SELECTION EFFECTS WITH HETEROGENEOUS FIRMS*

Monika Mrázová†
University of Surrey
and CEP, LSE

J. Peter Neary‡
University of Oxford
and CEPR

December 23, 2012

Abstract

We characterize how firms select between alternative ways of serving a market. "First-order" selection effects, whether firms enter or not, are extremely robust. "Second-order" ones, how firms serve a market conditional on entry, are less so: more efficient firms will select into the entry mode with lower market-access costs, if and only if firms' maximum profits are supermodular in production and market-access costs. Supermodularity holds in many cases but not in all. Exceptions include FDI (both horizontal and vertical) when demands are "sub-convex" (i.e., less convex than CES), fixed costs that vary with access mode, and R&D with threshold effects.

Keywords: Foreign Direct Investment (FDI); Heterogeneous Firms; Proximity-Concentration Trade-Off; R&D with Threshold Effects; Super- and Sub-Convexity; Supermodularity.
JEL Classification: F23, F15, F12

---

*An earlier version was circulated as Department of Economics Working Paper No. 588, University of Oxford, December 2011. We are grateful to Arnaud Costinot for extensive discussions, and to Richard Baldwin, Paola Conconi, Jonathan Dingel, Carsten Eckel, Peter Egger, Gene Grossman, Dermot Leahy, Marc Melitz, Giordano Mion, Lindsay Oldenski, Emanuel Ornelas, Gianmarco Ottaviano, Mathieu Parenti, Robert Ritz, Jacques Thisse, Rick van der Ploeg, Adrian Wood, Krešimir Žigić, and participants at various seminars and conferences for helpful comments. Monika Mrázová wishes to thank the ESRC, grant number PTA-026-27-2479. Peter Neary wishes to thank the European Research Council for funding under the EU's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement no. 295669.

†School of Economics, University of Surrey, Guildford, Surrey GU2 7XH, UK; e-mail: m.mrazova@surrey.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

Why do different firms choose to serve particular markets in different ways? Ten years ago, economists had little theory to guide them in thinking about such questions, though a growing body of empirical work had already documented systematic patterns in firm-level data that were unexplained by traditional theory. In the intervening decade, a new and exciting body of theoretical work has emerged which has placed these empirical findings in context and inspired further extensions and elaborations. The starting point of this recent literature is the explicit recognition that firms differ in one or more underlying attribute, typically identified with their productivity; and its central prediction is that more productive firms select into activities with higher fixed costs but lower variable costs. The *locus classicus* for this pattern of behavior is Melitz (2003), who extended the theory of monopolistic competition with differentiated products in general equilibrium to allow for firm heterogeneity, and showed that more efficient firms select into exporting, whereas less efficient ones serve the home market only.[1] Subsequent work in the same vein has shown that more efficient firms select into many different activities, such as producing in-house rather than outsourcing, as in Antràs and Helpman (2004); serving foreign markets via foreign direct investment (FDI) rather than exports, as in Helpman, Melitz, and Yeaple (2004); paying higher wages as in Egger and Kreickemeier (2009) and Helpman, Itskhoki, and Redding (2010); and producing with more skill-intensive techniques as in Bustos (2011). Exploring the implications of firm heterogeneity has already had a profound effect on the study of international trade, and is increasingly being extended to other fields, for example by Ghironi and Melitz (2005) to international macroeconomics, by Davies and Eckel (2010) to international tax competition, and by Forslid, Okubo, and Ulltveit-Moe (2011) to environmental economics.

This literature on heterogeneous firms prompts a number of observations. First, international trade is not the only field in economics where it has been noted that a firm's

[1] A related result of this recent literature, also due to Melitz (2003), is a new source of gains from trade: trade liberalization encourages exit by less productive firms and entry by more productive ones, and so, even when the productivity of each individual firm is unchanged, aggregate productivity rises. However, recent work by Arkolakis, Costinot, and Rodríguez-Clare (2012) suggests that this effect operates in a similar fashion to the gains from trade in traditional models with homogeneous firms.

superiority in one dimension may be associated with enhanced performance in others. The same idea, though expressed in very different ways, can be found in Milgrom and Roberts (1990), who argued that such a complementarity or "supermodularity" between different aspects of firm performance is typical of modern manufacturing. They also advocated using the mathematical tools of robust comparative statics to examine the responses of such firms to exogenous shocks, especially in contexts where variables may change by discrete amounts. This suggests that it may be worth exploring possible links between these two literatures, and possible payoffs to adapting the tools of robust comparative statics to better understand the behavior of heterogeneous firms.

Second, the question immediately arises whether the results derived to date in the literature on heterogeneous firms and trade are robust. One dimension of robustness is that of functional form. All the papers cited above assume that consumers have Dixit-Stiglitz or constant-elasticity-of-substitution (CES) preferences, and all but Melitz (2003) assume that firm productivities follow a Pareto distribution. These assumptions have been relaxed in some papers; for example, Melitz and Ottaviano (2008) show that more efficient firms also select into exports when preferences are quadratic rather than CES. However, with existing techniques each small change in assumptions requires that the model be solved again in full, and, as a result, the question of robustness with respect to functional form has been relatively little explored. A different dimension of robustness is symmetry: existing studies typically assume that countries are identical, both in size and in the distribution of firm productivities. Does this matter for the results? Yet another dimension of robustness is market structure. All the literature on heterogeneous firms to date assumes that the industry is monopolistically competitive, so firms produce a unique product but are infinitesimal in their market. However, if successful firms are indeed large in every dimension, then monopolistic competition may not be the best way of modeling market structure. At least in some markets, it may be more plausible to allow for the emergence of a small number of large firms, competing strategically against each other, and possibly coexisting with a "monopolistically competitive fringe" as in Neary (2010). Clearly it would be desirable to know if the selection effects that have been derived assuming monopolistic competition are also likely to hold in oligopolistic markets. Finally, turning robustness on its head, we can ask whether the fact that more efficient firms engage in more activities is a universal tendency. Should we always expect more productive firms to engage in more and more complex activities? Or are there interesting counter-examples?

In this paper we seek to illuminate these issues both substantively and technically. At a substantive level, we distinguish between two different classes of selection effects, one much more robust than the other. On the one hand, what we call "first-order selection effects" arise when a firm faces a zero-one choice of either engaging or not in some activity, such as production or exporting. On the other hand, "second-order selection effects" arise when a firm faces a choice between different ways of pursuing some goal, such as serving a foreign market via either foreign direct investment or exporting.

We first show that first-order selection effects are extremely robust, requiring only a restriction on the first derivative of the ex post profit function, which as we show holds very widely. This allows us to generalize effortlessly existing results on firm selection into production and exporting, and also into spending on marketing and on worker screening. By contrast, we show that second-order selection effects are considerably less robust. Here, our main substantive contribution is a general result on firm selection which fully characterizes the conditions under which what we call the "conventional sorting" pattern occurs: more efficient firms select into activities with lower marginal costs. We first prove this result in a simple though canonical context: that of a single monopoly firm choosing between serving a foreign market by either exports or horizontal FDI. We then show that, with appropriate qualifications, the result extends to a wide variety of market structures, including both monopolistic competition and oligopoly; and to a wide variety of firm decisions, including vertical FDI, in-house production versus outsourcing, and choice of technique. In all cases, the key consideration is how a firm's own marginal cost of production interacts with the marginal cost of serving the market under different access modes. Our result reveals the unifying structure underlying a wide range of results in the literature, and also shows how they can easily be generalized in new and important ways.

From a technical point of view, our results on second-order selection effects contribute to the small but growing literature which uses the techniques of monotone comparative statics, and in particular the concept of supermodularity, to illuminate issues in international trade.[2] Supermodularity arises very naturally in our context. Our interest is in comparing firms whose production costs differ by a finite amount, and in particular in comparing their behavior under different modes of serving a market, whose marginal costs 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, which allows us to generalize existing results and derive new ones with remarkably few restrictions on technology, tastes, or market structure.

The plan of the paper is as follows. Section 2 shows that first-order selection effects arise naturally in a wide range of models, and are not sensitive to assumptions about functional form. Sections 3 and 4 turn to second-order selection effects, and focus on a single monopoly firm which faces the decision of how to serve a foreign market, trading off the proximity benefits of foreign direct investment against the concentration advantage of producing at home and exporting. Section 3 introduces the setting and explains the restrictions implied by supermodularity. Section 4 formalizes the gains from tariff-jumping and derives our main result on how firms of different productivities will select into one or other mode of serving the foreign market. The remainder of the paper shows that our approach applies in a wide range of contexts, both old (including some of the most widely-used models in international trade), and new. Sections 5 and 6 look at alternative market structures, considering monopolistic competition and oligopoly respectively. Section 7 turns to explore firm choices other than that between exports and FDI. It re-

[2]For more technical details on the application of monotone comparative statics to economics, see Milgrom and Roberts (1990), Milgrom and Shannon (1994), and Athey (2002). Other applications of supermodularity to international trade include Grossman and Maggi (2000), Costinot (2009), and Costinot and Vogel (2010), who use it to study problems of matching between different types of workers or between workers and sectors; Limão (2005), who considers links between trade and environmental agreements; and Costinot (2007), who shows how the assumption of log-supermodularity permits an elegant restatement of the results of Antràs and Helpman (2004) and Helpman, Melitz, and Yeaple (2004), assuming CES preferences and a Pareto distribution of firm productivities.

views a range of other applications, and considers the implications of heterogeneous and endogenous fixed costs of production. The overall message of the paper is that supermodularity holds in many cases but is not inevitable. Among the specific examples we give where supermodularity may be violated, and so the conventional assignment of firms to different modes of accessing foreign markets may be reversed, are FDI (both horizontal and vertical) when demand functions are less convex than the CES, fixed costs that vary with access mode, and R&D with threshold effects.

2 First-Order Selection Effects

Consider a profit-maximizing firm which contemplates serving a particular market. Doing so incurs a fixed cost $f_X > 0$, which, except where otherwise noted, we assume is exogenous and constant across firms. Potentially offsetting this are the firm's operating profits, which depend on various exogenous features of the market, such as market size, access costs, and the behavior of other firms: we assume that the firm takes all these as given.[3] Operating profits also depend on a range of decisions taken by the firm in this and all other markets, including prices and sales levels of each of its products, expenditure on marketing, input choice, whether to outsource or not, etc. We assume the firm takes these decisions optimally, and focus on its maximum or ex post operating profits, which we denote by $\pi(c)$. Here c denotes the one remaining determinant of profits: the firm's own intrinsic exogenous characteristics. In many applications we will follow the literature and identify c with the firm's marginal cost of production, the inverse of firm productivity. However, other interpretations will sometimes prove desirable. We focus on the case of a scalar c, though our results can easily be extended to allow for a vector of firm characteristics.[4]

We make only two assumptions about the maximum profit function:

Assumption 1. *$\pi(c)$ is continuous and strictly decreasing in c.*

[3] We discuss access costs in Section 4 and show how our approach extends to strategic interaction between firms in Section 6.

[4] Heterogeneous firm models with multiple firm characteristics have been considered by Antràs and Helpman (2004), discussed in Section 7.1 below, and Hallak and Sivadasan (2009).

This assumption is natural given our interpretation of c as a cost parameter, though it is non-trivial to establish that it holds in some models, as we will see.

Assumption 2. $\pi(0) > f_X$ *and* $\lim_{c \to +\infty} \pi(c) < f_X$.

By the intermediate value theorem, Assumptions 1 and 2 immediately imply the following result; though mathematically elementary, it has many important economic applications as we will see:[5]

Proposition 1. *There exists a* $c^* \in (0, +\infty)$ *such that* $\pi(c^*) = f_X$. *For any* $c \geq c^*$, $\pi(c) \leq f_X$, *and for any* $c_2 < c_1 \leq c^*$, $\pi(c_2) > \pi(c_1) \geq f_X$.

Thus, there is a threshold cost level c^* such that all firms with lower costs will enter the market and earn strictly positive profits, while those with higher costs will exit. This proposition can be applied either to a single monopoly firm in partial equilibrium or to a continuum of monopolistically competitive firms in industry equilibrium. In the monopoly context it can be interpreted in either a time-series or cross-section context: as a comparative statics statement about the effects of a change in cost, or as a *ceteris paribus* statement about the difference between two firms with different costs, operating in otherwise identical markets. In the monopolistically competitive context it should be interpreted only in the cross-section context: as a comparison between different firms in the same equilibrium.

To see the power of Proposition 1, we first apply it to a canonical example, the heterogeneous firms model of Melitz (2003), extended to a general demand function:

Example 1. [Selection into Production or Exports] Suppose the maximum operating profits of a firm in a particular market equal:

$$\pi(c) \equiv \underset{x}{Max}\ \tilde{\pi}(x; c), \qquad \tilde{\pi}(x; c) = \{p(x) - \tau c\}\, x \tag{1}$$

Here $p(x)$ is the inverse demand function taken as given, i.e., "perceived", by the firm: we impose no restrictions other than $p' < 0$; $\tau \geq 1$ is an iceberg transport cost, representing

[5] The wording of the Proposition is deliberately modeled on that of Proposition 1 in Arkolakis (2010b). See also Example 2 below.

the number of units which must be produced in order to deliver one unit to consumers; and $\tilde{\pi}$ denotes the ex ante operating profit function, maximization of which yields the ex post function π. Differentiating (1) and invoking the envelope theorem, it follows that profits are decreasing in c: $\pi_c = \tilde{\pi}_c = -\tau x < 0$. In the Appendix, Section 9.1, we express this and the relationship between firms' costs and sales in terms of proportional changes (denoted by a "hat" over a variable, e.g., $\hat{x} \equiv d \log x, x > 0$):

$$\hat{x} = -\frac{\varepsilon - 1}{2 - \rho}\hat{c}, \qquad \hat{\pi} = -(\varepsilon - 1)\hat{c} \tag{2}$$

Here $\varepsilon \equiv -\frac{p}{xp'}$ and $\rho \equiv -\frac{xp''}{p'}$ denote the elasticity and the convexity of the demand function respectively.[6] Equation (2) shows that both output and profits are strictly decreasing in c for $c > 0$.[7] This confirms that the Melitz model extended to general demands satisfies Assumption 1, and therefore Proposition 1 applies: more productive firms select into serving a market, whether domestic ($\tau = 1$) or foreign ($\tau > 1$), for all downward-sloping demand functions.

The point of Example 1 is not just that it extends the original Melitz model to arbitrary demand functions.[8] Even more important is what is missing: no assumptions are made about the distribution of costs across firms or about symmetry between countries. All that is needed is π decreasing in c: a very mild assumption. Why is our approach so simple? The standard approach in models of monopolistic competition is to compute the industry equilibrium, and then check that it exhibits selection effects. By contrast, our approach in effect *assumes* that an equilibrium exists, and then shows that π decreasing in c is sufficient for the conventional selection effects to emerge. Our approach parallels that of Maskin and Roberts (2008), who show that all the central theorems of normative general equilibrium theory can be proved using elementary methods once the existence of equilibrium has been established. Our approach cannot confirm that an equilibrium

[6] In the CES case, where σ is the constant elasticity of substitution, $\varepsilon = \sigma$ and $\rho = \frac{\sigma+1}{\sigma}$, so the first equation in (2) simplifies to $\hat{x} = -\sigma\hat{c}$.

[7] The firm's first-order condition requires that $\varepsilon \geq 1$, with $\varepsilon > 1$ for $c > 0$; the second-order condition requires that $\rho < 2$.

[8] Melitz (2003) demonstrated selection effects assuming CES preferences, while Melitz and Ottaviano (2008) showed they also hold under quadratic preferences.

exists.[9] However, 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 countries are symmetric; and it extends easily to considering firm choices in other models, as we shall see in the next two examples.

Finally, while our results apply to cross-section comparisons in a given equilibrium, they also extend to time-series comparisons between equilibria. In particular, we can invoke a result of Bertoletti and Epifani (2012), that selection effects in time series comparisons are robust to changes in functional form, specifically, a reduction of trade costs in an open economy reduces the threshold cost parameter for selection into exporting. The proof of this result relies on Assumption 1, which these authors adopt without discussion. As we have seen, this is natural in the Melitz model, but less immediate in other examples, to which we now turn.

Example 2. [Selection into Marketing] To see how our approach can be applied more widely, consider the case of a firm which must engage in marketing expenditure in order to reach consumers. Following Arkolakis (2010b), it faces two decisions: what price p to charge, or, equivalently, how much to sell per consumer, x; and what proportion n of consumers to target:[10]

$$\pi(c) \equiv \underset{x,n}{Max}\left[\tilde{\pi}(x,n;c)\right], \qquad \tilde{\pi}(x,n;c) = \{p(x) - c\}\, nx - f(n;c) \tag{3}$$

Increased spending on marketing targets a higher proportion of potential consumers n, but incurs a higher fixed cost $f(n;c)$, with $f(0;c) = 0$, $f_n > 0$, and $\lim_{n \to 1} f(n;c) = \infty$. (Note that this fixed cost is endogenous, so we include it in operating profits; as in Example 1, the firm may also incur an exogenous fixed cost.) We make the natural assumptions that the fixed cost of marketing is weakly higher for less productive firms, $f_c \geq 0$, and is

[9] Though this is not a major limitation of our analysis. Equilibrium in monopolistically competitive models of the kind considered in the applied theory literature is unlikely to be a problem. Negishi (1961) proved that equilibrium exists in a very general model of monopolistic competition, assuming that firms have convex production sets and perceive linear demand functions. Arrow and Hahn (1971), Section 6.4, relaxed these assumptions and also allowed for heterogeneous multi-product firms.

[10] To reduce inessential notation, we normalize the size of the market, transport costs, and wages to unity, we assume that consumers are homogeneous, and we assume that n is both the proportion of consumers targeted and the proportion who purchase the good. The latter feature is guaranteed with CES preferences, but may not hold if preferences are such that demand functions exhibit a choke price.

convex in the number of consumers targeted, $f_{nn} > 0$.[11] Arkolakis (2010b) assumes that preferences are CES, and that the marketing cost function takes a particular parametric form.[12] Now consider the effects of costs on profits. Using the envelope theorem as before we obtain:

$$\pi_c = \tilde{\pi}_c = -nx - f_c < 0 \tag{4}$$

It follows that the Arkolakis model, extended to general functional forms as here, exhibits unambiguous first-order selection effects: more efficient firms select into *both* exporting and marketing. This shows that Proposition 1 of Arkolakis (2010b) extends to a wider class of functional forms for demand and marketing costs. Moreover, we show in the Appendix that more productive firms have higher sales and engage in more marketing: (2) continues to hold, irrespective of the shape of the fixed-cost function;[13] while the effect of costs on marketing is also unambiguously negative:

$$\hat{n} = -\frac{\varepsilon - 1 + \frac{cf_{nc}}{f_n}}{\frac{nf_{nn}}{f_n}}\hat{c} \tag{5}$$

irrespective of the convexity of the demand function.[14]

Example 3. [Selection into Worker Screening] Our final example is one where workers have unobservable heterogeneous abilities. Following Helpman, Itskhoki, and Redding (2010), a firm must choose n, which in this example denotes the number of workers it screens for their ability, as well as a, the threshold ability level it will accept. Each of these incurs direct costs: search costs bn in the case of workers sampled, and screening costs $\frac{c_0}{\delta}a^{\delta}$ in the case of the hiring threshold.[15] These two variables in turn

[11]The latter assumption is necessary for an interior solution: the second-order condition for choice of n is $f_{nn} > 0$.

[12]Arkolakis (2010b) assumes that the marketing cost function takes the form: $f(n) = \frac{1-(1-n)^{1-\beta}}{1-\beta}$, $\beta \in (0, \infty)$, $\beta \neq 1$. As β approaches one, this can be shown, using L'Hôpital's Rule, to equal $f(n) = \log(1-n)$, a case explored by Butters (1977) and Grossman and Shapiro (1984). When β equals zero, the model reduces to the standard Melitz case. Arkolakis (2010a) allows for a more general marketing cost function similar to here, though retaining CES preferences.

[13]The expression for $\hat{\pi}$ in equation (2) holds with π interpreted as operating profits before the endogenous fixed costs $f(n; c)$ are paid.

[14]With the Arkolakis specification of the fixed-cost function given in footnote 12, this simplifies to $\hat{n} = -\frac{1-n}{n}\frac{\varepsilon-1}{\beta}\hat{c}$. A different special case is to assume the fixed cost is log-linear in c: $f(n; c) = \tilde{f}(n)c^{\alpha}, \alpha \geq 0$; Arkolakis (2010b) assumes $\alpha = 0$. In this case (5) simplifies to $\hat{n} = -\tilde{f}_n \frac{\varepsilon-1+\alpha}{n\tilde{f}_{nn}}\hat{c}$.

[15]The search cost b depends on the tightness of the labour market and so is endogenous in general

determine the number of workers hired, h, which incur wage costs of $w(\cdot)h$, where the wage is the outcome of a bargaining game to be explained below. Finally, fixed costs may also depend on firm productivity, denoted by the inverse of c as in previous examples.[16] All of these costs have to be subtracted from sales revenue $r(x) = p(x)x$ to yield operating profits, so the ex post profit function is:

$$\begin{aligned} \pi(c) &\equiv \underset{n,a}{Max}\left[\tilde{\pi}(n,a;c): \; x = c^{-1}h^{\gamma}\tfrac{k}{k-1}a, \; h = n\left(\tfrac{a}{\underline{a}}\right)^{-k}\right], \\ \tilde{\pi}(n,a;c) &= p(x)\,x - w(\cdot)h - bn - \tfrac{c_0}{\delta}a^{\delta} - f(c) \end{aligned} \tag{6}$$

Maximization of profits is subject to two kinds of constraints. First there are technological constraints on production and hiring, as indicated in the square brackets in (6): sales x are increasing in the number of workers hired h and the screening threshold a, while hires are increasing in the number of workers sampled but decreasing in a.[17] Second, the firm must bargain with its workers over the wage. As in Helpman, Itskhoki, and Redding (2010), we follow Stole and Zwiebel (1996) and assume that the firm cannot write binding contracts with its workers, but must incentivize them to stay with the firm. It does so by engaging in multilateral bargaining after all non-wage costs have been sunk, offering each worker a wage which just equals the reduction in surplus which the firm would suffer if the worker were to leave.[18] Since the firm's surplus equals its revenue less its wage costs, this implies a differential equation in the wage, the solution to which is the final constraint on the firm's profit maximization.

The effects of costs on profits cannot be established by inspection in this model.

---

equilibrium. However, it is taken as given by firms.

[16]While Helpman, Itskhoki, and Redding (2010) assume that fixed costs are common across firms, Helpman, Itskhoki, Muendler, and Redding (2012) in an empirical extension allow for heterogeneous fixed costs of exporting.

[17]The functional forms of both constraints reflect the assumption that worker abilities follow a Pareto distribution with a minimum ability level $\underline{a}$ and a shape, or inverse dispersion, parameter k. Hence setting a threshold ability a yields a truncated Pareto distribution of hired workers with average ability equal to $\frac{k}{k-1}a$.

[18]This assumes for simplicity that an individual worker's ability is unobservable, and that workers are risk-neutral, face an outside option of zero, and have equal bargaining power with the firm. For further discussion, see Stole and Zwiebel (1996), Acemoglu, Antràs, and Helpman (2007), and Helpman, Itskhoki, and Redding (2010).

However, we show in the Appendix that they take a simple form:

$$\pi_c = \tilde{\pi}_c = -\frac{wh}{c\gamma} - f_c \tag{7}$$

Therefore, as in the previous two examples, Assumption 1 holds and so the model exhibits unambiguous first-order selection effects, irrespective of the form of the demand function. We also derive in the Appendix the responses of other variables to differences in costs across firms:

$$\hat{x} = -\Gamma^{-1}\hat{c}, \;\; \hat{r} = \theta\hat{x}, \;\; \hat{n} = \frac{\gamma\theta - \omega}{\gamma\omega}\hat{x}, \;\; \hat{w} = \frac{k}{\delta}\hat{n}, \;\; \hat{h} = \frac{\delta - k}{\delta}\hat{n}, \;\; \hat{a} = \frac{1}{\delta}\hat{n} \tag{8}$$

where ω is the share of wages in revenue: $\omega \equiv \frac{wh}{r}$; and Γ is the inverse elasticity of sales with respect to costs: $\Gamma \equiv 1 - \frac{\gamma\theta-\omega}{\gamma\omega}\frac{1+\gamma(\delta-k)}{\delta}$.[19] Equation (8) shows that, if and only if Γ is positive, all variables are monotonically decreasing in c: more productive firms screen more workers, and also hire more, despite imposing a higher threshold ability level; as a result they have higher sales, revenue and profits, though at the same time they also pay higher wages. Crucially, all these results hold irrespective of the form of the demand function.

3 Operating Profits and Supermodularity

Having shown that first-order selection effects are extremely robust, we turn in the remainder of the paper to consider second-order selection effects. In this section we consider a firm located in one country which contemplates serving consumers located in a foreign country. The maximum operating profits the firm can earn in the foreign country equal

[19] As in Helpman, Itskhoki, and Redding (2010), the parameters must satisfy a number of constraints for the model to make sense and to accord with stylized facts. From the output and hiring constraints, $1-\gamma k$ is the elasticity of output with respect to the threshold ability level, for a given number of workers screened n; this must be positive if the firm is to have an incentive to screen. From the penultimate equation in (8), $\delta - k$ must be positive if the model is to exhibit an employer-size wage premium. From equations (54) and (57) in the Appendix, $\gamma\theta - \omega$ must be positive since sales are increasing in numbers of workers hired and in the threshold ability level. None of these conditions guarantees that Γ itself must be positive, though the model exhibits bizarre behaviour if it is not. See the Appendix for further discussion.

$\pi(t, c)$, where t is the access cost (tariffs and transport costs) it faces and c is an exogenous cost parameter, as in Section 2. The parameter c equals the firm's marginal production cost in many applications, though not in all: we will see exceptions in Example 5 and Section 7 below. We assume that π is non-increasing (though not necessarily continuous) in both t and c. As in Section 2, 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, while the latter are suppressed for convenience; we give some examples of each below.

We define Δ_c as the finite difference between the values of a function evaluated at two different values of c, c_1 and c_2, with the convention that c_1 is greater than or equal to c_2. Applying this to the operating profit function π gives:

$$\Delta_c \pi(t, c) \equiv \pi(t, c_1) - \pi(t, c_2) \qquad \text{when} \qquad c_1 \geq c_2 \tag{9}$$

So, $\Delta_c \pi(t, c)$ is the profit loss of a higher-cost relative to a lower-cost firm and is always non-positive. Note that, when $\pi(t, c)$ is differentiable in c, $\frac{\Delta_c \pi(t,c)}{c_1 - c_2}$ reduces to the partial derivative π_c as c_1 approaches c_2.[20]

We can now define what we mean by supermodularity in the context of our paper:[21]

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

$$\Delta_c \pi(t_1, c) \geq \Delta_c \pi(t_2, c) \qquad \text{when} \qquad t_1 \geq t_2. \tag{10}$$

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. 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

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

[21]More generally, following Milgrom and Roberts (1990) and Athey (2002), supermodularity can be defined in terms of vector-valued arguments: π is supermodular in a vector-valued argument when $\pi(x \vee y) + \pi(x \wedge y) \geq \pi(x) + \pi(y)$, 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 $y = \{c_1, t_2\}$.

that supermodularity is equivalent to:

$$\pi\left(t_2, c_2\right) - \pi\left(t_1, c_2\right) \geq \pi\left(t_2, c_1\right) - \pi\left(t_1, c_1\right) \geq 0 \qquad \text{when} \qquad t_2 \leq t_1 \ \text{ and } \ c_2 \leq c_1 \quad (11)$$

Thus, when supermodularity holds, a lower tariff is of more benefit to a more productive firm. This might seem like the natural outcome, 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 important counter-examples. When the first inequality in (10) is reversed, we say that the function is submodular.

Example 4. A simple case which helps to fix ideas is that of a single-product monopoly firm with constant marginal cost and specific tariffs. Let $p\left(x\right)$ denote the inverse demand function which the firm faces, where p and x denote its price and sales respectively. Its operating profits therefore equal:

$$\pi\left(t, c\right) \equiv \underset{x}{Max}\left[\left\{p(x) - c - t\right\} x\right] \tag{12}$$

It is easy to check that the profit function is supermodular in t and c in this case.[22] Intuitively, a firm with higher production costs c has lower sales; hence its profits are reduced less by a rise in the tariff.

Example 4 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 $\pi(t, c) = \pi(t + c)$ and π is differentiable, 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 5. Consider next the same example as above except that marginal cost varies

[22]It follows from the envelope theorem that the first derivative of π with respect to t is minus the initial level of sales: $\pi_t = -x\left(t, c\right)$. Hence the second cross-partial derivative of profits is minus the partial derivative of sales with respect to c: $\pi_{tc} = -x_c > 0$. To establish the sign of this term, differentiate the first-order condition $p - c - t + xp' = 0$ to get: $x_c = -H^{-1}$. The expression $H \equiv -\left(2p' + xp''\right)$ must be positive from the firm's second-order condition. Hence we have that $\pi_{tc} = -x_c = H^{-1} > 0$, and so π in (12) is supermodular in t and c.

with output.[23] Assume the firm's problem is as follows:

$$\pi\left(t,c\right)\equiv \underset{x}{Max}\left[\left\{p(x)-t\right\}x-C\left(c,x\right)\right] \tag{13}$$

Here c is not equal to marginal cost, but rather it is a parameter representing the firm's inverse productivity. The new expression $C\left(c,x\right)$ is the firm's total variable cost: it depends positively on c and on output x. Now the second cross-derivative of the profit function equals the following:[24]

$$\pi_{tc}=-x_c=H^{-1}C_{xc} \tag{14}$$

The term H is positive from the second-order condition for profit maximization, which works in favor of supermodularity. However, this term is offset, and the profit function is *submodular* in t and c, if C is submodular in its arguments $\{x,c\}$ so C_{xc} is 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 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.

The configuration shown in Figure 1, though not pathological, is somewhat contrived and of limited empirical relevance. In general, supermodularity will hold as long as static differences in efficiency between firms work in the same direction on average and at the margin, which seems the natural case. In later sections we will consider more plausible examples of submodularity.

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

[24] As in Example 5.3, the envelope theorem implies that $\pi_t=-x(t,c)$, and so π is supermodular in t and c if and only if x is decreasing in c: $\pi_{tc}=-x_c$. Direct calculation yields equation (14) where $H=-\left(2p'+xp''-C_{xx}\right)$.

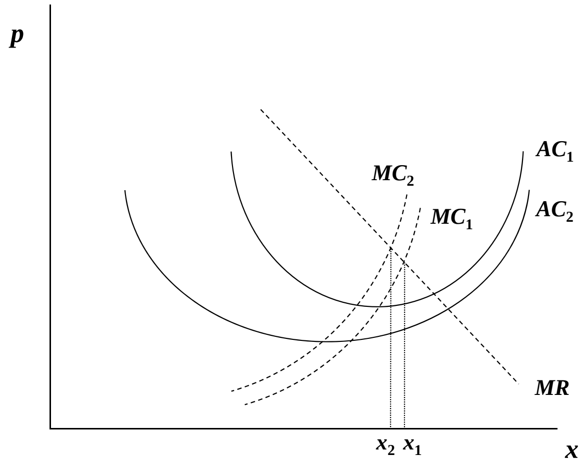


Figure 1: An Example of Submodularity

4 Selection into FDI versus Exporting

We now return to the general case where $\pi(t,c)$ is unrestricted, and compare the relative profitability of different modes of serving the foreign market. 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 or FDI.[25]

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 as a result total profits of exporting are:

$$\Pi^X = \pi(t,c) - f_X \tag{15}$$

When the firm engages in FDI and builds a plant in the market in question, the fixed cost equals f_F, which is not less than f_X. Assuming that access costs conditional on FDI are zero, the total profits from locating a plant in the target market are:

$$\Pi^F = \pi(0,c) - f_F \tag{16}$$

[25] Our formalization of the proximity-concentration trade-off follows Neary (2002).

We define the *tariff-jumping gain* γ as the difference between these two:[26]

$$\gamma(t, c, f) \equiv \Pi^F - \Pi^X = \pi(0, c) - \pi(t, c) - f \tag{17}$$

Here $f \equiv f_F - f_X$ is the excess fixed cost of FDI relative to exporting. For a proximity-concentration trade-off to exist, f has to be such that $\gamma(t, c, f)$ changes signs at least once on the range of parameters considered. On the one hand, f has to be strictly positive since $\pi(0, c) - \pi(t, c) > 0$ for all $t > 0$ and c. In other words, the fixed costs of FDI must be strictly greater than the fixed costs of exporting otherwise all firms would want to engage in FDI. On the other hand, f must not be such that, for all t and c, we have $\gamma(t, c, f) < 0$. In other words, the fixed cost of FDI must not be prohibitive otherwise no firms would want to engage in it. For $t > 0$, set $\bar{f} \equiv \max_c[\pi(0, c) - \pi(t, c)]$. In what follows, we assume that $f \in (0, \bar{f})$.[27]

We can now apply the finite difference operator Δ_c to the tariff-jumping gain:[28]

$$\Delta_c\gamma(t, c, f) = \Delta_c\pi(0, c) - \Delta_c\pi(t, c) \tag{18}$$

Recalling the definition of supermodularity in (10), we can sign this unambiguously, which gives our first result:

Lemma 1. *If and only if the profit function π is supermodular in t and c, $\Delta_c\gamma(t, c, f)$ 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 ones. Since γ measures the incentive to engage in FDI relative to exporting, we can go further and state one of the key results of our paper:

Proposition 2. *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 $f \in (0, \bar{f})$.*

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

[27]Note that $\bar{f}$ may be infinite if for example $\lim_{c\to 0}[\pi(0, c) - \pi(t, c)] = +\infty$.

[28]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 feature which will prove important below.

The sufficiency part of the proposition follows immediately from Lemma 1. The necessity part is more subtle and reflects the fact that we require the result to hold for all admissible fixed costs. A formal proof is in the Appendix.[29] Here we give an intuitive account.

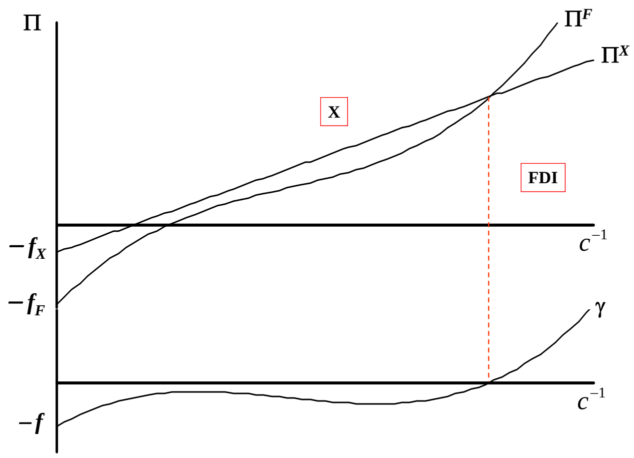


Figure 2: The Conventional Sorting can Hold without Supermodularity

The upper quadrant of Figure 2, based on Helpman, Melitz, and Yeaple (2004), gives a hypothetical illustration of total profits under the two modes of market access, as functions of inverse production costs.[30] The lower quadrant shows $\gamma(t, c, f)$, the difference between Π^F and Π^X. In the example shown, this curve is non-monotonic in c, and so the profit function is not supermodular in t and c. Despite this, the conventional sorting holds: there is a unique cost threshold, with all firms that have lower costs engaging in FDI and all those that have higher costs engaging in exports. This shows that supermodularity is not necessary for the conventional sorting to hold for a given fixed cost: the necessary condition, that the two profit curves cross only once, is weaker. However, Figure 3 shows that supermodularity is necessary if the conventional sorting is to hold for *any* fixed cost.

[29] The proof relies on the fact that Π^X and Π^F, though very general functions of t and c, are quasilinear with respect to f_X and f_F, respectively. It proceeds in a similar way to Proposition 10 of Milgrom and Shannon (1994).

[30] Helpman, Melitz, and Yeaple (2004) assume CES preferences, in which case the total profit curves are linear in a decreasing transformation of costs; see Section 5 below for further discussion. More generally, the total profit curves must satisfy only two restrictions: (i) both must be upward-sloping, reflecting the assumption that operating profits are non-increasing in c; and (ii) Π^F must lie everywhere above Π^X when $f_F = f_X$, reflecting the assumption that operating profits are non-increasing in t.

If γ is not monotonic in c, we can always find an f which leads to a violation of the conventional sorting. In the case shown in Figure 3, the level of f_F, the fixed cost of FDI, is lower than in Figure 2, shifting the $\gamma(t, c, f)$ curve upwards such that the conventional sorting no longer holds: both the lowest- and the highest-cost active firms engage in FDI. Hence we can conclude that supermodularity of the profit function in t and c is necessary as well as sufficient for higher-cost firms to select into exports, and lower-cost firms to select into FDI, for all admissible fixed costs f.

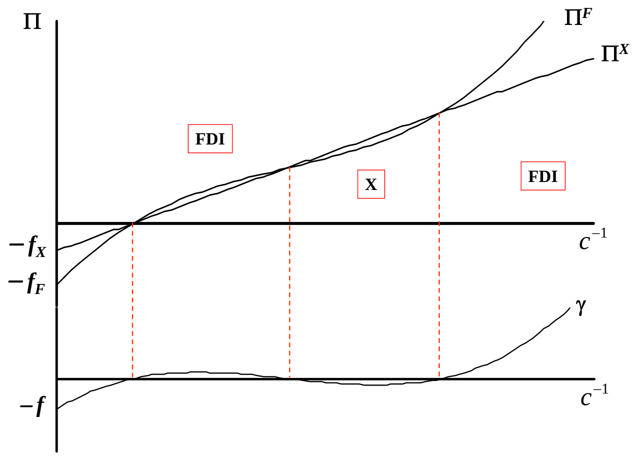


Figure 3: Absent Supermodularity, the Conventional Sorting Fails for Some Fixed Costs

A striking feature of Proposition 2 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 necessarily predict 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.[31] 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

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

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 cross-section comparisons between firms. Matters are different if they depend on either t or c, as we shall see in Section 7. First, we turn to compare Proposition 1 with the result obtained by Helpman, Melitz, and Yeaple (2004). Our result is more general than theirs in that it places no restrictions on the functional form of the demand function: only mild restrictions on the maximized profit function $\pi(t, c)$ are needed. However, at first sight our result seems more special since it holds only for the case of a single monopoly firm, whereas Helpman, Melitz, and Yeaple (2004) proved their result in a general equilibrium model with monopolistic competition. In the next section we show that this apparent limitation of our result is illusory. With suitable reinterpretation, our result holds in a large class of monopolistically competitive models, including that of Helpman, Melitz, and Yeaple (2004).

5 Monopolistic Competition

To see how our result extends to models of monopolistic competition, we need to address the issue of market structure per se, to explore demand systems other than the CES, and to examine the specification of transport costs. We consider these issues in turn.

5.1 Exports versus FDI with CES Preferences

As already noted, Helpman, Melitz, and Yeaple (2004) were the first to consider how firms of different costs will select into different modes of serving foreign markets. They considered this issue in a model of heterogeneous firms in monopolistic competition with CES or Dixit-Stiglitz preferences, iceberg transport costs, and a Pareto distribution of firm productivities. In our notation the variable-profit function for a typical firm in such a model is:

$$\pi(t, c) = (\tau c)^{1-\sigma} B \tag{19}$$

where $\tau \equiv 1 + t \geq 1$ is an iceberg transport cost;[32] 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 total expenditure and the prices of all other goods.[33] Consider first the partial equilibrium or firm-level case where B is taken as given. In that case, the profit function is clearly supermodular in t and c:

$$\pi_{tc} = (\sigma - 1)^2 (\tau c)^{-\sigma} B > 0 \tag{20}$$

Hence, from Proposition 2, the ranking of firms by their mode of serving foreign markets established by Helpman, Melitz, and Yeaple (2004) 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 at a minimum the number of active firms serving this market from every country i, the distribution of firm costs $g(c)$, and the transport cost τ:

$$B = \tilde{B}(E, P) \qquad P = \tilde{P}\left[\{n_i\}, g(c), \tau\right] \tag{21}$$

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 transport costs which disturb the full equilibrium. But our concern is not with such a time-series comparison, rather with characterizing the pattern of firm selection between alternative modes of serving a foreign market which incur *different* transport costs. Since any pair of firms is infinitesimal relative to the mass of all firms, we can compare their choices

[32] For continuity with previous sections we continue to write ex post profits as a function of t. This is not a restriction since $\partial\pi/\partial\tau = \pi_t$.

[33] In typical specifications, $B = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} A$, where A is the constant term in the demand function $x = Ap^{-\sigma}$. A in turn depends on nominal expenditure E and the aggregate price index P: $A = EP^{\sigma-1}$.

while holding constant the actions of all other firms. Hence, partial equilibrium is the appropriate framework for the cross-section comparisons between different firms in the same equilibrium that we want to make.

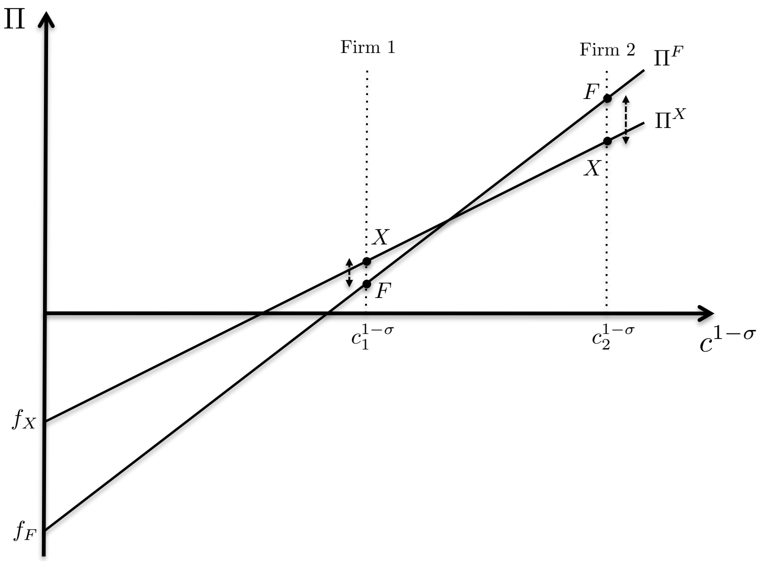


Figure 4: Inferring Selection Effects from Supermodularity

This key point can be made differently by considering Figure 4, which is based on Helpman, Melitz, and Yeaple (2004). Their approach, 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 4. 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:

$$\begin{aligned} & \pi(t, c_1) - \pi(t, c_2) > \pi(0, c_1) - \pi(0, c_2) \\ \Leftrightarrow \quad & \pi(0, c_2) - \pi(t, c_2) > \pi(0, c_1) - \pi(t, c_1) \\ \Leftrightarrow \quad & \Pi^F(c_2) - \Pi^X(c_2) > \Pi^F(c_1) - \Pi^X(c_1) \end{aligned} \tag{22}$$

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, just as we saw for first-order selection effects in Section 2.

5.2 General Demands

While the result in the last sub-section has already been derived by Helpman, Melitz, and Yeaple (2004), the strength of our approach is that it allows us to sign selection effects into FDI for any demand system, not just the CES. Write the demand function facing the firm in inverse form, $p = p(x)$, with no restrictions other than that consumers' willingness to pay is decreasing in price, $p' < 0$; and write the elasticity of demand as a function of sales: $\varepsilon(x) \equiv -\frac{\partial x}{\partial p}\frac{p}{x} = -\frac{p}{xp'}$. To determine which specifications of demand favor the conventional sorting, we introduce the term "superconvex" demand: we define a superconvex demand function as one for which $\log p$ is convex in $\log x$.[34] As we show in the Appendix, this is equivalent to the demand function being more convex than a constant-elasticity CES demand function (for which ε equals σ), and to one whose elasticity of demand is increasing in output, so ε_x is non-negative. The case where demand is not superconvex, so ε is decreasing in x, we call subconvex. Subconvexity is sometimes called "Marshall's Second Law of Demand", as Marshall (1920) argued it was the normal case, a view echoed by Krugman (1979). It implies plausibly that consumers are more responsive to price changes the greater their consumption; and it encompasses many of the most

[34] For a formal definition, and proofs of the statements that follow, see the Appendix, Section 9.3. The term "superconvexity" seems to be used, if at all, as a synonym for log-convexity, i.e., $\log x$ convex in p. (See Kingman (1961).) For related discussions, see Bertoletti and Epifani (2012), Neary (2009), and Zhelobodko, Kokovin, Parenti, and Thisse (2012). For the most part, these papers assume that preferences are additively separable, though this is not necessary for our approach, since we only consider the demand function from the firm's perspective.

widely-used non-CES specifications of preferences, including quadratic (to be considered further below), Stone-Geary, and translog preferences.[35] Strict superconvexity is less widely encountered; an example is where the inverse demand function has a constant elasticity relative to a displaced or "translated" level of consumption: $p = (x-\beta)^{-1/\sigma}$ with β strictly positive.[36] It is shown in Lemma 8 of the Appendix that superconvex demands come "closer" than subconvex demands to violating the firm's second-order condition for profit maximization. Note that super- and subconvexity are local properties, and in particular ε need not be monotonic in x; both ε and ε_x are variable in general, and the latter could be negative for some levels of output and positive for others. However, monotonicity holds for many special cases, including those of quadratic and Stone-Geary preferences.

The importance of superconvexity in this context is shown by the following result:

Proposition 3. *With iceberg transport costs, a sufficient condition for the profit function to be supermodular in t and c for all levels of output is that the demand function is weakly superconvex, i.e., the elasticity of demand is non-decreasing in output, $\varepsilon_x \geq 0$.*

The proof, given in the Appendix, follows by expressing the cross-partial derivative of the profit function in terms of the elasticity of demand and its responsiveness to output:

$$\pi_{tc} = \frac{(\varepsilon-1)^2 + x\varepsilon_x}{\varepsilon - 1 - x\varepsilon_x} x \tag{23}$$

The denominator $\varepsilon - 1 - x\varepsilon_x$ must be positive from the second-order condition. Hence the numerator shows that, whenever ε_x is strictly negative, submodularity may hold for sufficiently high x. We can be sure that supermodularity holds for all output levels only in the CES and strictly superconvex cases.[37]

Intuitively, the result follows from another implication of superconvexity. A positive

[35]Feenstra (2003) shows how the translog can be adapted to allow for a variable number of varieties consumed, and so used in models of monopolistic competition with free entry. He also shows that it implies an elasticity of demand which is always increasing in price, and so, from Lemma 7 in the Appendix, the translog demand function is subconvex. (See Feenstra (2003), p. 85).

[36]We are grateful to Rob Feenstra for suggesting this example.

[37]In the CES case, when ε_x is zero, (23) reduces to $\pi_{tc} = (\sigma-1)x$ and is always positive. This is equivalent to equation (20), using the fact that output with CES preferences equals: $x = \left(\frac{\sigma}{\sigma-1}\tau c\right)^{-\sigma}$.

value of ε_x means that larger firms face a higher elasticity of demand. Since output is decreasing in c in this model ($x_c < 0$), this implies that, if and only if ε_x is positive, more productive firms face more elastic demand. Hence, they also have lower mark-ups, as measured by the Lerner Index, $L \equiv \frac{p-\tau c}{p}$, since $L = \frac{1}{\varepsilon}$.[38] This implies that a more productive firm will have an incentive to expand output more in order to maximize profits. As a result, the Matthew Effect is stronger when ε_x is positive, sufficiently so that supermodularity is guaranteed. By contrast, when ε_x is negative, the Matthew Effect is weaker and so more productive firms may not benefit as much from avoiding the tariff by engaging in FDI.

Proposition 3 is important for highlighting which classes of demand function are consistent with super- or submodularity, but it is only a sufficient condition. To determine whether a particular demand function exhibits supermodularity, we can use the necessary and sufficient condition given by the following:

Proposition 4. *With iceberg transport costs, a necessary and sufficient condition for the profit function to be supermodular in t and c is that the sum of the elasticity and convexity of demand is greater than three.*

The proof (given in the Appendix) proceeds by showing that equation (23) can be reexpressed as follows:

$$\pi_{tc} = \frac{\varepsilon + \rho - 3}{2 - \rho} x, \qquad \rho \equiv -\frac{xp''}{p'} \tag{24}$$

where ρ is our measure of convexity of demand. So, submodularity is more likely when demand is less elastic and more concave. In particular, it may arise for any linear or concave demand system, and even for demands that are "not too" convex.

To illustrate how these results can be applied in practice, we consider two subconvex demand systems, one of which implies that profits are always supermodular and the other which implies submodularity for high values of output. The first is a version of the translated CES case already mentioned:

Lemma 2. *The demand function $p = (x - \beta)^{-1/\sigma}$, with β strictly negative, is always*

[38] It can be checked that the Lerner index falls as costs fall if and only if ε_x is positive: $\frac{dL}{dc} = -\frac{\varepsilon_x}{\varepsilon^2}\frac{dx}{dc}$.

subconvex, but the implied profit function is supermodular at all levels of output provided $\sigma \geq \frac{5}{4}$.

In this case, higher sales are associated with a lower demand elasticity and thus a higher markup, implying that more productive firms do not exhibit such a large difference in output. Nevertheless, the elasticity of demand never falls sufficiently low to allow submodularity to emerge.[39]

Our second example is the case of quadratic preferences, which have been studied in the context of heterogeneous firms by Melitz and Ottaviano (2008). Their model has been extended to the choice between exports and FDI by Nefussi (2006), but only by solving for the full general equilibrium. Using our approach it is easy to establish its properties and to show that its predictions for firm selection into FDI are ambiguous. Writing the inverse demand function as $p = A - bx$, we can express the elasticity of demand as a function of output:[40]

$$\varepsilon(x) = \frac{A - bx}{bx} \tag{25}$$

This is monotonically decreasing in x, so from Proposition 3 we know that for high values of output the profit function may be submodular. However, we need to check that this will happen for values of x that are admissible in equilibrium. To confirm this we specialize Proposition 3 to the quadratic case:

Lemma 3. *With quadratic preferences and iceberg transport costs, the profit function* π *is supermodular when defined on the interval* $c \in \left[\frac{A}{2\tau}, \infty\right]$ *and submodular when defined on the interval* $c \in \left[0, \frac{A}{2\tau}\right]$.

(The proof is in the Appendix.) Unlike the CES case, the profit function is therefore *submodular* for low-cost exporters, although it continues to be *supermodular* for high-cost ones. Hence, provided both exporting and FDI are profitable in the relevant range,

[39] As shown in the Appendix, Section 9.6, $\varepsilon = \frac{x-\beta}{x}\sigma$ and $\rho = \frac{x}{x-\beta}\frac{\sigma+1}{\sigma}$. Though the demand function is subconvex for $\beta < 0$, it is always strictly convex. It follows that $\varepsilon + \rho - 3 = \varepsilon + \frac{\sigma+1}{\varepsilon} - 3$, which can be negative only for very low σ.

[40] As always in monopolistic competition, the demand parameters A and b are taken as given by firms, but are endogenous in general equilibrium. For example, in the Melitz-Ottaviano framework, $A \equiv \frac{\gamma\alpha+\eta N\bar{p}}{\gamma+\eta N}$ and $b \equiv \frac{\gamma}{L}$, where α, γ and η are demand parameters, L is market size, N is the mass of firms, and $\bar{p}$ is the aggregate price index.

we can expect a threefold selection effect in this model: the highest-cost firms select into exporting, but so do the lowest-cost ones, while intermediate-cost firms select into FDI.[41] Figure 5 illustrates this configuration.[42]

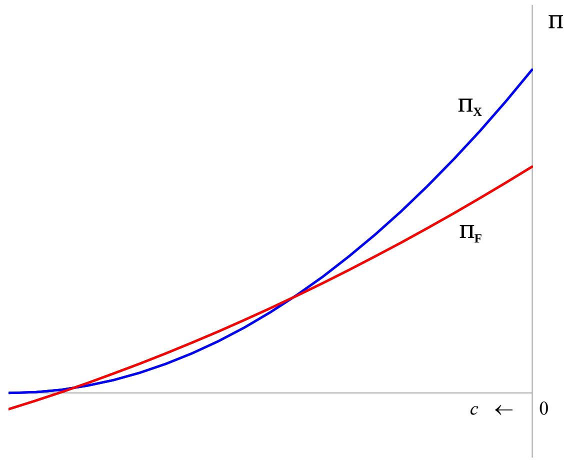


Figure 5: Selection Effects with Quadratic Preferences and Iceberg Transport Costs

As already noted, the corollary given here generalizes the result of Nefussi (2006), dispensing with the assumptions of symmetric countries and a Pareto distribution of firm productivities which he makes. Our result also extends easily to explain the pattern of firm selection into exporting, export-platform FDI, and multi-market FDI when there is more than one foreign market. This is particularly convenient because, as Behrens, Mion, and Ottaviano (2011) show, it does not seem to be possible to compare two different FDI equilibria analytically when preferences are quadratic. 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).[43]

[41] This case holds provided a number of boundary conditions are met: (i) exporting must be profitable, $\Pi^X \equiv \pi(t,c) - f_X > 0$, which requires: $c < \frac{1}{\tau}\left(A - 2\sqrt{bf_X}\right)$; (ii) FDI must be profitable, $\Pi^F \equiv \pi(0,c) - f_F > 0$, which requires: $c < A - 2\sqrt{bf_F}$; and (iii) some selection must take place, i.e., the quadratic equation in c defined by $\Pi^X = \Pi^F$ must have two real roots, which requires: $(\tau-1)A^2 > 4(\tau+1)b\left(f_F - f_X\right)$. Note that we allow for a non-zero fixed cost of exporting, unlike Melitz and Ottaviano (2008). To solve their model in full, they have to assume that exports do not incur any fixed costs, in which case the demand parameter A equals the marginal cost of the threshold firm in equilibrium. Our approach can accommodate fixed costs of exporting, so this property does not necessarily hold here.

[42] To facilitate comparison with earlier figures, c is measured from right to left, starting at zero.

[43] Though this is only possible if there is no fixed cost of exporting, an assumption which our approach

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 our result on supermodularity to justify which mode is relatively more profitable for any pair of firms, and, by extension, for all firms.

5.3 General Transport Costs

The result in the previous sub-section that the largest firms select into exporting for a wide class of demand functions is not necessarily paradoxical. It may simply be viewed as yet another example of large firms' "supermodular superiority."[44] To the extent that the most efficient firms are more productive in all the activities in which they engage, then it is reasonable to assume that they also incur the lowest per unit transport costs. Perhaps they are able to avail of economies of scale in transportation, or to negotiate better terms with transport contractors. From that perspective, the assumption of iceberg transport costs can be seen as a convenient reduced-form way of modeling this superiority of more efficient firms. On the other hand, the suspicion remains that this result is an artifact of iceberg transport costs. It is stretching credulity to assume that the most efficient firms produce the cheapest icebergs, and, in particular, that highly efficient firms, with production costs close to zero, also incur negligible transport costs irrespective of distance. But this is what is implied by the iceberg assumption: to sell x units it is necessary to produce and ship τx units, so the technology of transportation is identical to that of production: $(p - \tau c)x = px - c(\tau x)$.

To see how alternative specifications of transport costs affect the outcome, consider a general specification of the ex post profit function as the outcome of choosing output x to maximize $\tilde{\pi}\,(x; \tau, c)$, the firm's operating profits as a function of the exogenous variables τ and c and an arbitrary level of output:

$$\pi\,(t, c) \equiv \underset{x}{Max}\;\; \tilde{\pi}\,(x; \tau, c) \tag{26}$$

does not require, as discussed in the previous footnote.

[44]We are grateful to Adrian Wood for suggesting this line of reasoning.

We can now express the desired cross-partial derivative of π in terms of second derivatives of $\tilde{\pi}$:[45]

$$\pi_{tc} = \tilde{\pi}_{\tau c} + \tilde{\pi}_{\tau x}\frac{dx}{dc} = \tilde{\pi}_{\tau c} - \tilde{\pi}_{\tau x}\left(\tilde{\pi}_{xx}\right)^{-1}\tilde{\pi}_{xc} \tag{27}$$

This shows that supermodularity of the profit function in t and c depends on the balance between two effects: a direct effect given by $\tilde{\pi}_{\tau c}$, which is the effect of a difference in production costs on the profit disadvantage of higher transport costs at a *given* level of output; and an indirect effect given by the second term on the right-hand side. The expression $\tilde{\pi}_{xx}$ is negative from the firm's second-order condition, so the sign of the indirect effect depends on the product $\tilde{\pi}_{\tau x}\tilde{\pi}_{xc}$. This is presumptively positive; for example it must be so in the case of constant production costs and iceberg transport costs, when $\tilde{\pi}_{\tau x} = -c$ and $\tilde{\pi}_{xc} = -\tau$. This is the Matthew Effect from Section 3: it arises because a higher-cost firm is less vulnerable to a rise in transport costs since it has presumptively lower sales: both $\tilde{\pi}_{\tau x}$ and $\frac{dx}{dc}$ are negative, so their product is positive. By contrast, the direct effect is less robust. In the case of iceberg transport costs it simply equals $\tilde{\pi}_{\tau c} = -x$ and is clearly the source of the potential for submodularity identified in the previous sub-section. It reflects the fact that a higher-cost firm loses more from a rise in transport costs ($\tilde{\pi}_{\tau}$ is more negative) since its cost of shipping one unit of exports is $(\tau - 1)\, c$.

It is immediate that the direct effect vanishes if transport costs and production costs are separable in the profit function $\tilde{\pi}$. This corresponds to the case where exports do not melt in transit, but trade costs are levied instead on the value of sales:

$$\tilde{\pi}\left(x; \tau, c\right) = R\left(x, \tau\right) - cx \tag{28}$$

Here net sales revenue accruing to the firm, R, depends in a very general way on the transport cost parameter. However, there is no direct interaction between transport costs and production costs. As a result, there is no direct effect in the supermodularity expression given by (27): total transport costs and hence $\tilde{\pi}_{\tau}$ do not depend directly on

[45] To derive this we use the envelope theorem to set $\pi_t = \tilde{\pi}_{\tau}$, and totally differentiate the first-order condition $\tilde{\pi}_x = 0$ to obtain $\frac{dx}{dc} = -\left(\tilde{\pi}_{xx}\right)^{-1}\tilde{\pi}_{xc}$.

c, implying that the direct effect $\tilde{\pi}_{tc}$ is zero. By contrast, the indirect effect is positive as before. Hence, profits are supermodular in t and c for all levels of output and all specifications of demand when transport costs and production costs are separable in this way.

Specific transport costs, already considered in Example in Section 3 above, provide one example of (28). Another is where transport costs are ad valorem or proportional to price, so net sales revenue becomes: $R(x, \tau) = \frac{xp(x)}{\tau}$. Relative to the case of iceberg transport costs, the firm's first-order condition is unchanged, but profits are deflated by τ: $\tilde{\pi}(x; \tau, c) = \left[\frac{p(x)}{\tau} - c\right] x$. Similar derivations to those already given shows that equation (27) now becomes: $\pi_{tc} = -c(2p' + xp'')^{-1} > 0$. Thus the full effect is unambiguously positive for *all* demand systems, and so the profit function is always supermodular. Figure 6 illustrates the case of quadratic preferences and proportional transport costs. Clearly, the conventional sorting is now restored, and the model predicts that the most efficient firms will always engage in FDI rather than exporting.

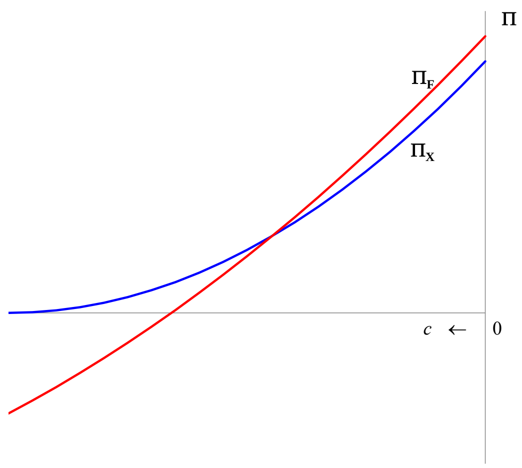


Figure 6: Selection Effects with Quadratic Preferences and Ad Valorem Transport Costs

6 Selection Effects in Oligopoly

The previous section, like almost all the recent literature on trade with heterogeneous firms, assumed that markets are monopolistically competitive. Rare exceptions to this generalization include Porter (2012), who shows that the more efficient firm in a duopoly is more likely to engage in FDI than exporting, and Leahy and Montagna (2009) who show a similar result for outsourcing. It is desirable to establish whether similar results hold more generally when firms are large enough to exert market power over their rivals, so markets are oligopolistic. As already noted, this is of interest both as a check on the robustness of the results and also because, to the extent that more successful firms are likely to engage in a wider range of activities, the assumption that they remain atomistic relative to their smaller competitors becomes harder to sustain.

If individual firms are no longer of measure zero then the arguments used in Section 5.1 no longer hold. If we wish to compare a firm's profits under exporting and FDI, we can no longer assume that the industry equilibrium is unaffected by its choice. However, our earlier result still holds when we take behavior by rival firms as given. To illustrate with a simple example, consider the case where there are two rival U.S. firms, labeled "1" and "2", both of which consider the choice between exporting to the EU and locating a foreign affiliate there. The payoffs to firm 1, conditional on different choices of firm 2, are given in Table 1. Thus, the first entry in the first row, $\pi(t, c, X) - f_X$ gives the operating profits which it will earn if it exports to the foreign market, conditional on the rival firm 2 also exporting. We would expect this to be always less than the second entry, $\pi(t, c, F) - f_X$, which is conditional on firm 2 engaging in FDI: better market access by the rival presumably reduces firm 1's profits, *ceteris paribus*. However, what matters for firm 1's choice is the comparison between different entries in the same column, and it is clear that, conditional on a given mode of market access by firm 2, firm 1's choice will reflect exactly the same considerations as in previous sections. Hence, provided supermodularity holds in each column, and in the columns of the corresponding table for firm 2, our earlier result goes through: when that is the case, more efficient firms will select into FDI and less efficient ones into exporting.

Choice of Firm 2:	Export	FDI
Export:	$\pi(t,c,X)-f_X$	$\pi(t,c,F)-f_X$
FDI:	$\pi(0,c,X)-f_F$	$\pi(0,c,F)-f_F$

Table 1: Payoffs to Firm 1 Given Choices of Firm 2

While the central result derived earlier still holds, it has to be applied with care. One issue is that boundary cases have to be considered in detail. Depending on the configuration of the two firms' costs, in the Nash equilibrium only one of them may serve the market at all, or do so via FDI. There may be no equilibria in pure strategies, in which case mixed strategy equilibria have to be considered. Finally, the necessity part of Proposition 2 does not survive. This is because, even when we allow for all values of fixed costs as in the proof of Proposition 2, supermodularity of the profit function conditional on rivals' responses is necessary for the conventional sorting only at those points which are relevant to a particular Nash equilibrium. Thus it is conceivable that supermodularity might not hold over a range of the profit function; but if that range was never relevant for any value of fixed costs, then the conventional sorting would still apply.

7 Other Applications

So far, our focus has been on the choice between exports and FDI.[46] However, as we show in this section, the same approach applies to a wide range of other firm choices. We first look at the issues of both choice of location and choice of organizational form which arise when firms can vertically disintegrate. We then turn to show how our approach extends to the case where fixed costs differ between locations and between firms. Finally, we consider how selection effects can also be inferred in models where fixed costs are endogenous, determined by prior investments in variables such as technology, research and development (R&D), or marketing. In all cases, results analogous to those derived

[46]In Mrázová and Neary (2010) we show that similar considerations can be used to predict whether firms choosing how to serve a number of foreign markets will select into export-platform or multi-plant FDI.

above apply: supermodularity between the firm's own cost parameter and a parameter representing the marginal cost of the mode of accessing a market is necessary and sufficient for the standard selection effect, whereby more productive firms select into the access mode with lower marginal cost.

7.1 Vertical Disintegration and Choice of Organizational Form

Our discussion of FDI in previous sections concentrated on the horizontal kind, where the firm is considering how to serve a foreign market, and FDI involves effectively reproducing abroad the production facilities which are already located in the home country. This archetypal problem is also one in which differences between the two countries are not central: in particular, we assumed for convenience that the marginal cost of production was the same whether the firm engaged in exporting or in FDI. A different problem arises in the case of a firm whose goal is to serve its home market, but which faces two distinct choices about its organizational form. On the one hand, it has the option of producing either at home or in a lower-cost location abroad. On the other hand, it can choose either to produce in-house or to outsource: the choice of whether or not to vertically integrate arises irrespective of where production is located. The classic treatment of this issue in a model with heterogeneous firms is by Antràs and Helpman (2004), and we draw on their work in what follows. However, since our specification is a reduced-form one, it is also consistent with other ways of modeling the choice of organizational form.[47]

Ignoring fixed costs for the present, the choice of production location and organizational form will depend on the total operating profits that are realized in each case. We write this as:

$$\pi\left(w, \psi, c\right) \equiv \underset{x}{Max}\left(1-\psi\right)\left[p(x)-wc\right]x \tag{29}$$

Unlike in previous sections, we abstract from transport costs. In other respects the model is more complicated. First, the wage which the firm must pay differs between locations. If production takes place in the home country, which we will henceforth refer to as "North", the wage equals w_N, while if it takes place in "South" it equals w_S, with

[47]Kohler and Smolka (2011) present a similar approach to ours.

$w_S < w_N$. Second, the firm owner or "headquarters" must use the services of the supplier of an intermediate input, the quality of which, though observable to both parties, is not contractible.[48] This leads to a profit loss due to incomplete contracting between the firm owner and the intermediate-input supplier, represented by the parameter ψ. Structural microfoundations for this parameter are provided in Antràs and Helpman (2004).[49] Here we need only assume that it may differ both between location and organization form; in particular, we assume that, irrespective of location, it is lower when the firm vertically integrates than when it outsources and must contract with an outside supplier: $\psi_j^V < \psi_j^O$, $j = N, S$.[50]

Considering only variable costs, the firm owner has an incentive to locate production in the South, and to produce in-house rather than outsource. Offsetting these differences in variable costs, the fixed costs of locating in the South are higher then in the North, irrespective of the choice of organizational form: $f_S^i > f_N^i$, $i = O, V$; and the fixed costs of vertical integration or in-house production are higher then those of outsourcing, irrespective of the choice of location: $f_j^V > f_j^O$, $j = N, S$. This configuration of fixed and variable costs ensures that there is a trade-off between different modes of organization, and opens up the possibility that some selection by firms will take place. However, just as in previous sections, it does not predict the direction of selection effects, except in the special case of CES preferences considered by Antràs and Helpman (2004).

To see this, we consider in turn the two choices which the firm must make. Consider first the simplest form of the choice of location, where we assume that there is no efficiency cost of incomplete contracting, so ψ in (29) equals zero. This corresponds to the choice between outsourcing to a Northern contractor and offshoring to a Southern contractor

[48] Either the costs of writing a comprehensive contract are infinite, or the outcomes cannot be observed by a third-party arbitrator. In either case a complete contract cannot be enforced.

[49] They assume that the input supplier's outside option is zero, so it must be paid a fixed amount to persuade it to participate. The headquarters maximizes $(1-\psi)\pi - T - f$ where T is the fixed amount that it must pay the supplier. This is equivalent to maximizing $(1-\psi)\pi - f$.

[50] Our reduced-form specification covers both the case where internalization eliminates all costs of incomplete contracting, as in Williamson (1975) and Grossman and Helpman (2002), and the case where even vertical integration in the North incurs some cost, as in Grossman and Hart (1986), Hart and Moore (1990), and Antràs and Helpman (2004). Our approach can easily be extended to allow the efficiency cost of incomplete contracting to depend on either firm productivity or wages or both: just replace ψ in (29) by $\Psi(\psi, c, w)$, where ψ now represents a structural parameter that determines the cost of incomplete contracting for given values of c and w.

Firm's Decision	Profits	$\varepsilon_x \geq 0$	$\varepsilon_x < 0$
HFDI with iceberg transport costs:	$\tilde{\pi}(x; t, c) = [p(x) - \tau c]\, x,\ \tau \geq 1$	Yes	No
HFDI with separable transport costs:	$\tilde{\pi}(x; \tau, c) = R(x; \tau) - cx,\ \tau \geq 1$	Yes	Yes
Produce in North or South:	$\tilde{\pi}(x; w, c) = [p(x) - wc]\, x$	Yes	No
Produce in-house or outsource:	$\tilde{\pi}(x; \psi, c) = (1 - \psi)\, [p(x) - wc]\, x$	Yes	Yes
Invest in low- or high-tech technology:	$\tilde{\pi}(x; \xi, c) = [p(x) - \xi c]\, x,\ \xi \geq 1$	Yes	No

Table 2: Is the Profit Function Supermodular at all Levels of Output?

in Antràs and Helpman (2004). Comparing the first and third row of Table 2, it is clear that the profit function in this case has exactly the same form as the profit function with iceberg transport costs in the horizontal FDI case, with the wage w playing the role of the iceberg cost parameter τ.[51] Hence the results of previous sections apply immediately. From Proposition 2, if selection takes place, then more efficient firms will offshore and less efficient ones will outsource at home if and only if the profit function is supermodular in w and c; while from Proposition 3 we can only be sure that the profit function is supermodular for all output levels if the elasticity of demand is constant or increasing in sales. Thus, without the need for any further analysis, we can conclude that more efficient firms select into outsourcing at home when preferences are CES, but not necessarily otherwise. For example, the sorting predicted by Antràs and Helpman (2004) need not hold under quadratic or Stone-Geary preferences, even with all of their other assumptions retained: very efficient firms are likely to choose to outsource at home rather than to offshore, since they are less affected by the higher wage in the North. On the other hand, with CES preferences, their results go through irrespective of how firm productivities are distributed or whether countries are symmetric or not, provided only that an equilibrium exists.[52]

Consider next the simplest type of choice of organizational form, where we hold lo-

[51] We defer consideration of the fifth row of the table until Sub-Section 7.3.

[52] Note a further generalization which our approach makes possible. Antràs and Helpman (2004) assume that fixed costs are always incurred in the North, so $\Pi = \pi - w_N f$. Suppose instead that fixed costs are incurred in the country where production is located. In that case we can redefine variable profits as $\pi(w, c) = (p - wc)x - wf$, $w = w_N, w_S$, and we can see that this makes no difference to whether π is supermodular or not in w and c. Hence the assumption made by Antràs and Helpman is not needed for the results. (The only complication is that, for the necessity part of the proof of Proposition 2, we need to assume that there is some component of fixed costs which is independent of w.)

cation constant and hence the wage is given. From the fourth row of Table 2, selection effects now depend on whether the profit function is supermodular in ψ and c. It turns out that this is always the case. The direct effect of a difference in costs between firms is straightforward: the higher-cost firm has lower profits, and as a result is less affected by a higher profit leakage due to incomplete contracts. In fact, the direct effect is simply $\tilde{\pi}_{\psi c} = x$. As for the indirect effect, it equals zero: although the higher-cost firm has lower sales, it is neither more nor less vulnerable to a rise in the profit leakage parameter ψ, since this is analogous to a uniform profits tax on all firms: $\tilde{\pi}_{\psi x} = -\tilde{\pi}_x = 0$. Thus, adapting equation (27) to this case, we have:

$$\pi_{\psi c} = \tilde{\pi}_{\psi c} - \tilde{\pi}_{\psi x} \left(\tilde{\pi}_{xx}\right)^{-1} \tilde{\pi}_{xc} = x > 0 \tag{30}$$

This implies that the selection effect for outsourcing is very robust: more efficient firms select into the lower-ψ organizational form (i.e., producing in-house rather than outsourcing) for *all* preferences and productivity distributions.

Finally, selection effects into vertical FDI are ambiguous in general. Relative to producing as an integrated firm at home, vertical FDI faces conflicting incentives. On the one hand, it incurs a higher cost of incomplete contracting, $\psi_S^V > \psi_N^V$, since in the event of the relationship breaking down, the headquarters cannot expect to retain as large a share of profits when production is in the South. On the other hand, it incurs a lower wage rate, $w_S < w_N$. Thus the profit function is unlikely to be unambiguously supermodular or submodular in c and the cost vector $[\psi \;\; w]$, and the predicted selection effects are likely to be highly sensitive to the specification of the model.

7.2 Heterogeneous Fixed Costs

Up to this point we have followed most of the literature on heterogeneous firms in assuming that fixed costs are the same for all firms and in all foreign locations. This is clearly unrealistic, and we need to examine whether our approach can be extended to the case where fixed costs differ between firms or locations.

The previous analysis is unaffected if fixed costs vary with trade costs t only, so f

becomes $f(t)$. For example, Kleinert and Toubal (2010) allow the fixed costs of a foreign plant to increase with its distance from the parent country, and show that this change in assumptions rationalizes a gravity equation for FDI, while Kleinert and Toubal (2006) show that it also avoids the counter-factual prediction that falling trade costs lower FDI. These are important insights, but the model's predictions about selection effects are unchanged. The reason is simple: although the fixed cost varies with trade costs, the finite difference operator applied to the gain from FDI relative to exporting γ eliminates the fixed cost since $\Delta_c f_F(t) = 0$. While differences in fixed costs between locations clearly affect locational choice, they do so in the same way for all firms.

Matters are more complicated if fixed costs vary with both production costs c and trade costs t. Technically, our approach can still be applied, but some care is needed. We now need to include any firm-specific fixed costs in the definition of operating profits.[53] Thus, let $\tilde{\pi}(t,c)$ denote operating profits, and define total operating profits $\pi(t,c)$ as operating profits net of firm-specific fixed costs:

$$\pi(t,c) \equiv \tilde{\pi}(t,c) - [1 - \mathbb{1}(t)]\, f_F(c), \quad \mathbb{1}(t) \equiv \begin{cases} 1 & t > 0 \\ 0 & t = 0 \end{cases} \tag{31}$$

The indicator function $\mathbb{1}(t)$ equals one when t is positive (the exporting case), and equals zero when t is zero (the FDI case). Thus the fixed cost of a foreign plant must be subtracted to get total operating profits in the case of FDI, and this varies with the firm's productivity. Now, there is an additional reason why supermodularity may not hold, depending on how fixed costs vary with productivity. Applying the finite difference operator to the total operating profits function (31) gives:

$$\Delta_c \pi(t,c) - \Delta_c \pi(0,c) = [\Delta_c \tilde{\pi}(t,c) - \Delta_c \tilde{\pi}(0,c)] + \Delta_c f_F(c) \tag{32}$$

The first term in parentheses on the right-hand side is the same as in previous sections. The second term is new, and shows that supermodularity is more likely to hold if fixed

[53] If the necessity part of Proposition 2 is to hold, we must also assume that there is a component of fixed costs which is common to all firms, as before.

costs are higher for less efficient firms.

Two examples illustrate how this effect can work in different directions. The first is from Behrens, Mion, and Ottaviano (2011), who assume that a firm's fixed costs are proportional to its variable costs, $f_F(c) = cf$, so more efficient firms incur lower fixed costs of establishing a foreign plant. In this case, the final term in (32) becomes $\Delta_c f_F(c) = (c_1 - c_2)f$, which is strictly positive for $c_1 > c_2$. Hence, supermodularity of π and so the conventional sorting pattern are reinforced in this case.

A second example comes from Oldenski (2012), who develops a model of task-based trade in services. Because they use knowledge-intensive tasks disproportionately, higher-productivity firms in service sectors are more vulnerable to contract risk when located abroad. This implies that their fixed costs of FDI are decreasing in c: $f'_F < 0$. As a result, the final term in (32) becomes $\Delta_c f_F(c) = f_F(c_1) - f_F(c_2)$, which is strictly negative for $c_1 > c_2$, so π may be submodular. In this case the conventional sorting may be reversed, as higher-productivity firms may find it more profitable to locate at home. Oldenski presents evidence for this pattern in a number of U.S. service sectors.

7.3 Endogenous Fixed Costs

The previous sub-section considered fixed costs that differ exogenously between firms. By contrast, there are many ways in which a firm can influence the level of its fixed costs as well as its variable costs in each market: R&D, marketing, and changing its product line are just three examples. (For simplicity we focus on the case of R&D in what follows.) It is desirable to explore whether our approach extends to these cases, where firms faces more complex trade-offs. We can distinguish between two kinds of decisions. First, conditional on serving a market, does the firm engage in R&D? Second, conditional on engaging in R&D, how much does it invest? In each case we want to understand how differences in productivity between firms affect their choices.

Consider first the participation decision. Following Bustos (2011), it is natural to model this as a choice between two technologies: "high" has higher fixed cost but lower variable cost than "low". Extending the notation used in previous sections, operating profits can be written as follows:

$$\pi(\xi, c) \equiv \underset{x}{Max}\left[\{p(x) - \xi c\}\, x\right], \quad \xi \geq 1 \tag{33}$$

When the "low" technology is adopted, ξ is strictly greater than one, marginal production cost is $\xi c > c$, and fixed cost equals f_l. By contrast, ξ equals one when the "high" technology is adopted, reducing marginal cost to c but incurring a higher fixed cost of f_h. Writing the variable profit function in this way shows that it is formally identical to the case of exports versus FDI considered in Section 4. In particular, the gain from adopting the "high" technology can be written as follows:

$$\gamma(\xi, c) = \pi(1, c) - \pi(\xi, c) \tag{34}$$

As in Section 4, the fixed costs of each technology are the same for all firms, so they do not affect selection at the margin. All that matters for selection is whether profits are supermodular:[54]

Corollary 1. *If and only if the variable profit function $\pi(\xi, c)$ is supermodular in ξ and c, higher-cost firms will select into the "low" technology, while lower-cost firms will select into the "high" one, for all $f \in (0, \bar{f})$.*

The implications of this are summarized in the final row of Table 2. In particular, very efficient firms that already have a low variable cost of production have less incentive to invest in reducing it, and may not do so if their sales are high and they face a sub-convex demand function.

Consider next the case where the firm has decided to invest in R&D or marketing, assumed to be specific to a particular foreign market, but faces the choice of how much to invest and whether to locate its investment at home or in the target market. The earlier derivations go through with relatively little modification, provided we redefine the maximized profit function as the outcome of the firm's choice of *both* its sales and its level of investment. To fix ideas, consider the case of investment in cost-reducing R&D. (Other

[54]Here $f \equiv f_h - f_l$, and $\bar{f}$ is defined as in Proposition 2.

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. This 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 (though it can no longer be interpreted as the inverse of its productivity), while k is chosen endogenously. $C(c,k)$ is increasing in c and decreasing in k, while fixed costs $F(k)$ are increasing in k. The maximum profits which the firm can earn in a market, conditional on t and c, are:[55]

$$\pi(t,c) \equiv \underset{x,k}{Max}\ \tilde{\pi}(x,k;\tau,c), \quad \tilde{\pi}(x,k;\tau,c) = [p(x) - C(c,k) - t]\,x - F(k) \tag{35}$$

As we will see, 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 (35) 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. We show in the Appendix, Section 9.8, that it equals:

$$\pi_{tc} = -x_c = \underset{+}{D}^{-1}\Big[\underbrace{C_c\left(xC_{kk} + F''\right)}_{+} - x\underset{-}{C_k}C_{kc}\Big] \tag{36}$$

The second-order conditions imply that the determinant D and the first term inside the brackets are positive, as indicated, which work in favor of supermodularity of π.

[55] To highlight the new features which arise from investment in R&D, we focus in the text on the case of ad valorem transport costs only. If instead we assume iceberg transport costs, then the ex ante variable profit function becomes: $\tilde{\pi}(x,k;\tau,c) = \{p(x) - \tau C(c,k)\}\,x - F(k)$. Supermodularity of the ex post profit function now depends on $\pi_{tc} = \tilde{\pi}_{\tau c} + \tilde{\pi}_{\tau\nu}\tilde{\pi}_{\nu\nu}^{-1}\tilde{\pi}_{\nu c}$, where $\nu = [x\ k]'$, so submodularity can arise if *either* the demand function or the investment cost function exhibits "too little" convexity.

The second could work either way. In particular, the term could be negative, 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. Ruling out this case gives a sufficient condition for supermodularity of π:

Proposition 5. *$\pi(t,c)$ is supermodular in $\{t,c\}$ if $C(c,k)$ is supermodular in $\{c,k\}$.*

Proposition 5 applies to one of the most widely-used models of R&D:

Example 6. [Linear-Quadratic Costs] d'Aspremont and Jacquemin (1988) assume that the marginal cost function is linear while the investment cost function is quadratic in k:[56]

$$C(c,k) = c_0 - c_1^{-1}k \qquad F(k) = \frac{1}{2}\gamma k^2 \tag{37}$$

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.[57] Hence, the right-hand side of (36) is positive and supermodularity is assured for this specification of R&D costs.

What if the cost function is not supermodular in $\{c,k\}$? We can get a necessary and sufficient condition for supermodularity of π with the following specification:

$$C(c,k) = c_0 + c\phi(k),\ \phi' < 0, \quad \text{and} \quad F'' = 0 \tag{38}$$

This cost function is always submodular: $C_{ck} = c\phi' < 0$; despite which, we can state the following:

Proposition 6. *Given* (38)*, $\pi(t,c)$ is supermodular in $\{t,c\}$ if and only if $\phi(k)$ is log-convex in k.*

(The proof is in the Appendix, Section 9.9.) Just as, in previous sections, supermodularity of the profit function was less likely the less convex the demand function, so here it is less

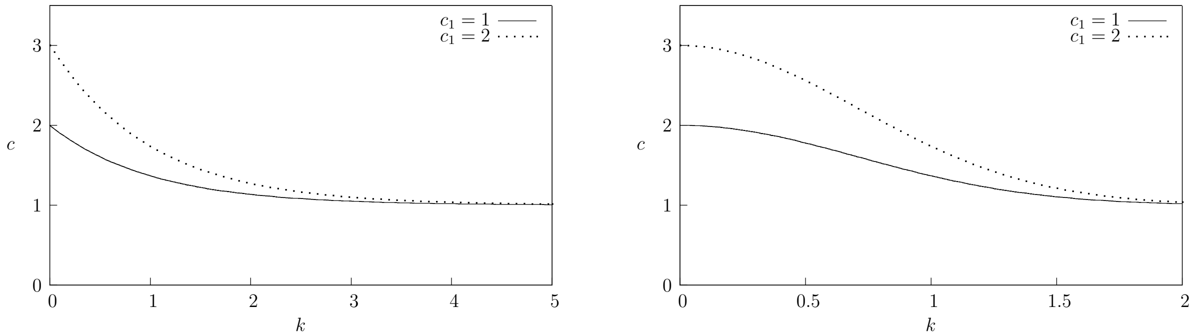
[56] This specification has been applied to the study of FDI by Petit and Sanna-Randaccio (2000). Both they and d'Aspremont and Jacquemin also allowed for spillovers between firms.

[57] When $c = c_1$, $C_{ck} = c^{-2} > 0$.

likely the less convex the marginal cost function. Two examples illustrate the applicability of Proposition 6:

Example 7. [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.[58] A more attractive and only slightly less tractable alternative due to Spence (1984) is also widely used:[59]

$$C(c, k) = c_0 + c_1 e^{-\theta k} \qquad F(k) = k \tag{39}$$



(a) Spence Model: $a = 1$

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

Figure 7: 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 7(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 we set $c_1 = c$ from now on). Now, 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.[60] Expressed in terms of Proposition 6, equation (39) is a special case of (38), with $\phi(k) = e^{-\theta k}$. Hence $\frac{d \log \phi}{dk} = -\theta$, so ϕ is log-linear in k, implying that equation (36)

[58] 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.

[59] 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.

[60] 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 $-\theta$.

is zero and so $\pi(t, c)$ is *modular*, i.e., both supermodular and submodular: the expression in Definition 1 holds with equality. It follows that, other things equal, two firms with different cost parameters produce the same output. The implications for how two firms of different productivities will assess the relative advantages of exporting and FDI are immediate. 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, and so they earn the same operating profits.[61] Hence 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 that differ in c_1, all firms will adopt the same mode of serving the foreign market, so no selection effects will be observed.

Example 8. [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 implies from Proposition 6 that a less convex marginal cost function would yield submodularity. Such a specification is found by generalizing that of Spence as follows:

$$C(c, k) = c_0 + ce^{-\theta k^a},\ a > 0 \qquad F(k) = k \tag{40}$$

In this case $\phi(k) = e^{-\theta k^a}$ and so $\frac{d^2 \log \phi}{dk^2} = -\theta a(a-1)k^{a-2}$, which is negative for $a > 1$, so profits are submodular in $\{c, t\}$. This case is illustrated in Figure 7(b) (drawn for $c_0 = \theta = 1$ and $a = 2$).[62] For values of a greater than the Spence case of $a = 1$, the marginal cost function is initially concave and then becomes convex.[63] This justifies

[61] From (68), the effect of a difference in the cost parameter c on the level of investment is given by: $Dk_c = (2p' + xp'')\, xC_{kc} + C_c C_k$. In general 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: $k_c = \theta c$.

[62] The case of $a = 2$ is the Gaussian distribution.

[63] From (40), $C_{kk} = -\theta ack^{a-2}e^{-\theta k^a}(a - 1 - \theta ak^a)$. 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 $k = 1/\sqrt{2}$. No firm will produce positive output below the inflection point, since C_{kk} must be positive from the second-order conditions. Note that, while the function is concave at some points and convex at others, it is log-concave everywhere.

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 foreign affiliate 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. Hence the conventional sorting is reversed.

8 Conclusion

This paper has provided a novel approach to one of the central questions in recent work on international trade and other applied theory fields: how do different firms select into different modes of serving a market? As well as presenting many new results, we give a unifying perspective on a large and growing literature, identify the critical assumptions which drive existing results, and develop an approach which can easily be applied to new ones.

Our first main contribution is to emphasize an important but hitherto unnoticed distinction between what we call first-order and second-order selection effects. First-order selection effects exhibit a "To be or not to be" feature: firms face a zero-one choice between engaging in some activity (such as serving a market) and not doing so. By contrast, second-order selection effects exhibit a "Scylla versus Charybdis" feature: firms face a choice between two alternative ways of serving a market, each incurring different costs, but both profitable in themselves. The distinction matters because the two types of effects have very different determinants, and the first kind has much more robust predictions than the second.

We show that first-order selection effects between firms with different productivities depend only on the *first* derivative of the ex post profit function with respect to marginal cost. This derivative is presumptively negative in all models (though proving it is positive is non-trivial in some cases), which immediately implies that the standard selection effect holds: the most efficient firms will select into serving the market, the least efficient ones will not. This result applies irrespective of the form of the demand function faced by firms, and requires no assumptions about the distribution of firm productivities. Thus it generalizes substantially a wide range of results: these include the original result of Melitz (2003), that more efficient firms will choose to produce for the home market and to export, less efficient ones will not; as well as the prediction that more efficient firms will engage in marketing, as in Arkolakis (2010b); and that more efficient firms will engage in worker screening, paying higher wages as a result of ex post bargaining with workers hired, as in Helpman, Itskhoki, and Redding (2010).

By contrast, our second main contribution is to show that second-order selection effects are much less robust, and depend on the *second cross*-derivative of the profit function with respect to the firm's marginal cost of production and the marginal cost of the higher-cost activity. If and only if profits are supermodular in these two cost variables, firms exhibit the conventional sorting pattern: more efficient firms select into the lower-variable-cost mode of serving the market, whereas less efficient firms select into the higher-variable-cost mode. We first proved this result for a special case where a single monopoly firm chooses between exporting to a foreign market and engaging in foreign direct investment there. We then showed that our approach generalizes to other market structures, both oligopolistic and monopolistically competitive; and to a wide range of other firm choices, including between outsourcing and producing in-house, and between producing with more or less skill- or R&D-intensive techniques.

The key criterion of supermodularity that we highlight is extremely parsimonious: all that needs to be checked is whether the function giving the maximum profits a firm can earn in a market is supermodular in the firm's own cost parameter, and in a second parameter measuring the marginal cost of accessing the market. Our criterion is simple both in what it includes and in what it omits: no special assumptions are required about the structure of demand, about the distribution of firm productivities, nor about whether countries are symmetric. We are able to dispense with such assumptions because our approach sidesteps the key issue of existence of equilibrium. As Maskin and Roberts (2008) show in a different context, conditional on an equilibrium existing, its properties can often be established relatively easily.

Since the impact effect of both production costs and market access 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 we have shown that there are many plausible examples where it does not hold. In an important subset of cases, where production and market-access costs affect profits multiplicatively, supermodularity and hence the conventional sorting pattern is only assured if the demand function is "superconvex," meaning that it is more convex than a CES demand function with the same elasticity. By contrast, most widely-used demand systems, except the CES, exhibit "subconvexity." Thus, for example, if preferences are quadratic or Stone-Geary, and if selection is observed, it is likely that the most efficient firms will select into exporting rather than FDI. Surprisingly, this multiplicative-costs class includes the canonical case of horizontal FDI where exports incur iceberg transport costs. In this case, the source of the anomalous result can be traced to the assumption of iceberg transport costs: when higher productive efficiency translates into lower transport costs, the most efficient firms suffer a lower transport penalty and so will select into exporting rather than FDI. However, in the case of choosing between outsourcing at home versus offshoring to a lower-wage location, our result continues to hold, even in the absence of transport costs. It implies that for most non-CES preferences more productive firms will select into outsourcing at home, where their greater efficiency offsets the higher wage penalty they incur. We have also identified other plausible cases where supermodularity may fail, such as fixed costs which are higher for more efficient firms, and market-specific investment costs which are subject to threshold effects.

Our results cast the role of fixed costs as determinants of selection effects in a new light. For example, in the choice between FDI and exports, a fixed cost of FDI is essential for a proximity-concentration trade-off to exist: for a firm to face the luxury of choosing between the two modes of market access, it must be sufficiently efficient to afford the additional fixed cost of FDI in the first place. However, conditional on facing the choice, fixed costs do not determine which firms will choose which mode. What matters for this is the difference-in-differences 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 first highlighted the implications of this insight for selection into FDI, and then noted that the general point applies to other cases, including selection by more efficient firms into offshoring as in Antràs and Helpman (2004), or into more skill-intensive techniques as in Bustos (2011). There are likely to be many other models which can be illuminated by our approach, and other contexts where the assumption of supermodularity helps to bound comparative statics responses.

9 Appendix

9.1 Derivations for Section 2

Example 1: The first-order condition sets marginal revenue equal to marginal cost: $p + xp' = \tau c$. Totally differentiating gives: $(2p' + xp'')\, dx = \tau dc$. Reexpressing in terms of proportional changes, using the first-order condition to eliminate c, and using the definitions of ε and ρ, gives $\hat{x}$ in equation (2). A similar elimination of τc yields the expression for $\hat{\pi}$.

Example 2: The first-order condition for sales per consumer is unchanged from Example 1. The first-order condition for the number of consumers equates the net revenue from selling to an additional consumer to the marginal cost of targeting that consumer: $(p - \tau c)\, x = f_n$. Totally differentiating gives: $-\tau x dc = f_{nn} dn + f_{nc} dc$. Collecting terms gives equation (5).

Example 3: We first review the implications of the Stole-Zwiebel bargaining rule with a general demand function. Define revenue as a function of sales $r(x)$; and, via the production function, revenue as a function of hires, the screening threshold, and the firm's cost $R(h, a; c)$:

$$r(x) \equiv xp(x), \qquad R(h, a; c) \equiv r\left[x(h, a; c)\right] \tag{41}$$

Next, define $S(h, a; c)$ as the surplus retained by the firm after wages are paid out of sales revenue:

$$S(h, a; c) \equiv R(h, a; c) - w(h, a; c)h \tag{42}$$

This is less than operating profits, $\pi + f$, because of hiring and screening costs, which are sunk before the bargaining stage. With equal bargaining weights, the wage of the marginal worker must equal the additional surplus to the firm from hiring her: $w(h, a; c) = S_h(h, a; c)$. Substituting from (42) yields a differential equation in $w(h, a; c)$:

$$2w(h, a; c) = R_h(h, a; c) - w_h(h, a; c)h \tag{43}$$

The solution to this is the wage schedule as a function of the number of workers hired:

$$w(h, a; c) = \frac{1}{h^2} \int_0^h R_\xi(\xi, a; c)\xi d\xi \tag{44}$$

In the CES case, R is a power function of h and so (44) can be integrated directly. With general demands, integrate by parts and rearrange to obtain Proposition 3 in Stole and Zwiebel (1996):

$$S\left(h, a; c\right) = \frac{1}{h} \int_0^h R\left(\xi, a; c\right) d\xi \tag{45}$$

Thus the surplus retained by the firm when h workers are hired is an unweighted mean of the revenues generated by all workforces $\xi \in [0, h]$.

With general demands, equation (45) cannot be expressed in closed form. However, all we need are the partial derivatives of the surplus function:

$$
\begin{array}{c}
S_h = -\frac{S}{h} + \frac{1}{h} R(h, a; c) = w \\
S_a = \frac{1}{h} \int_0^h R_a(\xi, a; c) \, d\xi = \frac{1}{h} \int_0^h r' \left[x(\xi, a; c) \right] x_a(\xi, a; c) \, d\xi = \frac{1}{ha} \int_0^h r' \left[x(\xi, a; c) \right] x(\xi, a; c) \, d\xi = \frac{wh}{a\gamma} \\
S_c = \frac{1}{h} \int_0^h R_c(\xi, a; c) \, d\xi = \frac{1}{h} \int_0^h r' \left[x(\xi, a; c) \right] x_c(\xi, a; c) \, d\xi = -\frac{1}{hc} \int_0^h r' \left[x(\xi, a; c) \right] x(\xi, a; c) \, d\xi = -\frac{wh}{c\gamma}
\end{array} \tag{46}
$$

To derive the second and third of these, we use the derivatives of the production function, $x_a = \frac{1}{a}x$, $x_c = -\frac{1}{c}x$, and $x_h = \frac{\gamma}{h}x$, as well as equation (44):

$$
\int_0^h r' \left[x(\xi, a; c) \right] x(\xi, a; c) \, d\xi = \frac{1}{\gamma} \int_0^h r' \left[x(\xi, a; c) \right] x_\xi(\xi, a; c) \, \xi d\xi = \frac{1}{\gamma} \int_0^h R_\xi(\xi, a; c) \xi d\xi = \frac{wh^2}{\gamma} \tag{47}
$$

Summarizing, the total derivative of the surplus function is:

$$
\hat{S} = \frac{\omega}{1-\omega} \left[\hat{h} + \frac{1}{\gamma} (\hat{a} - \hat{c}) \right] \tag{48}
$$

where $\omega \equiv wh/r$ is the share of wages in sales revenue.

We can now restate the firm's problem from (6) as an unconstrained maximization problem:

$$
\underset{n,a}{Max} \left[\tilde{\pi}(n, a; c) \right], \quad \tilde{\pi}(n, a; c) = S\left[h(n, a), a; c \right] - bn - \frac{c_0}{\delta} a^\delta - f(c), \quad h(n, a) = n \left(\frac{a}{\underline{a}} \right)^{-k} \tag{49}
$$

Applying the envelope theorem yields equation (7) in the text, with $\pi_c = \tilde{\pi}_c = S_c - f_c = -\frac{wh}{c\gamma} - f_c$. To derive the other properties of the model, we begin by differentiating (49) to obtain the first-order conditions for the number of workers screened n and the threshold ability level a:

$$
\begin{array}{rcl}
\tilde{\pi}_n(n, a; c) = S_h h_n - b = 0 & \Rightarrow & wh = bn \\
\tilde{\pi}_a(n, a; c) = S_h h_a + S_a - c_0 a^{\delta-1} = 0 & \Rightarrow & \frac{1-\gamma k}{\gamma} wh = c_0 a^\delta
\end{array} \tag{50}
$$

Thus, as in Helpman, Itskhoki, and Redding (2010), wage costs wh equal hiring costs bn and a multiple $\frac{\gamma\delta}{1-\gamma k}$ of screening costs $\frac{c_0 a^\delta}{\delta}$. Totally differentiating the first-order conditions gives:

$$\hat{w}+\hat{h}=\hat{n} \quad \text{and} \quad \delta\hat{a}=\hat{w}+\hat{h} \tag{51}$$

The next equation comes from totally differentiating the bargaining rule $r = wh + S(h, a; c)$:

$$\hat{r}=\omega\left(\hat{w}+\hat{h}\right)+(1-\omega)\,\hat{S}=\omega\left[\hat{w}+2\hat{h}+\frac{1}{\gamma}\left(\hat{a}-\hat{c}\right)\right] \tag{52}$$

making use of (48). The remaining equations comes from totally differentiating the revenue function and the production and hiring constraints:

$$\hat{r}=\theta\hat{x}, \quad \hat{x}=\gamma\hat{h}+\hat{a}-\hat{c}, \quad \text{and} \quad \hat{h}=\hat{n}-k\hat{a} \tag{53}$$

Here the sales-elasticity of revenue, θ, is a simple transformation of the elasticity of demand: $\theta = \frac{\varepsilon-1}{\varepsilon}$.

All that remains is to solve the six equations in (51), (52), and (53) for changes in the six variables, r, x, h, a, w and n, as functions of changes in c. We first combine the two first-order conditions from (51) with the total derivative of the hiring constraint from (53) to solve for changes in h, a and n as functions of $\hat{w}$:

$$\hat{h}=\frac{\delta-k}{k}\hat{w}, \quad \hat{a}=\frac{1}{k}\hat{w}, \quad \text{and} \quad \hat{n}=\frac{\delta}{k}\hat{w} \tag{54}$$

Substituting for $\hat{h}$ and $\hat{a}$ into the total derivative of the production function in (53) gives:

$$\hat{x}=\frac{1+\gamma(\delta-k)}{k}\hat{w}-\hat{c} \tag{55}$$

Substituting from this and the expression for $\hat{h}$ from (54) into the total derivative of the bargaining rule (52) yields:

$$\hat{r}=\omega\left[\hat{w}+\hat{h}+\frac{1}{\gamma}\hat{x}\right]=\omega\left[\frac{\delta}{k}\hat{w}+\frac{1}{\gamma}\hat{x}\right] \tag{56}$$

Equating this to the total derivative of the revenue function from (53) shows that sales and wages are monotonically related:

$$\hat{x} = \frac{\gamma}{\gamma\theta - \omega}\frac{\omega\delta}{k}\hat{w} \tag{57}$$

Using this we can eliminate $\hat{w}$ from (55) to get the expression for $\hat{x}$ in equation (8). The other expressions in that equation follow immediately.

Note how these equations simplify in the case of CES preferences considered by Helpman, Itskhoki, and Redding (2010). Now, the demand elasticity, ε, is a constant, so the sales-elasticity of revenue, θ, is also a constant; the bargaining rule becomes: $wh = \frac{\gamma\theta}{\gamma\theta+1}r$, which when totally differentiated implies $\hat{r} = \hat{w} + \hat{h}$; and the fixed wage share $\omega = \frac{\gamma\theta}{\gamma\theta+1}$ implies that $\frac{\gamma\theta-\omega}{\gamma\omega} = \theta = \frac{1}{\gamma}\frac{\omega}{1-\omega}$. As a result, the inverse elasticity of sales with respect to costs, Γ, simplifies to: $\Gamma_{CES} = 1 - \theta\frac{1+\gamma(\delta-k)}{\delta}$. This parameter also equals the ratio of operating profits to firm surplus: $\left.\frac{\pi+f}{S}\right|_{CES} = \Gamma_{CES}$. The latter property does not extend to the general case: $\frac{\pi+f}{S} = \Gamma + \left(\theta - \frac{1}{\gamma}\frac{\omega}{1-\omega}\right)\frac{1+\gamma(\delta-k)}{\delta\omega}$.

9.2 Proof of Proposition 2

Sufficiency, SM $\Rightarrow$ CS, is trivial. To prove necessity, SM $\Leftarrow$ CS, we proceed by contrapositive and prove $\neg SM \Rightarrow \neg CS$: if π is not supermodular then there exists a fixed cost $f \in (0, \bar{f})$ for which there is unconventional sorting.

Let $t > 0$. If π is not supermodular in t and c, then there exist some c_1 and c_2 such that $c_1 > c_2$ and $\pi(t, c_1) - \pi(t, c_2) < \pi(0, c_1) - \pi(0, c_2)$. Rearranging terms gives:

$$\pi(0, c_1) - \pi(t, c_1) > \pi(0, c_2) - \pi(t, c_2) \tag{58}$$

Let $\alpha_1 \equiv \pi(0, c_1) - \pi(t, c_1)$ and $\alpha_2 \equiv \pi(0, c_2) - \pi(t, c_2)$. Now choose f such that $f = \frac{1}{2}[\alpha_1 + \alpha_2]$. As operating profits are non-increasing in t, $\alpha_1 > \alpha_2 \geq 0$, hence $f > 0$. Also, $\max_c[\pi(0, c) - \pi(t, c)] \geq \alpha_1 > \alpha_2$, hence $f < \bar{f}$. Thus $f \in (0, \bar{f})$.

Now notice that, for this f, we have $\gamma(t, c_1, f) = \pi(0, c_1) - \pi(t, c_1) - f > 0$ and $\gamma(t, c_2, f) = \pi(0, c_2) - \pi(t, c_2) - f < 0$. Since γ measures the incentive to engage in FDI relative to exporting, the higher-cost firm will serve the foreign market via FDI while the lower-cost firm will serve it by exports. Thus, if π is not supermodular in t and c we can always find a fixed cost in $(0, \bar{f})$ such that the conventional sorting is reversed. It follows that supermodularity is necessary for the conventional sorting. $\square$

9.3 Superconvexity

Our formal definition of superconvexity is as follows:

Definition 2. $p(x)$ is superconvex if and only if $\log p$ is convex in $\log x$.

This can be compared with log-convexity:

Definition 3. The inverse demand function $p(x)$ is log-convex if and only if $\log p$ is convex in x. Analogously, the direct demand function $x(p)$ is log-convex if and only if $\log x$ is convex in p.

Some implications of superconvexity are easily established:

Lemma 4. *Superconvexity of the inverse demand function is equivalent to superconvexity of the direct demand function, and implies log-convexity of the inverse demand function, which implies log-convexity of the direct demand function, which implies convexity of both demand functions; but the converses do not hold.*

Proof. Direct calculation yields the entries in Table 3, expressed in terms of $\varepsilon \equiv -\frac{p}{xp'}$ and $\rho \equiv -\frac{xp''}{p'}$.

	Direct Demand	Inverse Demand
Convexity	$\frac{d^2x}{dp^2} = \frac{x}{p^2}\varepsilon\rho \geq 0$	$\frac{d^2p}{dx^2} = \frac{p}{x^2}\frac{\rho}{\varepsilon} \geq 0$
Log-convexity	$\frac{d^2 \log x}{dp^2} = \frac{\varepsilon^2}{p^2}(\rho - 1) \geq 0$	$\frac{d^2 \log p}{dx^2} = \frac{1}{x^2\varepsilon}\left(\rho - \frac{1}{\varepsilon}\right) \geq 0$
Superconvexity	$\frac{d^2 \log x}{d(\log p)^2} = \varepsilon^2\left(\rho - \frac{\varepsilon+1}{\varepsilon}\right) \geq 0$	$\frac{d^2 \log p}{d(\log x)^2} = \frac{1}{\varepsilon}\left(\rho - \frac{\varepsilon+1}{\varepsilon}\right) \geq 0$

Table 3: Criteria for Convexity of Direct and Inverse Demands

The Lemma follows by inspection. Note that the log-convexity ranking of the direct and inverse demand functions requires that $\varepsilon > 1$, whereas the others require only that $\varepsilon > 0$. $\square$

Lemma 5. *A demand function is superconvex if and only if it is more convex than a CES demand function with the same elasticity.*

Proof. Differentiating the CES inverse demand function $p = \alpha x^{-1/\sigma}$ gives: $p' = -\alpha \frac{1}{\sigma} x^{-(1+\sigma)/\sigma}$; and $p'' = \alpha \frac{\sigma+1}{\sigma^2} x^{-(1+2\sigma)/\sigma}$. Hence we have $\varepsilon^{CES} = \sigma$ and $\rho^{CES} = \frac{\sigma+1}{\sigma}$. From the final row of Table 3, it follows that an arbitrary demand function which has the same elasticity as a CES demand function at their point of intersection is superconvex at that point if and only if its convexity exhibits $\rho > \frac{\varepsilon+1}{\varepsilon} = \frac{\sigma+1}{\sigma} = \rho^{CES}$, which proves the result. □

Lemma 6. *A demand function is superconvex if and only if its elasticity is increasing in sales.*

Proof. Differentiating the expression for the elasticity of demand, $\varepsilon(x) = -\frac{p(x)}{xp'(x)}$, yields:

$$\varepsilon_x = -\frac{1}{x} + \frac{p\left(p' + xp''\right)}{\left(xp'\right)^2} = -\frac{1}{x}\left(1 + \varepsilon - \varepsilon\rho\right) \tag{59}$$

Comparison with the final row of Table 3 gives the required result. □

Super-convexity can also be expressed in terms of the direct demand function $x = x(p)$, with elasticity $e(p) \equiv -\frac{px'}{x} = \varepsilon\left[x\left(p\right)\right]$:

Lemma 7. *A demand function is superconvex if and only if its elasticity is decreasing in price.*

Proof. Differentiating the identity equating the two expressions for the elasticity of demand, $e(p) = \varepsilon\left[x\left(p\right)\right]$, yields $e_p = \varepsilon_x p'$. Hence the result follows from Lemma 6. □

Our final lemma relates superconvexity to the second-order condition:

Lemma 8. *Provided marginal cost is strictly positive, a demand function is superconvex if and only if the elasticity of marginal revenue is less than one in absolute value.*

Proof. Define revenue r as $r(x) \equiv xp(x)$. Clearly, $r_x = xp' + p = xp'(1 - \varepsilon) = \tau c$, which is strictly positive by assumption; and $r_{xx} = 2p' + xp''$, which must be negative from the second-order condition. Hence the elasticity of marginal revenue, or the convexity of revenue, equals:

$$\rho^r \equiv -\frac{xr_{xx}}{r_x} = \frac{2-\rho}{\varepsilon-1} = 1 - \frac{\varepsilon+\rho-3}{\varepsilon-1} \tag{60}$$

Recalling Proposition 4, it follows that, when $c > 0$, so $\varepsilon > 1$, superconvexity of the demand function is equivalent to $\rho^r < 1$. □

When marginal cost is strictly positive, the second-order condition requires that the profit function be strictly concave: $2p' + xp'' < 0 \Rightarrow \rho < 2 \Rightarrow \rho^r > 0$. Hence Lemma 8 formalizes the notion that superconvex demands come "closer" than subconvex demands to violating the second-order condition.

9.4 Proof of Proposition 3

Differentiating the profit function $\pi(t,c) = \underset{x}{Max}\,[p(x) - \tau c]\,x$ gives: $\pi_t = -cx$; and

$$\pi_{tc} = -x - c\frac{dx}{dc} = -x - \frac{\tau c}{2p' + xp''} \tag{61}$$

We want to express the right-hand side in terms of ε and ε_x. First, solve (59) for $p' + xp''$ in terms of ε_x, and add p' to it. Next, use the definition of ε to eliminate p', $p' = -\frac{p}{x\varepsilon}$, which gives the second-order condition in terms of ε and ε_x:

$$2p' + xp'' = -\frac{p}{x\varepsilon^2}\left(\varepsilon - 1 - x\varepsilon_x\right) \tag{62}$$

This confirms that the second-order condition $2p' + xp'' < 0$ is equivalent to $\varepsilon - 1 - x\varepsilon_x > 0$. The last preliminary step is to use the first-order condition $p - \tau c + xp' = 0$ to express τc in terms of p and ε: $\tau c = p + xp' = p - \frac{p}{\varepsilon} = \frac{\varepsilon-1}{\varepsilon}p$. (This is very familiar in the CES case.) Finally, substitute these results into (61):

$$\pi_{tc} = -x + \frac{\varepsilon - 1}{\varepsilon - 1 - x\varepsilon_x}\varepsilon x \tag{63}$$

Collecting terms gives the desired expression in (23). □

9.5 Proof of Proposition 4

The proof follows immediately by substituting for ε_x from (59) into the expression for π_{tc} in (23), and noting that, as in (23), the denominator must be positive from the second-order condition. $\square$

9.6 Proof of Lemma 2

For the demand function $p = (x-\beta)^{-1/\sigma}$, we have $p' = -\frac{1}{\sigma}(x-\beta)^{-\frac{\sigma+1}{\sigma}}$ and $p'' = \frac{\sigma+1}{\sigma^2}(x-\beta)^{-\frac{2\sigma+1}{\sigma}}$. Hence $\varepsilon = \frac{x-\beta}{x}\sigma$ and $\rho = \frac{x}{x-\beta}\frac{\sigma+1}{\sigma}$. It follows immediately that $\varepsilon_x = \frac{\beta}{x^2}\sigma$, and so the demand function is strictly subconvex ($\varepsilon_x < 0$) if and only if β is negative. To establish for which values of σ the profit function is supermodular, rewrite the elasticity as $\varepsilon = \frac{\sigma+1}{\rho}$. Hence we seek to minimize $\sigma = \varepsilon\rho - 1$ by choice of ε and ρ, subject to the supermodularity constraint from Proposition 4, $\varepsilon + \rho \geq 3$. Solving gives the boundary values $\varepsilon^* = \rho^* = 1.5$, which imply that the threshold value of σ is $\sigma^* = \left(\frac{3}{2}\right)^2 - 1 = 1.25$. $\square$

9.7 Proof of Lemma 3

Maximizing operating profits, $\pi = (p - \tau c)x$, yields the first-order condition, $A - 2bx = \tau c$, which can be solved for optimal output: $x = \frac{1}{2b}(A - \tau c)$. Substituting back into the expression for profits gives the maximized operating profit function:

$$\pi(t, c) = bx^2 = \frac{1}{4b}(A - \tau c)^2 \tag{64}$$

Hence the second cross-derivative is:

$$\pi_{tc} = -x + \frac{\tau c}{2b} = -\frac{1}{2b}(A - 2\tau c) \tag{65}$$

This is clearly positive for $c \geq \frac{A}{2\tau}$ and negative for $c \leq \frac{A}{2\tau}$, which proves the Lemma. $\square$

9.8 Proof of Proposition 5

The first-order conditions for output x and investment k are:

$$p - C - t + xp' = 0 \tag{66}$$

$$-xC_k - F' = 0 \tag{67}$$

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

$$\begin{bmatrix} 2p' + xp'' & -C_k \\ -C_k & -(xC_{kk} + F'') \end{bmatrix} \begin{bmatrix} dx \\ dk \end{bmatrix} = \begin{bmatrix} C_c dc + dt \\ xC_{kc} dc \end{bmatrix} \tag{68}$$

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 $D = -\left(2p' + xp''\right)\left(xC_{kk} + F''\right) - C_k^2$, must be positive. Solving for the effect of the cost parameter on output and substituting into π_{tc} gives equation (36). $\square$

9.9 Proof of Proposition 6

Specializing equation (36) to the investment cost functions in (38) gives:

$$\pi_{tc} = D^{-1} cx \left[\phi\phi'' - \left(\phi'\right)^2\right] \tag{69}$$

Since $\frac{d \log \phi}{dk} = \frac{\phi'}{\phi}$ and so $\frac{d^2 \log \phi}{dk^2} = \frac{\phi\phi'' - (\phi')^2}{\phi^2}$, a positive value for (69) is equivalent to ϕ being log-convex. $\square$

References

ACEMOGLU, D., P. ANTRÀS, AND E. HELPMAN (2007): "Contracts and technology adoption," *American Economic Review*, 97(3), 916–943.

ANTRÀS, P., AND E. HELPMAN (2004): "Global Sourcing," *Journal of Political Economy*, 112(3), 552–580.

ARKOLAKIS, C. (2010a): "Appendix to: Market Penetration Costs and the New Consumers Margin in International Trade," web appendix, Yale University.

——— (2010b): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118(6), 1151–1199.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New trade models, same old gains?," *American Economic Review*, 102(1), 94–130.

ARROW, K. J., AND F. HAHN (1971): *General Competitive Analysis*. San Francisco: Holden-Day.

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

BEHRENS, K., G. MION, AND G. I. OTTAVIANO (2011): "Economic Integration and Industry Reallocations: Some Theory with Numbers," in M.N. Jovanovic (ed.): *International Handbook of Economic Integration*, Volume II, Cheltenham, UK: Edward Elgar.

BERTOLETTI, P., AND P. EPIFANI (2012): "Monopolistic Competition: CES Redux," Discussion paper, University of Bocconi.

BUSTOS, P. (2011): "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, 101(1), 304–340.

Butters, G. R. (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44(3), 465–491.

Costinot, A. (2007): "Heterogeneity and Trade," University of California at San Diego, Economics Working Paper Series.

——— (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.

Davies, R. B., and C. Eckel (2010): "Tax Competition for Heterogeneous Firms with Endogenous Entry," *American Economic Journal: Economic Policy*, 2(1), 77–102.

Egger, H., and U. Kreickemeier (2009): "Firm Heterogeneity and the Labor Market Effects of Trade Liberalization," *International Economic Review*, 50(1), 187–216.

Feenstra, R. C. (2003): "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity," *Economics Letters*, 78(1), 79–86.

Forslid, R., T. Okubo, and K. H. Ulltveit-Moe (2011): "International Trade, CO2 Emissions and Heterogeneous Firms," CEPR Discussion Paper DP8583, Centre for Economic Policy Research.

Ghironi, F., and M. J. Melitz (2005): "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Quarterly Journal of Economics*, 120(3), 865–915.

Grossman, G. M., and E. Helpman (2002): "Integration Versus Outsourcing in Industry Equilibrium," *Quarterly Journal of Economics*, 117(1), 85–120.

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

GROSSMAN, G. M., AND C. SHAPIRO (1984): "Informative Advertising with Differentiated Products," *Review of Economic Studies*, 51(1), 63–81.

GROSSMAN, S., AND O. HART (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, pp. 691–719.

HALLAK, J., AND J. SIVADASAN (2009): "Firms' Exporting Behavior under Quality Constraints," NBER Working Paper No. 14928.

HART, O., AND J. MOORE (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, pp. 1119–1158.

HELPMAN, E., O. ITSKHOKI, M.-A. MUENDLER, AND S. J. REDDING (2012): "Trade and Inequality: From Theory to Estimation," NBER Working Paper No. 17991.

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

KINGMAN, J. F. C. (1961): "A Convexity Property of Positive Matrices," *Quarterly Journal of Mathematics*, 12(1), 283–284.

KLEINERT, J., AND F. TOUBAL (2006): "Distance Costs and Multinationals' Foreign Activities," Centre for Economic Institutions Working Paper No. 2006-6, Hitotsubashi University.

——— (2010): "Gravity for FDI," *Review of International Economics*, 18(1), 1–13.

KOHLER, W., AND M. SMOLKA (2011): "Sourcing premia with incomplete contracts: theory and evidence," *The BE Journal of Economic Analysis & Policy*, 11(1).

Krugman, P. R. (1979): "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9(4), 469–479.

Leahy, D., and C. Montagna (2009): "Outsourcing vs FDI in Oligopoly Equilibrium," *Spatial Economic Analysis*, 4(2), 149–166.

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.

Marshall, A. (1920): *Principles of Economics, An Introductory Volume*. London: Macmillan, eighth edn.

Maskin, E. S., and K. W. Roberts (2008): "On the Fundamental Theorems of General Equilibrium," *Economic Theory*, 35(2), 233–240.

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.

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, pp. 157–180.

Mrázová, M., and J. P. Neary (2010): "Firm Selection into Export-Platform Foreign Direct Investment," presented to the ETSG Annual Conference, Lausanne.

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

——— (2009): "Putting the 'New' into New Trade Theory: Paul Krugman's Nobel Memorial Prize in Economics," *Scandinavian Journal of Economics*, 111(2), 217–250.

——— (2010): "Two and a Half Theories of Trade," *The World Economy*, 33(1), 1–19.

Nefussi, B. (2006): "Exports versus FDI: Evidence from Two French Industries," Working paper, Crest-INSEE.

Negishi, T. (1961): "Monopolistic Competition and General Equilibrium," *Review of Economic Studies*, 28(3), 196–201.

Oldenski, L. (2012): "Export Versus FDI and the Communication of Complex Information," *Journal of International Economics*, 87(2), 312–322.

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.

Porter, L. A. (2012): "Asymmetric Oligopoly and Foreign Direct Investment: Implications for Host-Country Tax-Setting," *International Economic Journal*, 26(2), 229–246.

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

Stole, L. A., and J. Zwiebel (1996): "Intra-Firm Bargaining under Non-Binding Contracts," *Review of Economic Studies*, 63(3), 375–410.

Williamson, O. (1975): *Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization.* New York: Free Press.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012): "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," 80(6), 2765–2784.