs, $(d\tau)^{\eta}$ is the trade costs associated with the intermediate good, y_1, imported from the home country and $x(v)_M$ is the quantity supplied by the foreign subsidiary. The foreign affiliate has to face both the coordination costs, which rise with distance, and the trade costs that hit the imported intermediate.

Case (iv). If the market under consideration is the firm's home market, the profits from undertaking D-mode supply are

$$\pi_D(v) = [p_D(v) - a(v)]\, x(v)_D - f_D \tag{5}$$

where $x(v)_D$ represents the quantity supplied in the domestic market.

Using the intermediate results from consumers and firms optimization problems we calculate the operating profit for the three modes of supplying a market. The profit from serving the domestic market is a function of the demand shifter and the constant mark-up,

$$\pi_D^*(a, A, \eta) = Aa^{1-\sigma}\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} - f_D$$

where A and η are industry (and so country) specific. Using $B = \frac{A}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)}$ we obtain:

$$\pi_D^*(a, A, \eta) = Ba^{1-\sigma} - f_D \tag{6}$$

If a firm chooses the X-mode for a given foreign market, then its equilibrium net operating profit on sales in that market is

$$\pi_{X,d}^*(a, A, \eta) = B\,(d\tau a)^{1-\sigma} - f_X \tag{7}$$

If a firm chooses the M-mode for a given foreign market, then the equilibrium net operating profit it would earn is

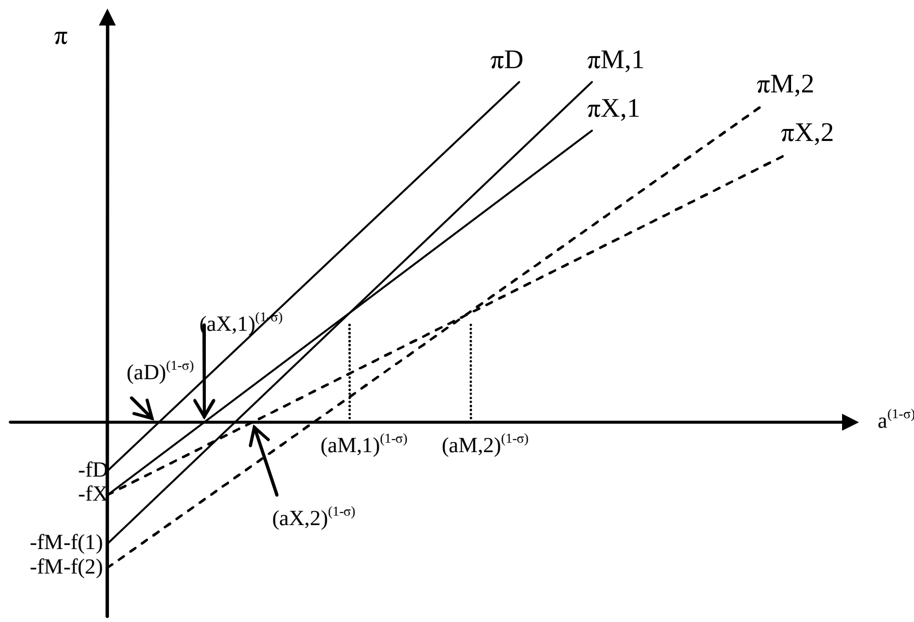
$$\pi_{M,d}^*(a, A, \eta) = Ba^{1-\sigma}\left[(d\tau)^{\eta}\right]^{1-\sigma} - f(d) - f_M \tag{8}$$

For what concerns the foreign markets, the two types of distance related costs, $d\tau$ and $f(d)$, imply that almost anything can happen. To focus on the central case, we set parameters so that we get the same ranking as in HMY when there are only two nations. Namely, firms with sufficiently high productivity will supply the foreign market at all, with the most productive supplying it via FDI rather than exports. In this way our model is in line with the HMY empirical findings. Hence, the regularity condition we need is,

$$f_D < d\tau^{(\sigma-1)} f_X < d\tau^{\eta(\sigma-1)}\left[f_M + f(d)\right]$$

In order to clarify the analysis, we represent the profit functions discussed above with the help of Figure 1.

Figure 1: Profits and Foreign Market d steps away.



In this figure, we represent how distance affects the modes of supply. On the horizontal axis we have $a^{1-\sigma}$; since $\sigma > 1$, this variable can be used as a firm-level productivity index. All the profits described in (6), (7) and (8) are increasing functions of $a^{1-\sigma}$. The diagram plots π_D, $\pi_{X,1}$ and $\pi_{M,1}$ which are the operating profits for a firm supplying a market locally (π_D), or supplying a market 1-step away (π_{X1}) or supplying the same market via M-mode ($\pi_{M,1}$). Independently of the type of activity, the more productive is a firm, the more profits it will make. The profit function $\pi_{M,1}$ is slightly flatter than π_D, due to trade costs, and its vertical intercept is lower due to higher fixed costs. The figure also plots the profits for a market 2-steps away as $\pi_{X,2}$ and $\pi_{M,2}$. Consider $\pi_{M,2}$. In this case, the slope of $\pi_{M,2}$ is flatter than $\pi_{M,1}$ and its vertical intercept is lower due to increased distance, i.e. $[(2\tau)^{\eta}]^{1-\sigma} < [(1\tau)^{\eta}]^{1-\sigma}$ and $f(2) > f(1)$. The profit function $\pi_{X,1}$ is flatter than π_D due to trade costs. In the supply mode via FDI only a part of the intermediate goods incurs in trade costs, this makes $\pi_{M,1}$ steeper than $\pi_{X,1}$; this condition is preserved for any further increase in distance: $\pi_{M,2}$ is also steeper than $\pi_{X,2}$.

From Figure 1 we see that there exist different productivity levels at which a firm is indifferent between supply modes; these productivity levels change with distance. The cutoff productivity level at which operating profits from domestic sales equal zero is $a_D^{1-\sigma}$. The productivity levels at which exporters and FDI just break even are generically $a_{X,d}^{1-\sigma}$ and $a_{M,d}^{1-\sigma}$. Greater distance will modify these cutoffs. For any given market, say one that is 1-step away, $a_{X,1}^{1-\sigma} < a_{M,1}^{1-\sigma}$. For market 2-steps away we have $a_{X,2}^{1-\sigma} < a_{M,2}^{1-\sigma}$. These conditions are ensured by the regularity condition. If $a_{X,d}^{1-\sigma}$ rises with the distance of the market "d", we cannot say the same for $a_{M,d}^{1-\sigma}$. In fact, $a_{M,d}^{1-\sigma}$ has an ambiguous behaviour with respect to distance, which depends on the freeness of trade. We cannot *a priori* rank the thresholds for X versus M, nor for M at different distances. More precisely, Figure 1 holds for sufficiently high freeness of trade and distance.

2.4 Equilibrium Conditions

We now turn to formal statements of the thresholds illustrate in Figure 1.

2.4.1 The Cutoff Conditions

Firms will choose the optimal supply mode for each market. To relate this choice to firms' marginal costs we define a threshold marginal cost, $a(v)$, for each destination and for each mode of supply. Using the equilibrium operating profit of serving the domestic market from (6), we derive the domestic cutoff condition,

$$a_D = \left(\frac{f_D}{B}\right)^{\frac{1}{1-\sigma}} \tag{9}$$

That is, firms with $a(v)$ below a_D will find it optimal to supply their local market; firms with $a(v) > a_D$ will expect negative profits and exit the industry.

The choice in foreign markets is more complex so we will structure the discussion with the help of Figure 1. As we see from the figure, the net operating profits of supplying the foreign market d-steps away rise under both modes of supply. Firms with $a_{X,d} < a(v) < a_D$ have positive operating profits from sales in the domestic market, but they lose money if they choose to supply foreign markets. Using the net operating profit from exporting (7), we can derive the X-mode cutoff,

$$a_{X,d} = \left(\frac{f_X}{B\left(d\tau\right)^{1-\sigma}}\right)^{\frac{1}{1-\sigma}} \tag{10}$$

Thus, only firms with $a(v) \leq a_{X,d}$ will consider export to the d market.

Notice from Figure 1 that at $a(v) = a_{X,d}$ exporting yields a higher net operating profit then FDI. This ordering switches, however, for firms with $a(v) \leq a_{M,d}$, where this is defined as the $a(v)$ where:

$$a_{M,d} = \left(\frac{f(d) + f_M - f_X}{B\left[(d\tau)^{\eta(1-\sigma)} - (d\tau)^{1-\sigma}\right]} \right)^{\frac{1}{1-\sigma}} \quad (11)$$

This M-mode cutoff is obtained by equating the operating profits from doing FDI, (8), with the operating profit from doing export, (7). This is because by construction, a firm will choose to supply the d-steps away country via FDI if and only if the FDI strategy is more profitable than the export strategy, i.e. if this holds:

$$\pi_{M,d} - \pi_{X,d} \geq f(d) + f_M - f_X$$

which can be rewritten as,

$$Ba_{M,d}^{1-\sigma}\left[(d\tau)^{\eta(1-\sigma)} - (d\tau)^{1-\sigma}\right] = f(d) + f_M - f_X$$

Notice that if $a(v) \leq a_{M,d}$, M-mode supply yields a higher net operating profit.

From the diagram it is clear that $a_{X,d} > a_{M,d}$ for every level of distance. Therefore, considering Figure 1, when $d = 2$, both $\pi_{M,d}$ and $\pi_{X,d}$ become flatter, but $\pi_{X,d}$ does always even more. Hence, the new crossing point defining the new equilibrium M-mode cutoff, $a_{M,2}$,will be at the right of $a_{X,2}$[12]. Nevertheless, some firms that were supplying market 1-step away via M-mode, switch to X-mode as a consequence of higher trade costs and coordination costs.

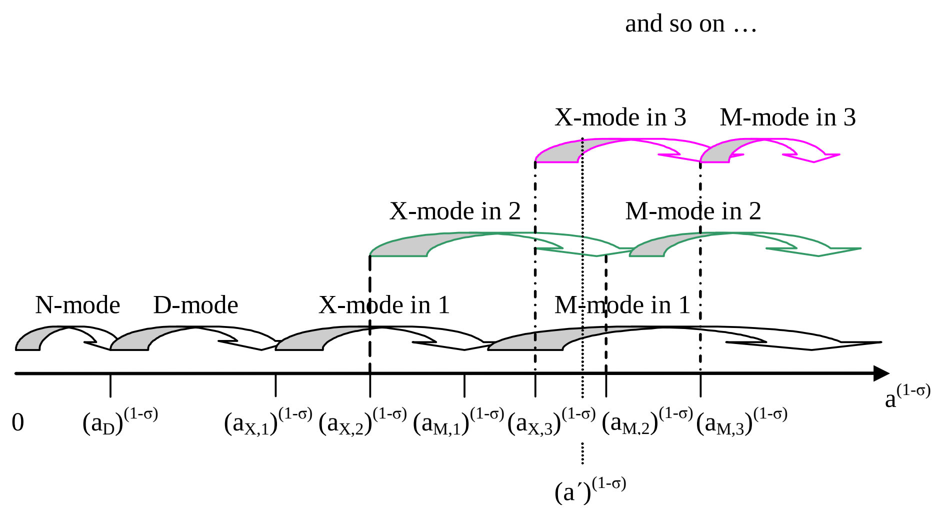
2.4.2 The Role of Distance

A key goal of our study is to characterize the spatial pattern of modes of supply. This is implicit in the cutoff conditions, but here we highlight the role of distance in explaining the variation by firms by market. The pattern of organizational forms could be characterized in two main steps. Markusen and Horstmann (1992) proposed a general-equilibrium model where MNFs arise due to a market-access motive to substitute for export flows, or what is termed "horizontal" FDI. Under the assumption of high trade costs, they found export nearby and FDI far away; this result is also called scale versus proximity result. More recently, HMY introduced firm heterogeneity in the pattern of organizational forms; here the nature of serving the market depends on the nature of the firm. Both papers are developed in a two-country framework. In the present work, the scale versus proximity is used in a model with heterogeneous firms and many locations, implying that there would not be a one to one

[12] This derives from the regularity condition assumed.

correspondence between firm type and firm characteristic. In this context, a firm could decide to supply via M-mode all markets up to a given distance and then as distance becomes too big, switch to X-mode supply. In Figure 2 we show the firm level characteristics with multiple destinations.

Figure 2: Modes of Supply and Destinations



In Figure 2 we represent the variation by firms by market with respect to the productivity level, $a^{1-\sigma}$, assuming a certain level of distance and trade openness. In this figure, we refer to the nearest destination country using the label 1. The origin country can reach the other N-1 countries with two different market access strategies: export or FDI. Points $a_D^{1-\sigma}$, $a_{X,1}^{1-\sigma}$ and $a_{M,1}^{1-\sigma}$ represent the classical HMY model, where location does not play any role: if a firm is doing FDI toward the nearest destination country, country 1, then it is also doing FDI toward the other $N-1$ countries. For this reason HMY consider types of firm: X-type, M-type. By contrast, in this paper we consider different supply modes per firm. A given firm may find it optimal to supply via exports to one market, i.e. X-mode, but via FDI to another, i.e. M-mode.

The existence of distance allows to highlight a new pattern. On the right hand side of $a_{M,1}^{1-\sigma}$ there is a new region which represents the spatial variation in M versus X mode supply, for each firm. In this analysis $a_D^{1-\sigma}$ represents the minimum productivity level in order to supply the local market; $a_{X,1}^{1-\sigma}$ is the minimum productivity level in order to become an exporter to country 1; $a_{M,1}^{1-\sigma}$ represents the productivity threshold in order to do FDI in country 1, and so on. These threshold levels change with distance. In relation to its own productivity, a single firm can undertake M-mode supply toward country 1 and then X-mode supply toward country 2 through N-1. This switching is determined by the increasing variable and fixed costs. Only the more productive firm will supply via local sales of foreign affiliates in every destination country.

Remark 1 *Under certain conditions, namely high distance and freeness of trade, the area on the right hand side of* $a_{M,1}^{1-\sigma}$ *shrinks with distance. This implies that few firms will supply the far away market via FDI. There is a critical value of d, specific for each firm, at which a FDI strategy is not as profitable as an export strategy. When this critical level is reached, this specific firm will undertake export activities, abandoning FDI.*

Remark 2 *If the above conditions are not respected, we will observe a switch from the X-mode to the M-mode.*

Remark 3 *If we aggregate all foreign markets into one, we would observe the HMY association between firm level efficiency and supply mode. However, when considering a single firm with efficiency* $(a')^{1-\sigma}$*, we see that this firm supplies nearby markets by M-mode, but further away markets by X-mode.*

2.4.3 Free Entry

It is possible to describe the equilibrium which characterizes this economy. In order to do so, we need to specify some other equilibrium equations, namely the free entry condition and the price index.

Free entry ensures equality between the expected operating profits of a potential entrant and the entry cost, $E(\pi) - f_I$. This condition holds for all type of firms. The cumulative density function is $G(a)$, with support: $[0, \dots, a_0]$, where for simplicity we can set $a_0 = 1$. The free entry condition can be defined as:

$$\int_0^{a_D} \pi_D dG(a) + 2\sum_{d=1}^{\frac{N-1}{2}} \{ \int_{a_{M,d}}^{a_{X,d}} \pi_{X,d} dG(a) + \int_0^{a_{M,d}} \pi_{M,d} dG(a) \} = f_I$$

Using the profit conditions (6)-(8), we obtain:

$$\int_0^{a_D} [\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} \frac{\beta E a^{1-\sigma}}{P^{1-\sigma}\sigma} - f_D] dG(a) + 2\sum_{d=1}^{\frac{N-1}{2}} \{ \int_{a_{M,d}}^{a_{X,d}} [\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} \frac{\phi\gamma a^{1-\sigma}\beta E}{P^{1-\sigma}\sigma} - f_X] dG(a) +$$

$$\int_{0}^{a_{M,d}} [\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} \frac{(\phi\gamma)^{\eta}\,\beta E a^{1-\sigma}}{P^{1-\sigma}\sigma} - f(d) - f_M]dG(a)\} \; = f_I \tag{12}$$

where $\phi = \tau^{1-\sigma}$ is freeness of trade, $\gamma = d^{1-\sigma}$ and d is the parameter that takes into consideration the different country locations; finally $P^{1-\sigma}$ is a weighted average of the marginal costs of all firms active in the market. Let's spend some more words on this term, $P^{1-\sigma}$.

In every country this weighted average, $P^{1-\sigma}$, is characterized by all the brands offered in that particular country. Brands offered by domestic firms, for which the consumer price is $a\sigma/\left(\sigma-1\right)$; brands offered by foreign exporters, for which the consumer price is $a\sigma d\tau/\left(\sigma-1\right)$; and finally, brands supplied by foreign subsidiaries, with consumer price $a\sigma\left(\tau d\right)^{\eta}/\left(\sigma-1\right)$. Therefore:

$$P^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} n \int_{0}^{a_D} a^{1-\sigma} dG(a/a_D) +$$

$$\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)} 2n \sum_{d=1}^{\frac{N-1}{2}} \left[\int_{0}^{a_{M,d}} [\phi\gamma]^{\eta}\, a^{1-\sigma} dG(a/a_D) + \int_{a_{M,d}}^{a_{X,d}} \phi\gamma a^{1-\sigma} dG(a/a_D) \right] \tag{13}$$

where n represents the measure of varieties available in the country. Notice that using (13) in (12) will make disappear the term $\left(\frac{\sigma}{\sigma-1}\right)^{(1-\sigma)}$.

2.4.4 Parameterization: Pareto Distribution

The fact that the free entry condition and the price index depend on the probability distribution implies that in order to have explicit solutions for this model, we need to assume a particular functional form for G(a). Following the empirical literature on firm size distribution (see Axtell 2001 and HMY), we use as an approximation the Pareto distribution. The cumulative distribution function of a Pareto random variable a is:

$$G(a) = \left(\frac{a}{a_0}\right)^{k} \tag{14}$$

where k and a_0 are the shape and scale parameter, respectively. Note that k=1 implies a uniform distribution on $[0, a_0]$. The shape parameter k represents the dispersion of cost draws. An increase in k would imply a reduction in the dispersion of firm productivity-draws. Hence, the higher is k the smaller is the amount of heterogeneity.

The support of the distribution $[0, ..., a_0]$, is identical for every country, where a_0 represents the upper bound of this distribution. The productivity distribution of surviving firms will also be Pareto with shape k. More precisely, since a firm will start producing only if it has at least a productivity of $1/a_D$, the probability distribution of supplying as an exporter, or as a foreign affiliate, is conditioned on the probability of successful entry in each market. Hence the truncated cost distribution is given by:

$$G(a/a_D) = \left(\frac{a}{a_D}\right)^k$$

where is exploited the fractal nature of the Pareto. Here the support is $[0, ..., a_D]$. Given the assumed parameterization, we can explicitly solve $P^{1-\sigma}$ and the free entry condition.

Price Index using the Pareto distribution

As we said, firms will offer a price only if they have at least a productivity of $1/a_D$. Hence, the cumulative distribution is defined on a support $[0, ..., a_D]$. Solving for the price index we will obtain:

$$P^{1-\sigma} = \frac{n}{1-\frac{1}{b}} a_D^{1-\sigma} [1 + 2T^{1-b} \sum_{d=1}^{\frac{N-1}{2}} (\phi\gamma_d)^b + 2 \sum_{d=1}^{\frac{N-1}{2}} V_d^{1-b} \left[(\phi\gamma_d)^\eta - \phi\gamma_d\right]^b \tag{15}$$

where $b = \frac{k}{\sigma-1}$; $\phi = \tau^{1-\sigma}$; $\gamma = (d)^{1-\sigma}$; $T = f_X/f_D$ and $V_d = (f(d) + f_M - f_X)/f_D$. In order for the integral to converge we assume that $b > 1$.

Free entry condition using the Pareto distribution

Rewriting the free entry condition using the Pareto distribution we obtain:

$$\frac{E}{\sigma P^{1-\sigma}} [\int_0^{a_D} (a^{1-\sigma} - f_D) dG(a) + 2 \sum_{d=1}^{\frac{N-1}{2}} \int_0^{a_{M,d}} (a^{1-\sigma} (\phi\gamma)^\eta - (f(d) + f_M) dG(a) +$$

$$+ 2 \sum_{d=1}^{\frac{N-1}{2}} \int_{a_{M,d}}^{a_{X,d}} (a^{1-\sigma} (\phi\gamma) - f_X) dG(a)] = f_I \tag{16}$$

We can now use (15), (16), and (9)-(11) to obtain closed form solutions.

3 General Equilibrium with N countries

In order to analyze the main implications of our model, we exploit the fact that all fixed coefficients are the same in every country and that the distribution function is the same. However, the existence of N countries located evenly around a circular trade route introduces a role for distance in generating heterogeneity in the supply mode decisions within the same firm. Using the expression in (15) inside the domestic cutoff condition (9), we find the equilibrium number of varieties (and so the number of surviving firms) consumed in a typical nation:

$$n^* = \frac{(b-1)\beta E}{\sigma b f_D[1 + 2T^{1-b}\sum_{d=1}^{\frac{N-1}{2}} (\phi\gamma)^b + 2\sum_{d=1}^{\frac{N-1}{2}} V_d^{1-b}\left[(\phi\gamma)^\eta - \phi\gamma_d\right]^b]}$$

We define $\Omega = 2T^{1-b}\sum_{d=1}^{\frac{N-1}{2}} (\phi\gamma)^b$, and, on the other hand, $\Psi = 2\sum_{d=1}^{\frac{N-1}{2}} V^{1-b}\left[(\phi\gamma)^\eta - \phi\gamma\right]^b$. Where Ψ and Ω could be considered as parameters that summarize the impact of the two types of trade barriers on exports and FDI activities. In particular, Ω represents the combined effect of higher fixed costs and variable distance costs on the export strategy. While Ψ measures the role of the difference in these costs when choosing between a FDI strategy and an export strategy. Using Ψ and Ω, the expression for n^* could be then simplified to:

$$n^* = \frac{(b-1)\beta E}{\sigma b f_D\left[1 + \Omega + \Psi\right]} \tag{17}$$

The equilibrium number of firms described by (17) represents the actual number of survivors in each country, which decreases with Ψ and Ω, hence it decreases with higher fixed and variable distance costs. Using the free entry condition in (16), and the cutoff conditions in (9)-(11), we get explicit closed form solutions for a_D, $a_{X,d}$, and $a_{M,d}$. In particular,

$$a_D^* = a_0\left[\frac{(b-1)f_I}{(f_D(1+\Psi+\Omega))}\right]^{\frac{1}{k}} \tag{18}$$

Using (20) inside the ratio between (10) and (9) we find:

$$a_{X,d}^* = a_0\left[\frac{(b-1)f_I}{f_X(1+\Psi+\Omega)}(\phi\gamma)^b T^{1-b}\right]^{\frac{1}{k}} \tag{19}$$

Finally, using (20) inside the ratio between (11) and (9) we obtain the equilibrium cutoff for the M-mode is:

$$a_{M,d}^* = a_0\left[\frac{(b-1)f_I}{(1+\Psi+\Omega)}\left[(\phi\gamma)^\eta - (\phi\gamma)\right]^b \frac{V_d^{1-b}}{f(d) + f_M - f_X}\right]^{\frac{1}{k}} \tag{20}$$

The index d inside these expressions is related to the geographical distance between the origin and a specific destination country.

Differently to HMY, these cutoffs change in relation to the geographical location of the destination country. Indeed, equations (18)-(20) change in relation to the number of countries belonging to this trade bloc and more importantly, (19) and (20) change with respect to the destination country we consider to reach. Since countries are evenly spaced along the circular trade route, the above equations are the same for whatever country we pick to be the origin country.

Remark 4 *The existence of different country locations generates distance-dependent cutoffs. Hence, the range of firms choosing the M-mode is shrinking the more distant is the destination country to reach. Therefore, some firms that are supplying the foreign country, j, via local sales of foreign affiliates (M-mode), could be forced to supply country $j+1$ via X-mode, where $d(j+1) > d(j)$.*

3.1 The Impact of Trade

In what follows, we observe the effect of opening to trade. Since f_I does not change in the transition from autarky to trade, the free entry condition is left unaffected by trade: regardless of profit differences across firms (relative to X-mode or M-mode), the expected value of future profits, in equilibrium, must be equal to the fixed investment cost f_I. Hence, in a graph where we represent the evolution of profits in relation to the marginal cost, the transition from autarky to open economy will move up the zero profit condition curve: the exposure to trade induces an increase in the cutoff productivity level, $((1/a_D)^T > (1/a_D)^A)$[13]. This will modify the productivity level of the least productive firms. In an open economy situation, a firm with a productivity level between $(1/a_D)^A$ and $(1/a_D)^T$ cannot earn positive profits and so will exit the market. Moreover, as discussed in Melitz (2003), another selection process acts: firms with productivity level above $1/a_X$ or above $1/a_M$ enter respectively as exporters or as subsidiary. These three effects are called *domestic market selection effect*, *export market selection effect* and *FDI market selection effect*. These effects reallocate market shares toward more efficient firms, and generate an increase in the overall productivity.

The transition toward the open economy situation generates a reduction in the number of surviving firms in every country[14]. The total number of firms selling in every country includes: total number of domestics firms, foreign exporters and multinationals. The number of surviving firms decreases as a consequence of the domestic market selection effect ($a_D \downarrow$). However, as the entrance of new foreign firms more than compensate this reduction, consumers typically enjoy a larger amount of varieties.

[13] Recall that in Melitz (2002) the ZPC are downward sloping and the FE conditions are upward sloping. However, since here we consider the marginal costs, the slope of these curves will be the opposite.

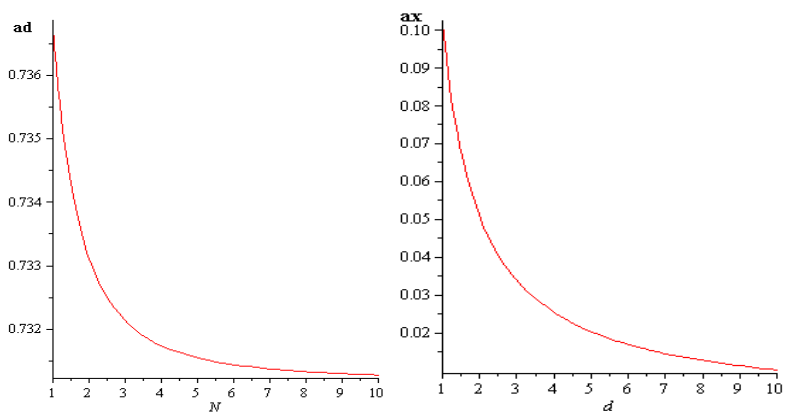
[14] As in Melitz, $n < n_A$, where n_A represents the number of firms in autarky.

3.2 The Effects of an Increase in Distance and Trade Liberalization

We now consider comparative statics with respect to d, N and ϕ: we study how these elements affect the equilibrium marginal costs. We have analytical solutions, but the analysis is facilitated by graphing the changes with respect to d, N and ϕ.

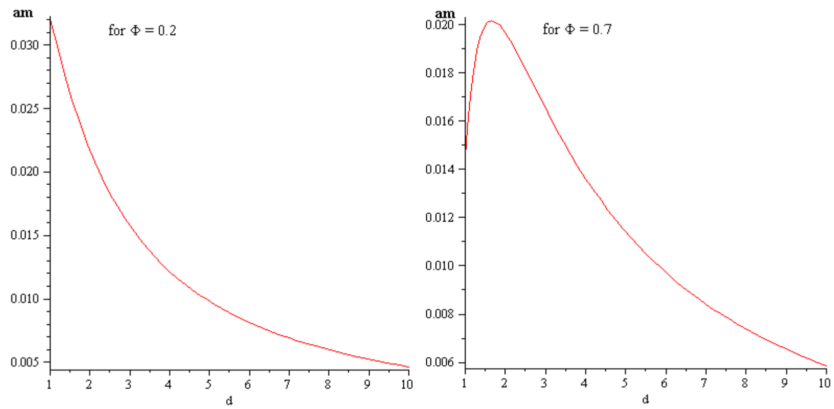
Increase in N and in Distance Since the domestic cutoff does not depend on distance, we examine the effect of a change in the dimension of the trade area, N, on a_D. Then we consider the effect of distance on the export cutoffs. These effects, as shown in Figure 3, are unambiguous.

Figure 3: Increase in N and in Distance



An increase in the overall dimension of the circle, N, implies a decrease in the domestic cutoff. In fact, the increase in the dimension of the economic area generates an increase in competition, and so in the expected profits. Moreover, the increase in distance between the origin and the destination country makes more expensive to reach the destination country through export. Therefore, the cutoff of the exporting firm is decreasing with distance. This means that when distance increases only the more productive firms can afford to export.

The effect of distance on MNF cutoff, $a_{M,d}$, is more complex; as mentioned in remark 2, it depends on the degree of openness.

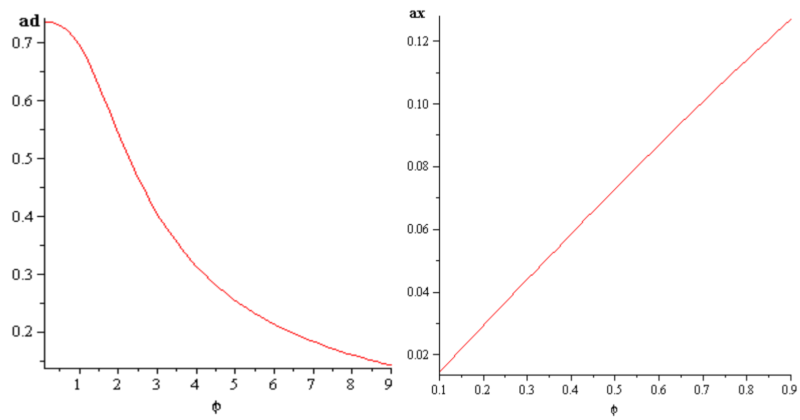


If ϕ is sufficiently low, the MNF cutoff, $a_{M,d}$, is a monotonic decreasing function of distance, d. This implies that the increase in distance is making it more difficult to choose M-mode in order to reach the foreign country. Differently, when ϕ is high, the MNF cutoff function does not have a monotonous behaviour: it increases for low level of distance and then it starts to decrease when distance becomes important. A possible interpretation of this result is that a large value of ϕ lowers the effects of distance on $a_{M,d}$.

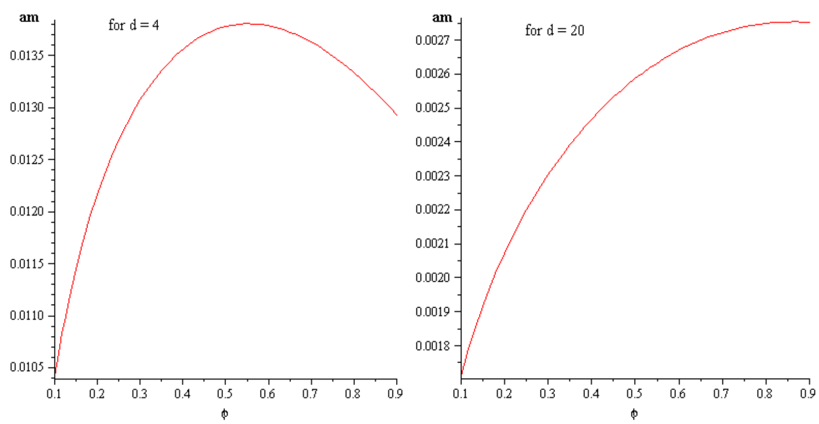
The intuition behind distance related comparative statics is that a decrease in distance reduces the fixed cost of FDI, but it also reduces the marginal costs of FDI and exporting. The proportional decrease in the marginal cost of exporting should be higher. However, it is possible for the reduction in the fixed cost of FDI to dominate, in which case FDI would become more attractive.

Increase in Trade Openness We now consider how a progressive exposure to trade affects the supply mode decisions of firms, via the effect on the equilibrium cutoffs conditions.

Figure 4: Increase in the freeness of trade



The increase in the exposure to trade, ($\phi \uparrow$ or $\tau \downarrow$), on the domestic and export cost cutoffs produces an effect similar to the one in Melitz (2003). The domestic cutoff is decreasing as a consequence of the market selection effect, and the export cutoff is increasing as a consequence of the reduction in τ. More complex results are obtained for what concern the MNF cost cutoff:

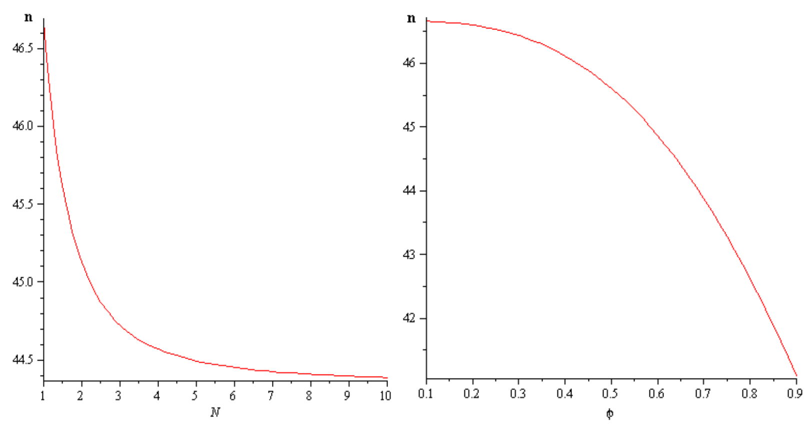


For a low level of distance, the MNF cost cutoff is not a monotonic function of ϕ. It increases for a low level of ϕ, and then it decreases. Hence, when distance is not too high, an initial increase in trade openness makes it easier to become a MNF. However, a further increase in trade openness makes the foreign market strategy too costly: the productivity required to supply via M-mode is now increasing with ϕ. On the contrary, when distance is sufficiently high, the MNF cost cutoff is a monotonic increasing function of phi. This last result confirms the classical MNF theory.

Why does a decrease in trade cost make FDI more attractive? We should remember that the marginal cost under FDI is always lower than the marginal cost under exports. Also, the fixed cost of FDI is higher than the fixed cost of exports. Further, the fixed cost of FDI is increasing in distance. An increase in ϕ should decrease the marginal cost of exporting more than the marginal cost of FDI, leaving unaffected the fixed costs. Therefore, while the ratios of the two fixed costs stay the same, the marginal cost of exporting decreases more than the marginal cost of FDI. Nevertheless, the final effect depends on the level of distance and on the share of the imported input in production. These two elements can allow FDI to become more attractive.

Number of Firms Also, the equilibrium number of firms does not depend on distance; therefore we examine the effect of a change in the dimension of the trade area, N and in trade openness, ϕ, on the number of active firms.

Figure 5



As a consequence of the increase in competition ($N \uparrow$), the number of active firms decreases. In the same way, a further exposure to trade reduces the number of active firms as a consequence of the increase of the domestic productivity cutoff (market selection effect).

3.3 Welfare Effects of Trade Liberalization

From the indirect utility function we can examine the welfare of consumers. Since the indirect utility function is given by $V = \beta E/P$, where P is the standard CES price index, we can examine the welfare effects simply by examining how P changes[15]. A greater openness will increase the welfare by lowering

[15] Without loss of generality, in this welfare analysis we are only concerned about the differentiated good.

the price index, as well as a decrease in the domestic cutoff[16]. Instead, a higher distance increases the price index, lowering welfare.

4 Distance and Aggregate Sales

In the following part we will use the analysis presented in Kleinert and Toubal (2006) in order to consider the role of distance on volume of sales. Following the standard scale versus proximity hypothesis, horizontal FDI is chosen when firms prefer to produce abroad through foreign affiliates so to save on trade costs. Hence, in the classical FDI framework we observe foreign affiliates' sales increase with trade costs. However, empirical findings based on aggregate data (Buch et al. 2005; Carr et al. 2001) find a negative relationship between affiliate sales and distance[17]. In order to fill this gap, we model a FDI activity which involves a local production of one intermediate good, y_2, and the imports of another intermediate good, y_1; so that it is possible to analyze how affiliate sales could be affected by distance. First of all, let's define the aggregate affiliate sales in the case of two countries:

$$\begin{aligned} S_A &= \int_0^{a_{M,d}} A a^{1-\sigma} \frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} (\phi\gamma)^\eta g(a) da \\ &= \left(\frac{a_{M,d}}{a_D} \right)^k a_{M,d}^{1-\sigma} \frac{k}{k-\sigma+1} A \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} (\phi\gamma)^\eta \end{aligned} \tag{21}$$

The first term, $(a_{M,d}/a_D)^k$, represents the cumulative probability of firms in the origin country to own an affiliate in the destination country. As we briefly mentioned in the previous section, if we multiply this term with the total mass of active firms from the origin country, $n^* G(a_{M,d}/a_D)$, we obtain the number of affiliates in the destination country. The remaining part of that expression represents average sales. Since we are dealing with N symmetric countries, and so with N-1 possible partners, the overall aggregate affiliate sales are:

$$S_A = 2 \sum_{d=1}^{\frac{N-1}{2}} \left(\frac{a_{M,d}}{a_D} \right)^k a_{M,d}^{1-\sigma} \frac{k}{k-\sigma+1} A \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} (\phi\gamma)^\eta \tag{22}$$

Without loss of generality, in the exercise of comparative statics, we consider only two countries; the analysis could be extended to N-1 countries, but the conclusions would not change. In observing the change in affiliate sales as a consequence of a change in d, we should keep in mind that the expression

[16] It can nevertheless happen that when trade costs are high and the number of foreign activities is strictly greater than the number of domestic firms, the effect of product varieties on welfare is negative (Melitz, 2002).

[17] In these empirical works, distance is used as a proxy for transportation costs.

above can be rewritten as:

$$S_A = \left(\frac{a_{M,d}}{a_D}\right)^k a_{M,d}^{1-\sigma} \frac{k}{k-\sigma+1} A \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left(\phi d^{1-\sigma}\right)^{\eta} \tag{23}$$

Deriving the expression above with respect to d, we realize that the way distance affects affiliate sales is double:

$$\frac{\partial S_A}{\partial d} = \underbrace{\frac{\partial S_A}{\partial d}}_{-} + \underbrace{\frac{\partial S_A}{\partial a_{M,d}}}_{+} \underbrace{\frac{\partial a_{M,d}}{\partial d}}_{?} \tag{24}$$

In the above expression we can see that distance affects S_A through a direct and an indirect effect. For what concern the direct effect, we see from (23) that there is a negative relationship between S_A and d. On the contrary, we have more ambiguous results for what concern the indirect effect. In particular the ambiguity is linked to the sign of the last partial derivative, $\partial a_{M,d}/\partial d$. Equation (23) shows us that $a_{M,d}$ is positively related to the number of affiliates producing in foreign countries, $n^* \left(\frac{a_{M,d}}{a_D}\right)^k$, while it is negatively related with the average size of foreign affiliates. Differentiating (23) with respect to $a_{M,d}$ we get:

$$\frac{\partial S_A}{\partial a_{M,d}} = \left(\frac{a_{M,d}}{a_D}\right)^k a_{M,d}^{-\sigma} k A \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} (\phi\gamma)^{\eta} > 0$$

The threshold marginal cost of being a MNF, $a_{M,d}$, is positively related to aggregate affiliate sales. Let's now analyze the other partial derivative in the second order effect of (24), i.e. the effect of distance on productivity. What is the sign of $\partial a_{M,d}/\partial d$? Firstly, we use the net operating profits condition of being a MNF with respect to an exporter. Hence,

$$\pi^*_{M,d}(a, A, \eta) = A a_{M,d}^{1-\sigma} \frac{1}{\sigma} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[(\phi\gamma)^{\eta} - (\phi\gamma)\right] - f(d) - f_M + f_X$$

where any particular functional form for $f(d)$ is assumed. Solving this expression by the threshold marginal cost, $a_{M,d}$, we get:

$$a_{M,d} = \left(\frac{f(d) + f_M - f_X}{B\left[(\phi d^{1-\sigma})^{\eta} - (\phi d^{1-\sigma})\right]}\right)^{\frac{1}{1-\sigma}}$$

where B is as defined before. It is now possible to derive this expression with respect to d. We find the following result:

$$\frac{\partial a_{M,d}}{\partial d} = \left[\frac{1}{1-\sigma} \left(\frac{f(d) + f_M - f_X}{B\left[(\phi d^{1-\sigma})^{\eta} - (\phi d^{1-\sigma}) \right]} \right)^{\frac{1}{1-\sigma}-1} \right] \times$$

$$\times \left[\frac{f'(d)}{B\left[(\phi d^{1-\sigma})^{\eta} - (\phi d^{1-\sigma}) \right]} - \frac{(f(d) + f_M - f_X)\,(1-\sigma) \left[\eta \frac{\left(\phi d^{(1-\sigma)}\right)^{\eta}}{d} - \phi d^{-\sigma} \right]}{B\left[(\phi d^{1-\sigma})^{\eta} - (\phi d^{1-\sigma}) \right]^2} \right] \overset{?}{\lesseqgtr} 0 \qquad (25)$$

The sign of the first term is straightforward: since $\sigma > 1$, the sign is negative; instead, the second term is more difficult to interpret. In fact, the sign of this second term depends on the behaviour of the following term:

$$(1-\sigma) \left[\eta \frac{\left(\phi d^{(1-\sigma)}\right)^{\eta}}{d} - \phi d^{-\sigma} \right] \qquad (26)$$

More precisely, the sign of this term is related to the degree of trade openness[18]. For low levels of trade openness, (26) is always negative. If it is so, the sign in the second bracket in (25) is positive[19]; thus the overall sign of (24) is negative: the affiliate sales are unambiguously decreasing with distance for low levels of trade openness. When ϕ starts to increase the sign of (24) is not so straightforward anymore, since the partial derivative $\partial a_{M,d}/\partial d$ could be positive. For sufficiently high trade openness and sufficiently small distance, the sign of (26) will be positive; hence the overall sign of (24) depends on the magnitude of its first and second term. However, when distance is not so small, the sign of (24) turns again to negative, since (25) is now negative.

It seems interesting to notice some peculiarities linked to higher level of trade openness. The change in the sign of (25) depends on the degree of trade openness. For example, when ϕ is very high, the change in the sign of (25) occurs at a higher level of distance than for a lower ϕ. Therefore, since overall aggregate sales are positively related to the threshold marginal cost, we conclude that when trade barriers are sufficiently high, aggregate sales are decreasing in distance: the overall effect of distance on S_A is negative. However, the magnitude of this reduction is strictly linked to the level of trade openness. Hence, when distance is sufficiently high, we expect more firms choosing to supply via X-mode. The size of the reduction in affiliates' sales and so of the increase in export strategies, becomes bigger the more distant is the affiliate's locations. This result is in line with recent empirical

[18] See Appendix 2 for a graphical representation.

[19] With respect to the second bracket, since the communication costs are increasing with distance and the denominator is smaller in the second ratio, the second term is bigger than the first.

findings.

Remark 5 *The reduction in affiliate sales due to increased distance is more relevant the less open trade is.*

It could be interesting to turn our attention to the role of distance on export activity, so as to compare the effect of distance on M-mode supply versus X-mode supply. Since export sales are affected by a combination of $a_{M,d}$ and $a_{X,d}$, we expect a complex relationship between S_X and distance. In order to put in evidence the effect of distance, we first define the aggregate export sales in the case of two countries:

$$\begin{aligned} S_X &= \int_{a_{M,d}}^{a_{X,d}} A\left(\phi\gamma\right) a^{1-\sigma} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \\ &= A\left(\phi d^{1-\sigma}\right)\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\frac{k}{a_D^k}\frac{\left(a_{X,d}^{k-\sigma+1}-a_{M,d}^{k-\sigma+1}\right)}{k-\sigma+1} \end{aligned} \tag{27}$$

The effect of distance on the aggregate exports is a combination of different effects,

$$\frac{\partial S_X}{\partial d} = \underbrace{\frac{\partial S_X}{\partial d}}_{-} + \underbrace{\frac{\partial S_X}{\partial a_{M,d}}}_{-}\underbrace{\frac{\partial a_{M,d}}{\partial d}}_{?} + \underbrace{\frac{\partial S_X}{\partial a_{X,d}}}_{+}\underbrace{\frac{\partial a_{X,d}}{\partial d}}_{-} \tag{28}$$

In what follows we confirm the sign of the above partial derivatives. Deriving S_X with respect to $a_{X,d}$ and $a_{M,d}$ we find that:

$$\frac{\partial S_X}{\partial a_{M,d}} = -\frac{\left(\phi\gamma\right)\frac{1}{a_D^k}a_{M,d}^{k-\sigma+1}k}{a_{M,d}}A\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} < 0 \tag{29}$$

and

$$\frac{\partial S_X}{\partial a_X} = \frac{\left(\phi\gamma\right)\frac{1}{a_D^k}a_{X,d}^{k-\sigma+1}k}{a_{X,d}}A\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} > 0 \tag{30}$$

As expected, export sales are increasing in the threshold marginal cost of being an exporter, $a_{X,d}$, and decreasing in $a_{M,d}$. In order to analyze the relationship between aggregate export sales and distance, we will use the net operating profits, so to derive the effect of distance on both the threshold $a_{M,d}$ and $a_{X,d}$. We already know from the analysis above that the sign of $\partial a_{M,d}/\partial d$ is ambiguous: it depends on the degree of trade openness. Thus, as long as distance is not too small the cutoff marginal cost

$a_{M,d}$ is negatively affected by distance, hence $\partial a_{M,d}/\partial d < 0$. On the other side for what concern $a_{X,d}$

$$a_{X,d} = \left(\frac{f_X}{B\left(\phi d^{1-\sigma}\right)} \right)^{\frac{1}{1-\sigma}}$$

hence

$$\frac{\partial a_{X,d}}{\partial d} = -\frac{\left(\frac{f_X}{B\left(\phi d^{1-\sigma}\right)} \right)^{\frac{1}{1-\sigma}}}{d} < 0 \tag{31}$$

the effect of distance, d, on the cutoff marginal cost $a_{X,d}$ is unambiguously negative. What could be concluded? Similarly to what we found for affiliate sales, the amount of aggregate exports does not linearly depend on distance. The effect of distance on S_X depends on the magnitude of the partial derivatives in (28). Nevertheless, when $\partial a_{M,d}/\partial d > 0$, since the sign of (30) is positive, so the overall sign of (28) will be negative[20]. The economic intuition behind this result is that, for high trade openness and low distance, the result of a slight increase in distance between the destination and the origin countries, is a decrease in the amount of export sales. Hence, when distance does not play an important role and trade is sufficiently open, the result obtained could be considered as a confirmation of the scale versus proximity hypothesis. On the contrary, when distance becomes important we cannot have a precise conclusion, because the second term in (28) will be positive, since $\partial a_{M,d}/\partial d < 0$. For high levels of distance, if the first and the third terms in (30) are smaller than the second term, the overall sign of that expression will be positive; meaning that when distance plays an important role the export sales are increasing with distance. To conclude:

Remark 6 *Under certain circumstances, namely sufficient trade openness, high distance and important reaction of S_X to changes in $a_{M,d}$, export sales are increasing with distance.*

5 Conclusion

In this paper we extend the standard Markusen MNF setting to enrich the ways distance can affect the choice between exporting and local production supply modes. We assume that distance affects the cost of the local production option in two ways. First, in a setting with N symmetric countries located evenly around a circular trade route, we assume that an essential intermediate good must be produced at home due to issues of intellectual property protection. Thus, distance from the home market raises the cost of local production but by less than it does for the exporting option. Second,

[20] When trade is sufficiently open and distance sufficiently low, the sign of $\partial a_M/\partial d$ will be positive.

we assume the cost of supplying a foreign market via local production rises with market's distance due to coordination costs. These elements allow a richer range of possible mode-of-supply outcomes. For instance, an increase in distance raises the cost of FDI, but it also raises the marginal costs of FDI and exporting. Depending upon the magnitudes, FDI may become the more attractive option for more distant markets. However, it is possible for the increase in the cost of FDI to dominate, in which case FDI would become less attractive with distance. Hence, we might see firms supply nearby markets with FDI and more distant market via exports. In line with recent empirical work, we show a richer pattern of modes of supply to foreign markets. Moreover, since we allow for firms heterogeneity in production costs, the fixed and variable costs involved in the FDI versus exporting decision will weigh differently large and small firms in a market at a given distance (as in HMY). But additionally, changes in distance also have heterogeneous effects.

To summary, this paper makes four contributions to the existing literature. First, we enrich the spatial pattern of FDI, so that it depends on firm characteristics. This generates a more complex outcome than the standard Markusen model. Second, by introducing heterogeneity by firms by markets we highlight a process through which one firm may supply some markets via FDI and others via exports. Third, we shed light on the non-monotonic relation between distance and the FDI supply-mode cutoff, which allows us to find a switching behaviour between FDI versus export modes of supply. Fourth, the predictions of the model are in line with the empirical evidence that shows a negative relation between distance and affiliate sales. Nevertheless, an empirical analysis to check how distance and trade openness impact FDI and whether the estimated patterns are indeed non-monotonic and interactive, as the model suggests, is needed.

In this paper we consciously avoided the literature on export platform FDI, since we did not allow for the possibility of re-export from a foreign affiliate. However, stepping slightly outside the model, Figure 2 makes it quite clear that the area after $a_{M,1}^{1-\sigma}$ could be interpreted in terms of an export platform. As we said, when distance increases, some firms stop building foreign affiliates abroad and start to undertake export as a foreign market access strategy. This export activity might be cheaper if it takes place between the last foreign country where it has been built the foreign affiliate and the new destination country. This latter case would imply an export platform strategy, where the foreign affiliate firm located in country j sells in that foreign domestic market, and also in third markets (j+1, j+2, ..., j+(N-1)/2) through export. In an extension of this paper we would like to model this export platform strategy.

References

[1] Alfaro, L., Charlton, A., 2007. Intra-Industry Foreign Direct Investment. CEP Discussion Paper No 825.

[2] Anderson, J.E., 1979. A Theoretical Foundation for the Gravity Equation. American Economic Review, 69 (1), 106-16.

[3] Antras, P., 2003. Firms, Contracts, And Trade Structure. The Quarterly Journal of Economics, 118 (4), 1375-1418.

[4] Antras, P., Helpman, E., 2004. Global Sourcing. Journal of Political Economy, 112 (3), 552-580

[5] Antras, P., Garicano, L., Rossi-Hansberg, E., 2006. Offshoring in a Knowledge Economy. The Quarterly Journal of Economics, 121 (1), 31-77.

[6] Axtell, R.L., 2001. Zipf Distribution of U.S. Firm Sizes. Science 293, September, 1818–20.

[7] Baldwin, R.E., Robert-Nicoud, F., 2007. Trade and growth with heterogeneous firms. Forthcoming in Journal of International Economics.

[8] Baldwin, R.E., 2005. Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model. NBER Working Paper No. 11471, National Bureau of Economic Research, Inc.

[9] Baltagi, B.H., Egger, P., Pfaffermayr, M., 2006. Estimating models of complex FDI: are there thirdcountry effects?. Journal of Econometrics, 140 (1), 260-281.

[10] Barba Navaretti, G., et al., 2004. Multinational Firms in the World Economy. Princeton, NJ: Princeton University Press.

[11] Bernard, A., Jensen, J. B., Redding, S., Schott, P., 2007. Firms in International Trade. Journal of Economic Perspective, 21 (3), 105-130.

[12] Blonigen, B.A., Davies, R.B., Naughton, H.T., Waddell, G.R., 2007. FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment. European Economic Review, 51 (5), 1303-1325.

[13] Brainard, S.L., 1993. A Simple Theory of Multinational Corporations and Trade with a Trade-Off Between Proximity and Concentration. NBER Working Papers 4269, National Bureau of Economic Research, Inc.

[14] Brainard, S.L., 1997. An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade. American Economic Review, 87 (4), 520-44.

[15] Buch, C.M., Kleinert, J., Lipponer, A., Toubal, F., 2005. "Determinants and effects of foreign direct investment: evidence from German firm-level data," Economic Policy, CEPR, CES, MSH, 20 (41), 52-110.

[16] Carr, D., Markusen, J.R., Maskus, K.E., 2001. Estimating the Knowledge-Capital Model of the Multinational Enterprise. American Economic Review, 91 (3), 693-708.

[17] Coase, R. (1937), The Nature of the Firm. Economica, 4 (16), 386-405.

[18] Dunning, J. H., 1994. Multinational enterprises and the globalization of innovatory capacity. Research Policy, Elsevier, 23 (1), 67-88.

[19] Ekholm, K., Markusen, J., Forslid, R., 2005. Export-Platform Foreign Direct Investment. Journal of European Economic Association, 5 (4), 776-795.

[20] Ethier, W.J, 1986. The Multinational Firm. The Quarterly Journal of Economics, 101 (4), 805-33.

[21] Feenstra, R.C., 1998. Integration of Trade and Disintegration of Production in the Global Economy. Journal of Economic Perspectives, 12 (4), pages 31-50.

[22] Grossman, G.M., Helpman, E., Szeidl, A., 2006. Optimal integration strategies for the multinational firm. Journal of International Economics, 70 (1), 216-238.

[23] Hanson, G.H., Mataloni, R.J., Slaughter, M.J., 2001. Expansion Strategies of U.S. Multinational Firms. NBER Working Papers 8433, National Bureau of Economic Research, Inc.

[24] Head, K.,Ries, J., 2003. Heterogeneity and the FDI versus export decision of Japanese manufacturers. Journal of the Japanese and International Economies, 17 (4), 448-467.

[25] Head, K., Mayer, T., 2004. Market Potential and the Location of Japanese Investment in the European Union. The Review of Economics and Statistics, 86 (4), 959-972.

[26] Helpman, E., 1984. A Simple Theory of International Trade with Multinational Corporations. Journal of Political Economy, 92 (3), 451-71.

[27] Helpman, E., Melitz, M.J., Yeaple, S.R., 2004. Export versus FDI with Heterogeneous Firms. American Economic Review, 94 (1), 300-316.

[28] Horstmann, I.J., Markusen, J.R., 1992. Endogenous market structures in international trade (natura facit saltum). Journal of International Economics, 32 (1-2), 109-129.

[29] Hummels, D., Ishii, Jun, Yi, K., 2001. The nature and growth of vertical specialization in world trade. Journal of International Economics, 54 (1), 75-96.

[30] Kleinert, J., Farid, T., 2006. Distance costs and Multinationals' foreign activities. CEI Working Paper.

[31] Krugman, P., 1980. Scale Economies, Product Differentiation, and the Pattern of Trade. American Economic Review, 70 (5), 950-59.

[32] Markusen, J.R., 1984. Multinationals, multi-plant economies, and the gains from trade. Journal of International Economics, 16 (3-4), 205-226.

[33] Melitz, M. J., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica, 71 (6), 1695-1725.

[34] Melitz, J.M., Ottaviano, G.I.P., 2005. Market Size, Trade, and Productivity. NBER Working Papers 11393, National Bureau of Economic Research, Inc.

[35] Neary, Peter J., 2006. Trade Costs and Foreign Direct Investment. CEPR Discussion Papers 5933, C.E.P.R. Discussion Papers.

[36] Ottaviano, G.I.P., Tabuchi, T., Thisse, J., 2002. Agglomeration and Trade Revisited. International Economic Review, 43 (2), 409-436.

[37] Tybout, R.J., 2001. Plant- and Firm-Level Evidence on "New" Trade Theories. NBER Working Papers NBER Working Paper 8418, National Bureau of Economic Research, Inc.

[38] Williamson, Oliver E., 1985. The Economic Institutions of Capitalism, New York: Free Press.

[39] Yeats, A.J., 2001. Just How Big is Global Production Sharing?. In: Arndt, Sven W. And Henryk Kierzkowski, eds., Fragmentation: New Production Patterns in the World Economy (Oxford: Oxford University Press).

6 Appendix

A1. Cost Minimization Problem In order to find the equilibrium operating profits, we solve the minimization problem of the firm. For example, the cost minimization problem for foreign affiliates:

$$\min_{y_1,y_2,\lambda} L = y_1 d\tau + y_2 + \lambda \left[x(v) - \frac{1}{a}\left(\frac{y_1}{\eta}\right)^{\eta}\left(\frac{y_2}{1-\eta}\right)^{1-\eta} \right]$$

where the Lagrangian multiplier λ represents the marginal cost of production. The Hicksian factor demands are:

$$y_1^* = x(v)\, a\eta \left(\frac{1}{d\tau}\right)^{1-\eta}$$

$$y_2^* = x(v)\, a(1-\eta)(d\tau)^{\eta}$$

Using the Hicksian demands, we can write the total cost of a subsidiary as a function of the final output, $x(v)$:

$$\begin{aligned} TC_{M,j} &= y_1^* d\tau + y_2^* + f(d) + f_M \\ &= x(v)\, a(d\tau)^{\eta} + f(d) + f_M \end{aligned} \tag{32}$$

Using (32) inside (5) it is possible to derive an expression for the multinational equilibrium profits, which depends only on the final output $x(v)$:

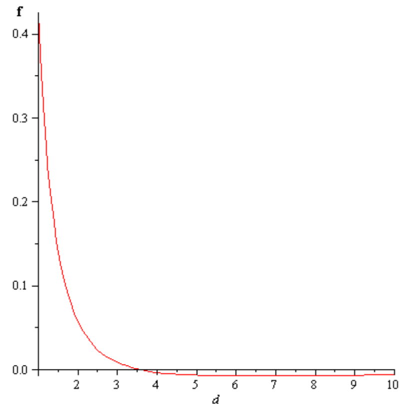
$$\pi_{M,d}(a, A, \eta) = A^{i\,\frac{1}{\sigma}} x(v)^{\frac{\sigma-1}{\sigma}} - x(v)\, a(d\tau)^{\eta} - f(d) - f_M \tag{33}$$

hence the optimal output for the affiliate located in the foreign country is:

$$x(v)^* = \frac{A^i \left(\frac{\sigma-1}{\sigma}\right)^{\sigma}}{\left(a(d\tau)^{\eta}\right)^{\sigma}} \tag{34}$$

Equations (33) and (34) refers to this specific multinational framework; the problem above can be solved for each different type of firm. More generically, the final good producer will choose the supply mode that maximizes $\pi_k^*(a, A, \eta)$ where $k = M$, X or D. For this reason, final good producers organize the production so as to minimize both variable and fixed costs.

A2. Distance and Aggregate sales The sign of $f = (1-\sigma)\left[\eta \frac{\left(\phi d^{(1-\sigma)}\right)^{\eta}}{d} - \phi d^{-\sigma}\right]$ changes in relation to the degree of trade openness. In particular, when $\phi = 0.9$



while when $\phi = 0.2$, its behaviour is the following: