FIRM SELECTION INTO EXPORT-PLATFORM FOREIGN DIRECT INVESTMENT*

Monika Mrázová[†]
MES and CEP,
London School of Economics

J. Peter Neary[‡]
University of Oxford
and CEPR

February 15, 2011

Abstract

We characterize how firms select into exporting or foreign direct investment, and how many plants they choose to establish, in a group of foreign countries. We show that, if and only if firms' maximum profits are supermodular in tariffs and production costs, then: the most efficient firms establish a branch plant in each country; firms of intermediate efficiency establish only one plant as an export platform; while the least efficient firms export. The results apply in a range of models and under a variety of assumptions about market structure, though R&D with threshold effects provides an interesting counter-example.

Keywords: Export-Platform FDI; Heterogeneous Firms; Proximity-Concentration Trade-Off; R&D with Threshold Effects; Supermodularity.
JEL Classification: F23, F15, F12

*We are grateful to Wolfgang Keller for advice on data sources, and to Paola Conconi, Arnaud Costinot, Gene Grossman, Dermot Leahy, Marc Melitz, Giordano Mion, Lindsay Oldenski, Emanuel Ornelas and Gianmarco Ottaviano for helpful comments. Monika Mrázová wishes to thank the ESRC [grant number PTA-026-27-2479].

[†]Managerial Economics and Strategy Group, London School of Economics, Houghton Street, London WC2A 2AE, UK; e-mail: m.mrazova@lse.ac.uk; URL: http://www.monikamrazova.com.

[‡]Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK; e-mail: peter.neary@economics.ox.ac.uk; URL: http://www.economics.ox.ac.uk/members/peter.neary/neary.htm.

1 Introduction

The incentives for a firm to serve a single foreign market through foreign direct investment (FDI) rather than exports are well understood from the "proximity-concentration trade-off" hypothesis, which can be traced back to the writings of Haberler (1936, 273–278) on "tariff factories". A large empirical literature broadly confirms the hypothesis that firms trade off the improved access to a foreign market stemming from proximity as a result of FDI, on the one hand, against the saving on fixed costs of concentrating production in their home plant and serving the foreign market by exports, on the other hand. The question of which firms will select the different modes of serving foreign markets has also been considered by Helpman, Melitz, and Yeaple (2004). They show that more productive firms are better able to engage in FDI and enjoy the benefits of serving foreign consumers at lower variable costs. However, they prove this result in only one framework, albeit a canonical one, that of free-entry monopolistic competition with CES preferences and a Pareto distribution of firm productivities. Whether the result holds more broadly is an open question. Furthermore, the incentives that drive export-platform FDI and the characteristics of firms that engage in it are even less well understood. This matters since export-platform FDI is quantitatively important. For example, in 2004, sales to other foreign countries accounted for 33.7% of the $2,387 billion worth of total foreign sales by foreign affiliates of U.S. firms.[1] Yet apart from a brief discussion in Helpman, Melitz, and Yeaple (2003), the issue of which firms will engage in export-platform FDI has received very little attention.[2]

In this paper we illuminate these issues by deriving a general result which characterizes the conditions under which a natural pattern of firm selection between different modes of serving foreign markets emerges. Building on Neary (2002, 2009), we develop a general model of how a firm will choose to serve a group of foreign markets by exports or FDI,

[1]The data are from the U.S. Bureau of Economic Analysis. The figures are: $1,583 billion "Total Local Sales" and $804 billion "Total Other Foreign Sales". In addition, U.S. sales by U.S.-owned foreign affiliates totalled $301 billion: these represent vertical FDI, which we do not consider in this paper. We also follow the bulk of the trade literature in concentrating on greenfield FDI, though much of FDI consists of cross-border mergers. See Neary (2009) and the references cited there for further discussion.

[2]See the appendix to Helpman, Melitz, and Yeaple (2003), the working paper version of Helpman, Melitz, and Yeaple (2004). We discuss this further in Section 6.1 below.

and, if by the latter, how many foreign plants it will want to establish.[3] We then derive a necessary and sufficient condition under which the most efficient firms establish a branch plant in each country; firms of intermediate efficiency establish one plant only and use it as an export platform; while least efficient firms export from their home country. Finally, we show that our results apply under a wide variety of assumptions about firm behaviour and market structure.

From a technical point of view, the paper contributes to the small but growing number of papers which uses the concept of supermodularity to illuminate issues in international trade.[4] Supermodularity arises very naturally in our context. Our interest is in comparing firms whose costs differ by a finite amount, and in particular in comparing their behaviour under different modes of serving foreign markets, in which the tariffs and transport costs they incur also differ by a finite amount. Supermodularity imposes a natural restriction on the finite "difference-in-differences" of the firm's profit function which we need to sign in order to make this comparison. As we show, the profit function exhibits supermodularity under a wide range of assumptions, and allows us to generalize our results with remarkably few restrictions on technology, tastes, or market structure.

The plan of the paper is as follows. Sections 2 to 5 focus on a single monopoly firm which faces the decision on how to serve a group of foreign markets. Section 2 introduces the setting and explains the restrictions implied by supermodularity. Sections 3 and 4 consider the relative profitability of different modes of serving the foreign markets, reviewing the cases of tariff-jumping and export-platform FDI respectively, while Section 5 derives our main result on how firms of different productivities will select into one or other mode. Section 6 then turns to applications, and shows that our approach applies in a wide range of contexts, including some of the most widely-used models in international trade. It also shows that supermodularity is not inevitable in all contexts. We introduce a new specification of investment costs subject to threshold effects and show that it may lead to a violation of supermodularity, and so to a reversal of the conventional assignment

[3]For other studies of export-platform FDI, see Motta and Norman (1996), Grossman, Helpman, and Szeidl (2006), and Ekholm, Forslid, and Markusen (2007).

[4]For other applications of supermodularity to international trade, see Grossman and Maggi (2000), Limão (2005), Costinot (2009), and Costinot and Vogel (2010).

of firms to different modes of accessing foreign markets.

2 Operating Profits and Supermodularity

Throughout we consider a firm located in country 0 (for concreteness, call it the U.S.) which contemplates serving consumers located in n symmetric foreign countries (for concreteness, call them the EU). The maximum operating profits the firm can earn in each of the n countries equal $\pi(t, c)$, where t is the access cost (tariffs and transport costs) it faces and c is an exogenous cost parameter, which we can think of as an inverse measure of productivity. The parameter c equals the firm's marginal production cost in many applications, though not in all: we will see exceptions in Example 2 and Section 6 below. Profits also depend on the firm's choice variables and on other exogenous variables. However, the former have been chosen optimally and so are subsumed into the π function (we give some examples below); while the latter are suppressed for convenience.

We define Δ_c as the finite difference between the values of a function evaluated at two different values of c. Applying this to the operating profit function π gives:

$$\Delta_c \pi(t, c) \equiv \pi(t, c_1) - \pi(t, c_2) \qquad \text{when} \qquad c_1 \geq c_2 \tag{1}$$

So, $\Delta_c \pi(t, c)$ is the profit loss of a higher-cost relative to a lower-cost firm. We assume that π is globally decreasing (though not necessarily continuous) in both t and c, so, under the convention we adopt, $\Delta_c \pi(t, c)$ is always non-positive. It is convenient to define it in this way, since it implies that, when $\pi(t, c)$ is differentiable in c, $\Delta_c \pi(t, c)$ reduces to the partial derivative $\pi_c(t, c)$ as c_1 approaches c_2.[5]

We can now define what we mean by supermodularity in the context of our paper:[6]

[5] We use subscripts to denote partial derivatives: e.g., $\pi_c = \partial \pi / \partial c$ and $\pi_{tc} = \partial^2 \pi / \partial t \partial c$.

[6] More generally, following Milgrom and Roberts (1990) and Athey (2002), supermodularity can be defined in terms of vector-valued arguments: $\pi(x \vee y) + \pi(x \wedge y) \geq \pi(x) + \pi(y)$, where $x \vee y \equiv \inf\{z \mid z \geq x, z \geq y\}$ and $x \wedge y \equiv \sup\{z \mid z \leq x, z \leq y\}$. This is equivalent to the definition in the text when we set: $x = \{c_2, t_1\}$ and $y = \{c_1, t_2\}$.

Definition 1. The function $\pi(t, c)$ is supermodular in t and c if and only if:

$$\Delta_c \pi(t_1, c) \geq \Delta_c \pi(t_2, c) \qquad \text{when} \qquad t_1 \geq t_2.$$

When the first inequality in the definition is reversed, we say that the function is submodular. Intuitively, supermodularity of π means that a higher tariff reduces in absolute value the cost disadvantage of a higher-cost firm. Putting this differently, the profit function exhibits the "Matthew Effect": "to those who have, more shall be given". Rewriting the definition we can see that supermodularity is equivalent to:

$$\pi(t_2, c_2) - \pi(t_1, c_2) \geq \pi(t_2, c_1) - \pi(t_1, c_1) \geq 0 \qquad \text{when} \qquad t_2 \leq t_1 \;\; \text{and} \;\; c_2 \leq c_1 \qquad (2)$$

Thus, when supermodularity holds, a lower tariff is of more benefit to a more productive firm. This might seem like the natural case, since a lower tariff contributes more to profits the more a firm sells, and we might expect a more productive firm to sell more. As we will see, this is often the case, but there are interesting counter-examples.

Note that, when $\pi(t, c)$ is differentiable in t and c, supermodularity of π implies that the second derivative π_{tc} is positive as t_1 approaches t_2 and c_1 approaches c_2. In the differentiable case, supermodularity is analogous to Hicksian complementarity in consumer theory or strategic complementarity in game theory.

Example 1. A simple case which helps to fix ideas is that of a single-product monopoly firm. Since the foreign countries are symmetric, the firm faces the same inverse demand function $p(x)$ in each, where p and x denotes its price and sales respectively. Its operating profits in each market therefore equal:

$$\pi(t, c) \equiv \underset{x}{Max} \left[\{p(x) - c - t\} x \right] \qquad (3)$$

It follows from the envelope theorem that the first derivative of π with respect to t is negative: $\pi_t = -x(t, c)$. We can also show that the second cross-partial derivative is positive: $\pi_{tc} = -x_c > 0$. To see this, differentiate the first-order condition $p - c - t + xp' = 0$

to get: $(2p' + xp'')\,dx = dc$, where the expression in brackets must be negative from the firm's second-order condition. Hence we have that $x_c = (2p' + xp'')^{-1} < 0$, and so π in (3) is supermodular in t and c.

This example exhibits two key features: π is continuous in trade and production costs, and it depends only on their sum. If both these conditions hold, then supermodularity in t and c is equivalent to *convexity* of π in both t and c: if $\pi(t, c) = \pi(t + c)$ then $\pi_{tc} = \pi_{cc}$. Our next example is a simple case where one of these conditions does not hold and as a result the profit function may not exhibit supermodularity.

Example 2. Consider next the same example as above except that marginal cost varies with output.[7] Assume the firm's problem is as follows:

$$\pi\,(t, c) \equiv \underset{x}{Max}\,[\{p(x) - C\,(c, x) - t\}\, x] \tag{4}$$

Here c is a parameter representing the firm's inverse productivity as before, while average variable cost $C\,(c, x)$ depends positively on c and also varies with output x.[8] (As we will see, the sign of the first derivative C_x, which indicates whether marginal cost rises or falls with output, is irrelevant to whether supermodularity obtains.) As in Example 1, π is supermodular in t and c if and only if x is decreasing in c. Direct calculation yields:

$$x_c = \frac{C_c + xC_{xc}}{2p' + xp'' - (2C_x + xC_{xx})} \tag{5}$$

The denominator is negative from the second-order condition for profit maximization, and C_c in the numerator is positive as already noted, both of which work in favour of supermodularity. However, these terms could be offset, and the profit function could be submodular, if the term C_{xc} is sufficiently negative; that is, if the cost of production falls faster (or rises more slowly) with output for a firm with higher c (i.e., a less productive firm). Figure 1 illustrates this possibility. Firm 1 is less productive than firm 2 overall, but it is *relatively* more productive at higher levels of output. As a result its marginal

[7]We are grateful to Dermot Leahy for suggesting this example.

[8]Total variable cost equals $xC\,(c, x)$, so marginal cost is $C + xC_x$, which varies with output.

cost curve MC_1 lies below that of firm 2 and so it has *lower* marginal cost and (facing the same marginal revenue curve) *higher* output. The profit function in this case is therefore submodular rather than supermodular.

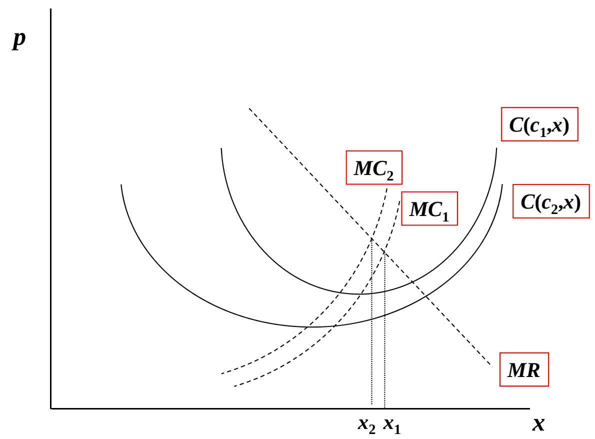


Figure 1: An Example of Submodularity

Admittedly, the configuration shown in Figure 1, though not pathological, is somewhat contrived. In general, supermodularity will hold as long as differences in efficiency between firms work in the same direction on average and at the margin, which seems the natural case. In Section 6 we will consider a more plausible example which can also exhibit submodularity.

3 The Tariff-Jumping Gain

We return to the general case where $\pi(t, c)$ is unrestricted, and compare the relative profitability of different modes of serving the n foreign markets. We first restate in our notation the familiar proximity-concentration trade-off, and then derive a general result on which firms will select into exporting and FDI.

Assume for the moment that the internal trade cost between EU countries is the same as the transatlantic cost, equal to t. Hence the only options the firm need consider are exporting from the U.S. to all EU countries, or setting up an affiliate plant in all. Exporting faces a higher access cost, so FDI has the advantage of proximity. However, it foregoes the benefits of concentration. In addition to operating profits, the firm must incur a fixed cost of serving the market, which differs depending on the mode of access. The fixed cost equals f_X if the firm exports and f_F if it engages in FDI and builds a plant in the market in question. We assume that $f_X < f_F$ and that access costs conditional on FDI are zero.

Total profits as a result of exporting from the U.S. to all n countries are:

$$\Pi^X = n\left[\pi\left(t, c\right) - f_X\right] \tag{6}$$

while total profits from locating a plant in each country are:

$$\Pi^{Fn} = n\left[\pi\left(0, c\right) - f_F\right] \tag{7}$$

We define the *tariff-jumping gain* γ as the difference between these two in a single market:[9]

$$\gamma(t, c, f) \equiv \pi(0, c) - \pi(t, c) - f = \frac{1}{n}\left[\Pi^{Fn} - \Pi^X\right] \tag{8}$$

Here $f \equiv f_F - f_X$ is the excess fixed cost of FDI relative to exporting, which by our earlier assumption is strictly positive. We can now apply the finite difference operator Δ_c to the tariff-jumping gain:[10]

$$\Delta_c\gamma(t, c, f) = \Delta_c\pi(0, c) - \Delta_c\pi(t, c) \tag{9}$$

Recalling the definition of supermodularity, we can sign this unambiguously, which implies the following:

Lemma 1. *If and only if the profit function π is supermodular in t and c, then $\Delta_c\gamma(t, c, f)$ is negative.*

The economic implications of this are immediate: if and only if π is supermodular in t and c, the tariff-jumping gain is lower for higher-cost firms and higher for more productive

[9]Strictly γ is a *trade-cost-jumping gain*, but the shorter title is traditional and simpler.

[10]To avoid confusion, we include f among the arguments of $\Delta_c\gamma(t, c, f)$. However, this finite difference is independent of f, a point which will become crucial below.

ones. Since γ measures the incentive to engage in FDI relative to exporting, this gives our first result:

Proposition 1. *Assume that some firms serve the foreign market by exports and the remainder by FDI, and that all internal and external tariffs equal t. Then, if and only if the profit function π is supermodular in t and c, higher-cost firms will select into exports, while lower-cost firms will select into FDI, for all admissible fixed costs f.*

This proposition extends the result obtained by Helpman, Melitz, and Yeaple (2004) to a more general framework than theirs, as we will see in more detail in Subsection 6.1. The sufficiency part of the proposition follows immediately from Lemma 1. The necessity part is slightly more subtle and reflects the fact that we require the result to hold for all admissible fixed costs. Details are given in the Appendix.

4 Export-Platform FDI

We turn next to consider the incentives to engage in export-platform FDI. Suppose that the n EU countries form a customs union, such that the external tariff on goods imported from country 0 remains at t, but that on goods produced in one union member and exported to another is strictly lower, equal to $t_U < t$. (Fixed costs are unaffected.) If the firm now establishes plants in m member countries and exports from them to the remaining $n - m$ countries, its profits will equal:

$$\Pi^{Fm} = m \left[\pi\left(0, c\right) - f_F\right] + \left(n - m\right) \left[\pi\left(t_U, c\right) - f_X\right] \tag{10}$$

In words, each of the m plants has tariff-free access to a single EU country, while they also sell to the remaining $n - m$ EU countries, facing the intra-union trade cost t^U rather than the transatlantic cost t, which would be incurred if those markets were served by exports from the U.S.[11]

[11]Since marginal production costs are identical and independent of scale in all plants, and since all n EU countries are identical, it makes no difference to the firm where in the EU the $n - m$ plants are located. For extensions to asymmetric countries, though under the assumption that the profit functions are differentiable everywhere, see Neary (2002).

We wish to show that the firm will establish plants in either one or all n countries, but not in m where $1 < m < n$. First, rewrite Π^{Fm} as follows:

$$\Pi^{Fm} = m\left[\pi\left(0, c\right) - f_F\right] + \left(n - m\right)\left[\pi\left(t, c\right) - f_X\right] + \left(n - m\right)\left[\pi\left(t_U, c\right) - \pi\left(t, c\right)\right] \tag{11}$$

Subtracting Π^X from this and using (6) and (8), we can write the difference between the profits from establishing m export-platform plants and the profits from exporting to all in the following way:

$$\Pi^{Fm} - \Pi^X = m\gamma\left(t, c, f\right) + \left(n - m\right)\phi\left(t, t_U, c\right) \tag{12}$$

where ϕ is the *export-platform gain*, which is strictly positive for $t_U < t$:

$$\phi(t, t_U, c) \equiv \pi\left(t_U, c\right) - \pi\left(t, c\right) > 0 \tag{13}$$

Equation (12) is linear in m, the number of plants. Hence the optimal number of plants must be a corner solution: the firm will either establish one in all n EU countries or in none. To see this differently, (12) implies that the gain from establishing m plants relative to establishing m', $m > m'$, is proportional to the *difference* between the tariff-jumping gain and the export-platform gain:

$$\Pi^{Fm} - \Pi^{Fm'} = \left(m - m'\right)\left[\gamma\left(t, c, f\right) - \phi\left(t, t_U, c\right)\right] \tag{14}$$

The expression in square brackets is independent of m and m'. If it is negative then $\Pi^{Fm} < \Pi^{F1}$ for all $m > 1$ and only one plant will be established; if it is positive then $\Pi^{Fn} > \Pi^{Fm'}$ for all $m' < n$ and n plants will be established. Summarizing:

Lemma 2. *With n identical countries, a firm that engages in FDI will establish either a single export-platform plant or n plants, one in each country.*

We can now summarize the options available to the firm. It has a choice between FDI in all markets and export-platform FDI, where the net gain from the former is proportional to the difference between the tariff-jumping gain and the export-platform gain:

$$\Pi^{Fn} - \Pi^{F1} = (n-1)\left[\gamma\left(t, c, f\right) - \phi\left(t, t_U, c\right)\right] \tag{15}$$

It also has a choice between export-platform FDI and exporting, where the net gain from the former equals the *sum* of the tariff-jumping gain in one market and the export-platform gain in $n-1$ markets:

$$\Pi^{F1} - \Pi^{X} = \gamma\left(t, c, f\right) + (n-1)\,\phi\left(t, t_U, c\right) \tag{16}$$

Both (15) and (16) can be either positive or negative, which gives our next result:

Proposition 2. *There are only three profitable modes of serving the n markets: exporting to all, export-platform FDI (with one plant), and multi-market FDI (with n plants).*

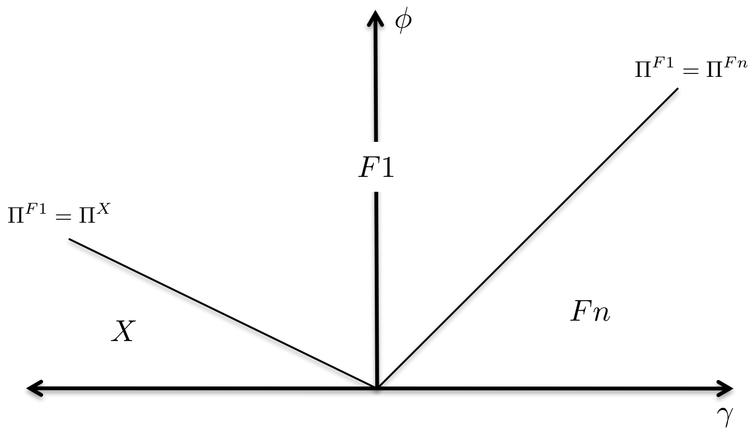


Figure 2: Alternative Modes of Serving Foreign Markets

Figure 2 illustrates the profitable modes of serving foreign markets in the space of $\gamma\left(t, c, f\right)$ and $\phi\left(t, t_U, c\right)$. When ϕ is zero (i.e., along the horizontal axis), exporting and multi-market FDI are the only feasible options. However, as the export-platform gain γ rises, that option becomes more attractive relative to both other modes of serving union markets. Moreover, it becomes increasingly more attractive relative to exporting: the slope of the $\Pi^{F1} = \Pi^{X}$ locus in the left-hand panel (equal to $1/(n-1)$ from (16)) is less in absolute value than the slope of the $\Pi^{F1} = \Pi^{Fn}$ locus in the right-hand panel (equal to one from (15)). Of course, $\gamma(t, c, f)$ and $\phi(t, t_U, c)$ depend on both market access and production costs, so we must turn to see how their variations with firm productivity influence firms' choice of access mode.

5 Firm Selection

Having identified the profitable modes of serving the markets, we now wish to explore how firms of different productivity will select between different modes. First, consider the effects of a difference in costs on the relative profitability of export-platform FDI and exports, as given by equation (16):

$$\Delta_c \left(\Pi^{F1} - \Pi^{X}\right) = \Delta_c \gamma(t, c, f) + (n-1)\,\Delta_c \phi(t, t_U, c) \tag{17}$$

We know from Lemma 1 that the finite difference of the tariff-jumping gain, $\Delta_c \gamma(t, c, f)$, is negative if and only if the profit function π is supermodular in t and c. Moreover, from the definition of the export-platform gain $\phi(t, t_U, c)$ in equation (13), its finite difference, $\Delta_c \phi(t, t_U, c)$, is also negative given supermodularity. Hence we can conclude that, if and only if π is supermodular, then (17) is negative, implying that more efficient firms will engage in export-platform FDI rather than in exports.

Next, consider how the choice between export-platform and multi-market FDI depends on costs. Note first that the net gain from multi-market FDI relative to export-platform FDI given by (15) can alternatively be written as:

$$\Pi^{Fn} - \Pi^{F1} = (n-1)\,\gamma(t_U, c, f) \tag{18}$$

The difference between the tariff-jumping gain from *outside* the union and the export-platform gain equals the tariff-jumping gain from *inside* the union.[12] Applying the finite

[12] All our results can be presented either in terms of $\gamma(t, c, f)$ and $\phi(t, t_U, c)$ or of $\gamma(t, c, f)$ and $\gamma(t_U, c, f)$. We prefer the former, which is more intuitive. While the alternative approach leads to a simpler form for (15), it leads to a less intuitive form for (16): $\Pi^{F1} - \Pi^{X} = n\gamma(t, c, f) - (n-1)\,\gamma(t_U, c, f)$, the negative sign reflecting the fact that export-platform FDI foregoes the gain from tariff-jumping into

difference operator to (18) we can conclude, once again invoking Lemma 1, that it too is negative provided profits are supermodular in t and c:

$$\Delta_c \left(\Pi^{Fn} - \Pi^{F1}\right) = (n-1)\,\Delta_c \gamma\left(t_U, c, f\right) \tag{19}$$

So, faced with a choice between multi-market and export-platform FDI, more efficient firms engage in the latter provided supermodularity holds.

Combining these implications from equations (17) and (19) gives our principal result:

Proposition 3. *If and only if the profit function π is supermodular in t and c, then the least efficient firms that serve the foreign markets will do so via exporting, the next most efficient via export-platform FDI, and the most efficient via multi-market FDI, for all admissible fixed costs f.*

A striking feature of Proposition 3 is that it does not depend directly on fixed costs. While fixed costs affect the *level* of the tariff-jumping gain γ, they vanish when we compare across two firms using the finite difference operator Δ_c. Fixed costs are essential for a proximity-concentration trade-off, and hence they are necessary for the *existence* of selection effects. However, they do not determine their *direction.* So statements like "Only the more productive firms select into the higher fixed-cost activity" are often true, but always misleading: they are true given supermodularity, but otherwise not.[13] What matters for the direction of selection effects is not a trade-off between fixed and variable costs, but whether there is a complementarity between variable costs of production and of trade. Putting this differently, for FDI to be the preferred mode of market access, a firm must be able to afford the additional fixed costs of FDI, but whether it can afford them or not depends on the cross-effect on profits of tariffs and production costs. When supermodularity prevails, a more efficient firm has relatively higher operating profits in the FDI case, but when submodularity holds, the opposite is true. Of course, all this assumes that fixed costs are truly fixed, both for a single firm as output varies, and for comparisons across firms. Matters are different if they depend on either t or c, as we

non-platform countries.

[13]The quoted statement is from Oxford graduate trade lecture notes in late 2009.

shall see in the next section.

6 Applications

We now wish to show that supermodularity of $\pi(t, c)$ holds in a wide range of models, not just in the simple monopoly example of Section 2. Our first application allows for alternative assumptions about market structure, and our second for alternative specifications of technology.

6.1 Monopolistic Competition with Heterogeneous Firms

As already noted, Helpman, Melitz, and Yeaple (2003) were the first to consider how firms of different costs will select into different modes of serving multiple foreign markets. They considered this issue in a model of heterogeneous firms in monopolistic competition with CES or Dixit-Stiglitz preferences and a Pareto distribution of firm productivities. In our notation the variable-profit function for a typical firm in such a model is:

$$\pi(t, c) = (ct)^{1-\sigma} B \tag{20}$$

Here σ is the elasticity of substitution in demand, which must be greater than one, and B is a catch-all term which summarizes the dependence of the demand for one firm's good on income and the prices of all other goods. Consider first what Helpman, Melitz, and Yeaple (2003) call the "partial equilibrium case" where B is taken as given. In that case, it is clear that the profit function is supermodular in t and c:

$$\pi_{tc}(t, c) = (1 - \sigma)^2 (ct)^{-\sigma} B > 0 \tag{21}$$

Hence, from Proposition 2, the ranking of firms by their mode of serving foreign markets established by Helpman, Melitz, and Yeaple (2003) follows immediately without any need to compare the levels of profits in different modes.

In full industry equilibrium, the demand term B is endogenous. It depends directly on the level of total expenditure E and on the overall price index P in the market in question, while P in turn depends on all the variables that affect the global equilibrium, including the number of active firms serving this market from every country i, the distribution of firm costs $g(c)$, and both tariffs t and t^U:

$$B = \tilde{B}(E, P) \qquad P = \tilde{P}\left[\{n_i\}, g(c), t, t^U\right] \tag{22}$$

However, for the comparisons we wish to make, this endogeneity is not relevant. The price index and hence the demand term B would be affected by *changes* in tariffs which disturb the full equilibrium. But our concern is rather with characterizing the pattern of firm selection between different modes of serving foreign markets which face *differences* in tariffs. Since any pair of firms is infinitesimal relative to the mass of all firms, we can compare their choices while holding constant the actions of all other firms. Hence, partial equilibrium is actually the appropriate framework for the cross-section comparisons between different firms in the same equilibrium that we want to make.

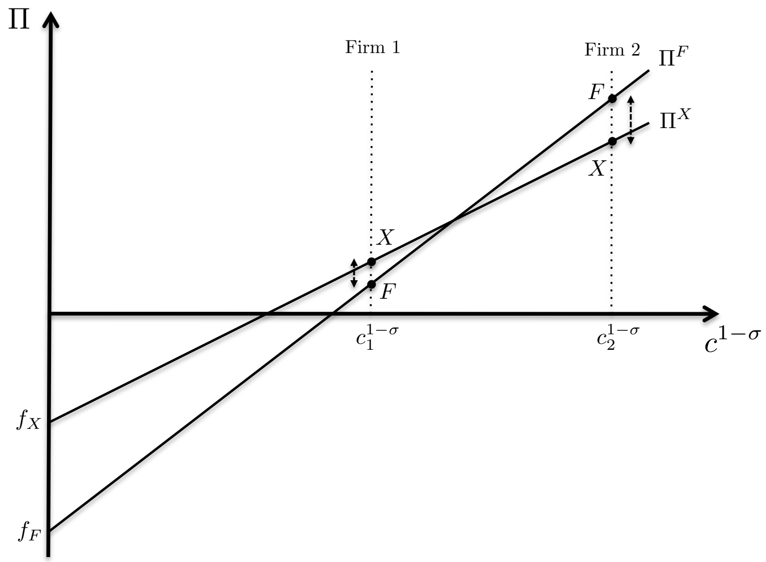


Figure 3: Inferring Selection Effects from Supermodularity

This key point can be made differently by considering the choice between exporting and FDI into a single market (so $n = 1$), using Figure 3, which is based on Helpman, Melitz, and Yeaple (2003). Their approach, which is now standard in the literature, is to compute the general equilibrium of the world economy and then to investigate what pattern of selection effects it exhibits. Thus they calculate not only the profit functions Π^F and Π^X, allowing for their dependence on expenditure and price indices in general equilibrium, but also their point of intersection, which is the threshold cost level at which a firm is indifferent between exports and FDI. By contrast, our approach is very different. We *assume* that an equilibrium exists, and that π is supermodular. We can then pick an arbitrary pair of firms, say those with the unit costs c_1 and c_2 in Figure 3. Rewriting the supermodularity condition $\Delta_c\pi(t,c) > \Delta_c\pi(0,c)$, and adding $-f_F + f_X$ to both sides gives a ranking of the two firms' total profits when they engage in FDI rather than exporting:

$$\begin{aligned} & \pi(t,c_1) - \pi(t,c_2) > \pi(0,c_1) - \pi(0,c_2) \qquad (23) \\ \Leftrightarrow\quad & \pi(0,c_2) - \pi(t,c_2) > \pi(0,c_1) - \pi(t,c_1) \\ \Leftrightarrow\quad & \Pi^F(c_2) - \Pi^X(c_2) > \Pi^F(c_1) - \Pi^X(c_1) \end{aligned}$$

Repeating this comparison for every pair of firms allows us to infer the qualitative properties of the Π^F and Π^X loci without the need to calculate the full equilibrium. While our approach cannot confirm that an equilibrium exists, by dispensing with computing one explicitly, it applies without specific restrictions on the functional forms of preferences, technology, or the distribution of costs, it avoids the need to assume that all countries are symmetric, and it extends easily to considering export-platform FDI, which cannot be illustrated in diagrams like Figure 3.

Supermodularity in models of monopolistic competition does not require that preferences are CES. For example, it can be shown that profits in the Melitz and Ottaviano (2008) model, which assumes quadratic preferences over a continuum of goods, are *sub-modular* in t and c over the relevant range of parameters. Hence, without the need for further analysis, we can state a new result: if preferences are quadratic and the industry is monopolistically competitive, then the pattern of firm selection into exporting, export-platform FDI, and multi-market FDI, is the reverse of that given by Proposition 3. This is particularly convenient because, as Behrens, Mion, and Ottaviano (2010) show, it is extremely difficult to compare two different equilibria analytically in the Melitz-Ottaviano model. The problem arises from the fact that all variables in any given equilibrium can be written as functions of the cost cutoff (the threshold level of marginal cost above which a firm finds it unprofitable to produce). However, comparing two different cutoffs is extremely difficult. Our approach makes it unnecessary to do so: we *assume* that an equilibrium exists in which firms select into different modes of serving the market, and can then invoke Proposition 2 to justify which mode is relatively more profitable for any pair of firms, and, by extension, for all firms.

6.2 Endogenous Fixed Costs

So far we have assumed that fixed costs are exogenous. It is insightful to explore whether our results also hold when they are chosen endogenously. Now, the firm can influence the level of its fixed costs in each market as well as its variable costs, so the trade-offs it faces are more complex.

Consider the case where endogenous fixed costs are market-specific. The earlier derivations go through with relatively little modification. All that is needed is to redefine the maximized profit function as the outcome of the firm's choice of *both* its output and its level of investment. To fix ideas, consider the case of investment in cost-reducing R&D. (Other forms of investment, such as in marketing or product innovation, can be considered with relatively minor modifications.) Let k denote the level of investment which the firm undertakes. For simplicity we concentrate on examples where k is a continuous variable. However, it is clear that our approach does not require this, and that supermodularity also applies to cases of a discrete choice between a finite number of techniques, as in Bustos (2011).

Turning to specifics, we assume that investment in R&D incurs an endogenous fixed cost $F(k)$ but reduces average production costs, now denoted $C(c, k)$. Here c is, just as in earlier sections, a parameter representing the firm's exogenous level of costs (the inverse of its productivity), while k is chosen endogenously. $C(c, k)$ is increasing in c and decreasing in k. The maximum profits which the firm can earn in a market, conditional on t and c, is:

$$\pi(t, c) \equiv Max_{x,k} \left[\{ p(x) - C(c, k) - t \} x - F(k) \right] \tag{24}$$

Formally, this is identical to (3) in Example 1, except that now the firm has two choice variables rather than one. It can be checked that π is supermodular in t and c for many commonly used specifications of the cost functions $F(k)$ and $C(c, k)$, so all our results apply in those cases too. However, there are also economically interesting examples where supermodularity is violated, and so the selection pattern of firms into different modes of serving foreign markets given by Proposition 2 is reversed.

To check whether the profit function (24) exhibits supermodularity in t and c, we proceed as in Example 1. The envelope theorem still applies, so the derivative of maximum profits with respect to the tariff equals minus the level of output: $\pi_t = -x(t, c)$. Hence it follows as before that: $\pi_{tc} = -x_c$. So, to check for supermodularity, we need only establish the sign of the derivative of output with respect to the cost parameter c. The first-order conditions for output x and investment k are:

$$p - C - t + xp' = 0 \tag{25}$$

$$-xC_k - F' = 0 \tag{26}$$

Totally differentiate these and write the results as a matrix equation:

$$\begin{bmatrix} 2p' + xp'' & -C_k \\ -C_k & -(xC_{kk} + F'') \end{bmatrix} \begin{bmatrix} dx \\ dk \end{bmatrix} = \begin{bmatrix} C_c dc + dt \\ xC_{kc} dc \end{bmatrix} \tag{27}$$

From the firm's second-order conditions, the diagonal terms in the left-hand coefficient matrix must be negative, and the determinant of the matrix, which we denote by D, must be positive. Solving for the effect of the cost parameter on output gives:

$$D.x_c = -C_c \left(xC_{kk} + F'' \right) + xC_k C_{kc} \tag{28}$$

The first term on the right-hand side is negative, which works in favour of supermodularity of π. The second could work either way. In particular, the term could be positive, and so supermodularity might not prevail, if C_{kc} is negative, so a lower-productivity firm benefits more from investment, in the sense that its costs fall by more; or, equivalently, if C_{ck} is negative, so investment lowers the cost disadvantage of a lower-productivity firm.

In some of the most widely-used models of R&D, the ambiguity of the sign of (28) can be resolved. We consider three examples in turn.

6.2.1 Linear-Quadratic Costs

Following d'Aspremont and Jacquemin (1988), a widely-used specification of R&D costs, also applied to the study of FDI by Petit and Sanna-Randaccio (2000), assumes that the marginal cost function is linear while the investment cost function is quadratic in k:[14]

$$C(c,k) = c_0 - c_1 k \qquad F(k) = \frac{1}{2}\gamma k^2 \tag{29}$$

Firms may differ in either the c_0 or c_1 parameters, but it is clear that in either case output must be decreasing in c: C_{ck} is zero if firms differ in c_0 and positive if they differ in c_1. Hence, the right-hand side of (28) is positive and supermodularity is assured for this specification of R&D costs.

6.2.2 Exponential Costs of R&D

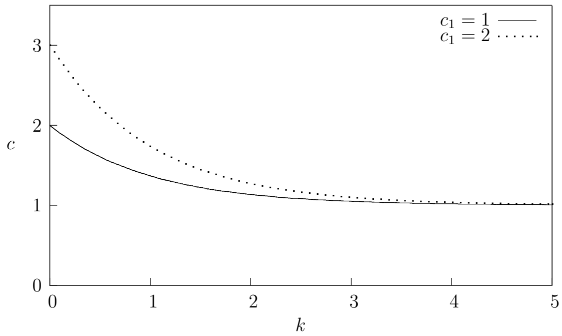
An implausible feature of the d'Aspremont-Jacquemin specification is that the returns to investing in R&D are constant.[15] A more attractive and only slightly less tractable alternative due to Spence (1984) is also widely used:[16]

$$C(c,k) = c_0 + c_1 e^{-\theta k} \qquad F(k) = k \tag{30}$$

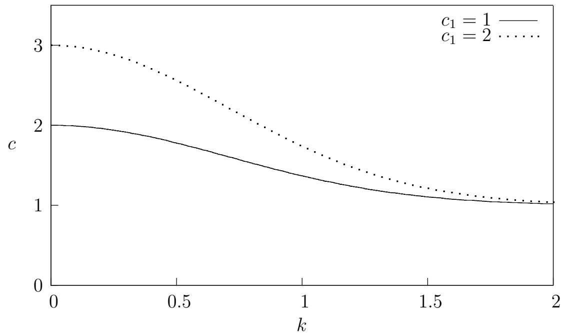
[14]d'Aspremont and Jacquemin and Petit and Sanna-Randaccio also allowed for spillovers between firms.

[15]The linearity of C in k also suggests that the cost of production can become negative, though second-order conditions ensure that this never happens in equilibrium.

[16]These specifications of $C(c,k)$ and $F(k)$ come from Section 5 and from equation (2.3) on page 104 of Spence (1984), respectively.



(a) Spence Model: $a = 1$



(b) Threshold Effects in R&D: $a = 2$

Figure 4: Marginal Cost of Production as a Function of Investment

In this case investment lowers marginal production costs ($C_k = -\theta c_1 e^{-\theta k} < 0$) but at a diminishing rate ($C_{kk} = \theta^2 c_1 e^{-\theta k} > 0$), as illustrated in Figure 4(a) (drawn for $c_0 = \theta = 1$); while the direct cost of investment increases linearly in k ($F'' = 0$). Once again, firms may differ in either the c_0 or c_1 parameters, and supermodularity is assured if they differ in c_0. However, matters are different if firms differ in c_1 (so $c_1 = c$ from now on). In this case, a lower-productivity firm benefits more from investment: $C_{ck} = -\theta e^{-\theta k} < 0$, and this effect is sufficiently strong that it exactly offsets the diminishing returns to investment.[17] As a result, equation (28) is zero under the Spence assumptions: other things equal, two firms with different cost parameters produce the same output.[18] Hence it follows that $\pi(t, c)$ is *modular*, i.e., both supermodular and submodular: the expression in Definition 1 holds with equality.

To see the implications of this for the decision to engage in FDI, consider how two firms of different productivities will assess the relative advantages of exporting and FDI. For any given mode of accessing a market, both firms will produce the same output, the less productive firm compensating for its higher *ex ante* cost by investing more.[19] Hence

[17]Formally, the semi-elasticities of both C_c and C_k with respect to k, C_{ck}/C_c and C_{kk}/C_k, are equal to $-\theta$.

[18]Evaluating (28) gives: $-C_c\left(xC_{kk} + F''\right) + xC_kC_{kc} = -\left(e^{-\theta k}\right)\left(x\theta^2 c e^{-\theta k}\right) + x\left(-\theta c e^{-\theta k}\right)\left(-\theta e^{-\theta k}\right) = -x\theta^2 c e^{-2\theta k} + x\theta^2 c e^{-2\theta k} = 0$.

[19]From (27), the effect of a difference in the cost parameter c on the level of investment is given in general by: $D.k_c = \left(2p' + xp''\right) xC_{kc} + C_cC_k$. The first term on the right-hand side is ambiguous in sign while the second is negative. In the Spence case, the first term is positive and dominates the second, and the expression as a whole simplifies to: $k_c = \theta c$.

their operating profits are also the same. It follows that both firms face exactly the same incentive to export or engage in FDI. We cannot say in general which mode of market access will be adopted, but we can be sure that both firms will always make the same choice. More generally, for any number of firms and for all three possible modes of serving n foreign markets, all firms will adopt the same mode, so no differences in firm selection will be observed.

6.2.3 R&D with Threshold Effects

The fact that the specification due to Spence is just on the threshold between super- and submodularity has implausible implications as we have seen. It also suggests that a less convex investment cost function would imply submodularity. Such a specification is found by adapting that of Spence as follows:

$$C(c,k) = c_0 + ce^{-\theta k^a},\ a > 0 \qquad F(k) = k \tag{31}$$

This is illustrated in Figure 4(b) (drawn for $c_0 = \theta = 1$ and $a = 2$). For values of a greater than the Spence case of $a = 1$, the cost function is initially concave and then becomes convex.[20] This justifies calling this specification one of threshold effects in R&D: low levels of investment have a relatively small effect on production costs whereas higher levels yield a larger payoff. In the FDI context this implies that firms will select into different modes of market access in exactly the opposite way to Proposition 2. Since profits are submodular in t and c, less efficient firms have a greater incentive to establish a branch plant and carry out their R&D investment locally. By contrast, more efficient firms gain relatively little from further investment in R&D, and find it more profitable to concentrate production in their home plant and serve foreign markets by exporting. Provided submodularity of the cost function holds for bilateral comparisons between every pair of active firms, firms of intermediate efficiency engage in export-platform FDI just

[20] From (31), $C_{kk} = -\theta ack^{a-2}e^{-\theta k^a}(a - 1 - \theta ak^a)$. For $0 < a \leq 1$ this is always positive. However, for $a > 1$ it is negative for low k and then becomes positive. The point of inflection occurs where the expression in brackets is zero, which is independent of c and so (for given a and θ) is the same for all firms. In the case illustrated, with $\theta = 1$ and $a = 2$, this occurs at $k = 1/\sqrt{2}$. Note that, while the function is concave at some points and convex at others, it is log-concave everywhere.

as in the supermodular case.

7 Conclusion

This paper characterizes the determinants of firm selection into exporting or foreign direct investment, when firms wish to serve a group of foreign countries. Our key result is that supermodularity of a firm's maximum profits in tariffs and production costs is necessary and sufficient for the configuration of productivities first identified by Helpman, Melitz, and Yeaple (2003): the most efficient firms establish a branch plant in each country; firms of intermediate efficiency establish one plant only and use it as an export platform; while least efficient firms export from their home country. Whereas Helpman, Melitz and Yeaple demonstrated this result in a heterogeneous-firms monopolistic competition model with CES preferences, symmetric countries, and a Pareto distribution of firm productivities, we show that it applies in a much broader range of models and under a variety of assumptions about market structure.

Since the impact effect of both tariffs and production costs is to lower profits, it is not so surprising that there are many cases where their cross effect is positive, so that supermodularity holds. Nevertheless, the restriction of supermodularity is a non-trivial one, and it is possible to find plausible examples where it does not hold. In particular, a further contribution of our paper is a novel specification of investment costs with threshold effects which introduces the possibility of submodularity. In that case, a less efficient firm may have a greater incentive to invest in market-specific R&D than a more efficient one, so as a result it finds engaging in FDI relatively more attractive than exporting.

Our results cast the role of fixed costs as determinants of selection effects in a new light. A fixed cost of FDI is essential for a proximity-concentration trade-off to exist. For FDI to be the dominant model of market access, a firm must be able to afford the additional fixed cost of FDI. However, fixed costs are irrelevant to the determination of which firms are in that category. What matters is the difference-in-difference effect on profits of the marginal costs of production and trade. When supermodularity prevails, a more efficient firm has relatively higher profits in the low-tariff case, but when submodularity holds, the opposite is true. In this paper we have highlighted the implications of this insight for selection into FDI, but it is clear that the general point applies to other cases: selection by more efficient firms into exports as in Melitz (2003), into outsourcing as in Antràs and Helpman (2004), into more skill-intensive techniques as in Bustos (2011), or into more efficient screening of potential workers as in Helpman, Itskhoki, and Redding (2010). There are likely to be still other models which can be illuminated by our approach, and other contexts where the assumption of supermodularity helps to bound comparative statics responses.

Finally, our results should also lend themselves to empirical testing. The predictions of firm selection into either exporting or FDI in one market have received empirical confirmation. (See for example Arnold and Hussinger (2010).) Hopefully the predictions of this paper for selection into export-platform FDI will also prove helpful in empirical work.

References

Antràs, P., and E. Helpman (2004): "Global Sourcing," *Journal of Political Economy*, 112(3), 552–580.

Arnold, J. M., and K. Hussinger (2010): "Exports versus FDI in German Manufacturing: Firm Performance and Participation in International Markets," *Review of International Economics*, 18(4), 595–606.

Athey, S. (2002): "Monotone Comparative Statics Under Uncertainty," *The Quarterly Journal of Economics*, 117(1), 187–223.

Behrens, K., G. Mion, and G. I. Ottaviano (2010): "Economic Integration and Industry Reallocations: Some Theory with Numbers," forthcoming in: Jovanovic, M.N. (ed.) *International Handbook of Economic Integration*, Cheltenham, UK: Edward Elgar.

Bustos, P. (2011): "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms," *American Economic Review* (forthcoming).

Costinot, A. (2009): "An Elementary Theory of Comparative Advantage," *Econometrica*, 77(4), 1165–1192.

Costinot, A., and J. Vogel (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 747–786.

d'Aspremont, C., and A. Jacquemin (1988): "Cooperative and Noncooperative R&D in Duopoly with Spillovers," *American Economic Review*, 78(5), 1133–1137.

Ekholm, K., R. Forslid, and J. R. Markusen (2007): "Export-Platform Foreign Direct Investment," *Journal of the European Economic Association*, 5(4), 776–795.

Grossman, G. M., E. Helpman, and A. Szeidl (2006): "Optimal Integration Strategies for the Multinational Firm," *Journal of International Economics*, 70(1), 216–238.

GROSSMAN, G. M., AND G. MAGGI (2000): "Diversity and Trade," *American Economic Review*, 90(5), 1255–1275.

HABERLER, G. (1936): *The Theory of International Trade with its Applications to Commercial Policy.* William Hodge, London.

HELPMAN, E., O. ITSKHOKI, AND S. REDDING (2010): "Inequality and Unemployment in a Global Economy," *Econometrica*, 78(4), 1239–1283.

HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2003): "Export versus FDI," NBER Working Paper No. 9439.

——— (2004): "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 94(1), 300–316.

LIMÃO, N. (2005): "Trade Policy, Cross-Border Externalities and Lobbies: Do Linked Agreements Enforce More Cooperative Outcomes?," *Journal of International Economics*, 67(1), 175–199.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

MELITZ, M. J., AND G. I. OTTAVIANO (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295–316.

MILGROM, P., AND J. ROBERTS (1990): "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, 80(3), 511–528.

MOTTA, M., AND G. NORMAN (1996): "Does Economic Integration Cause Foreign Direct Investment?," *International Economic Review*, 37(4), 757–783.

NEARY, J. P. (2002): "Foreign Direct Investment and the Single Market," *Manchester School*, 70(3), 291–314.

——— (2009): "Trade Costs and Foreign Direct Investment," *International Review of Economics & Finance*, 18(2), 207–218.

Petit, M.-L. P., and F. Sanna-Randaccio (2000): "Endogenous R&D and Foreign Direct Investment in International Oligopolies," *International Journal of Industrial Organization*, 18(2), 339–367.

Spence, M. (1984): "Cost Reduction, Competition, and Industry Performance," *Econometrica*, 52(1), 101–122.