SELECTION EFFECTS WITH HETEROGENEOUS FIRMS*

Monika Mrázová†
University of Geneva
and CEPR

J. Peter Neary‡
University of Oxford,
CEPR and CESifo

January 10, 2017

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 much less so: more efficient firms always select the entry mode with lower market-access costs if firms' maximum profits are supermodular in production and market-access costs, but not necessarily otherwise. We derive microfoundations for supermodularity in a range of canonical models. Notable exceptions include horizontal and vertical FDI with "sub-convex" demands (i.e., less convex than CES), fixed costs increasing with productivity, 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

---

*We are very grateful to Arnaud Costinot for extensive comments and discussions, and to Pol Antràs, Richard Baldwin, Paola Conconi, Jonathan Dingel, Pawel Dziewulski, Carsten Eckel, Peter Egger, Rob Feenstra, Gene Grossman, Willy Kohler, Oleg Itskhoki, Dermot Leahy, Marc Melitz, Giordano Mion, Toshihiro Okubo, Lindsay Oldenski, Emanuel Ornelas, Gianmarco Ottaviano, Mathieu Parenti, Robert Ritz, Jacques Thisse, Rick van der Ploeg, Chunan Wang, Adrian Wood, Stephen Yeaple, Krešimir Žigić, and participants at various seminars and conferences for helpful comments. Monika Mrázová thanks the ESRC, grant number PTA-026-27-2479, and the Fondation de Famille Sandoz for funding under the "Sandoz Family Foundation - Monique de Meuron" Programme for Academic Promotion. Peter Neary thanks the European Research Council for funding under the EU's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement no. 295669.

†Geneva School of Economics and Management (GSEM), University of Geneva, Bd. du Pont d'Arve 40, 1211 Geneva 4, Switzerland; e-mail: monika.mrazova@unige.ch.

‡Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK; e-mail: peter.neary@economics.ox.ac.uk.

1 Introduction

Why do different firms choose to serve particular markets in different ways? Not much more than 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.[1] In the intervening period, 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. 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 Lileeva and Trefler (2010) and 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, including international macroeconomics, international tax competition, and environmental economics.[2]

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 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 monotone comparative statics to examine the responses of such firms to exogenous shocks, especially

[1] See, for example, Clerides, Lach, and Tybout (1998) and Bernard and Jensen (1999).

[2] See Ghironi and Melitz (2005), Davies and Eckel (2010), and Forslid, Okubo, and Ulltveit-Moe (2011), respectively.

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 monotone 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, existing approaches typically solve the model in full with each new set of assumptions, and, as a result, the question of robustness of selection effects to functional form has been relatively little explored. A different dimension of robustness is symmetry: existing studies often assume that countries are identical, both in size and in the distribution of firm productivities. Does this matter for the results? 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 introduce a distinction 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 either by exporting or by foreign direct investment.

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 second substantive contribution is a general result on firm selection which provides a sufficient condition for what we call the "conventional sorting" pattern: more efficient firms select into activities with lower marginal costs. To fix ideas, we first present this result in a simple though canonical context, where firms choose between serving a foreign market by either exports or horizontal FDI. We then show that the result extends to a wide variety of other firm decisions, including vertical FDI and choice of technique. In all cases, the key consideration is whether firms' optimal profits are supermodular in each firm's own marginal cost of production and in 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.

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. In this context, our third main contribution is to provide microfoundations for supermodularity in a wide range of models.

Our paper is the first to take a systematic approach to second-order selection effects. However, the non-robustness of standard results when demands are not CES has been independently pointed out in particular models on at least three occasions. Mukherjee (2010) appears to be the first published paper to show that the Helpman, Melitz, and Yeaple (2004) result on selection effects when firms choose between exports and FDI may not hold with linear demands. Nefussi (2006) shows the same result by solving for industry equilibrium in the linear-demand monopolistically competitive model of Melitz and Ottaviano (2008), extended to the choice between exports and FDI. Finally, Spearot (2012) and Spearot (2013) show that the largest firms may choose not to invest in new capital or to engage in domestic acquisitions, deriving in these contexts one of our general results in Section 5 below.[3]

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

[3] See Proposition 4, Condition 2 below. The text of Spearot's papers focuses on the case of linear demands, and the appendices extend the results to general demands.

in particular the concept of supermodularity, to illuminate issues in international trade. 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; and Limão (2005), who considers links between trade and environmental agreements. Closer to our approach is Section 7 of Costinot (2007), the unpublished version of Costinot (2009), which shows that the models of Antràs and Helpman (2004) and Helpman, Melitz, and Yeaple (2004), that assume CES preferences and a Pareto distribution of firm productivities, exhibit log-supermodularity, which permits a compact and elegant restatement of their results. Finally, a recent paper by Bache and Laugesen (2015) explores how firm-level complementarities affect adjustment to shocks at firm and industry level, though without considering the microfoundations for supermodularity which is our main focus.

The plan of the paper is as follows. To set the scene, Section 2 gives an intuitive introduction to first- and second-order selection effects, and summarizes the results from the literature on monotone comparative statics that we use in the paper. Section 3 shows that first-order selection effects arise naturally in a wide range of models, and are not sensitive to assumptions about functional form. Sections 4 turns to second-order selection effects, and derives our central result which relates supermodularity of the ex post profit function to firms' choices between alternative modes of serving a market. The remainder of the paper explores the microfoundations of supermodularity in a wide range of contexts, both old (including some of the most widely-used models in international trade), and new. Section 5 shows that the case where production and marginal access costs are multiplicative includes many important models, including horizontal FDI with iceberg transport costs, vertical FDI, and dichotomous choice of technique. This section then characterizes the properties of demand which are necessary and sufficient for supermodularity to hold in such cases. Section 6 explores the implications of separability between production and access costs, and relates this to the specification of transport costs in models of FDI. Section 7 turns to consider the implications of heterogeneous and endogenous fixed costs of production. Section 8 concludes, while the appendices provide proofs of all results and note some extensions. In particular, Appendix H shows that similar results apply in oligopoly.

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 increase with productivity, and R&D with threshold effects.

2 Preliminaries

Section 2.1 gives some intuition for our main findings, while Section 2.2 reviews some useful results on monotone comparative statics.

2.1 First- and Second-Order Selection Effects

Figure 1 reproduces a well-known diagram from Helpman, Melitz, and Yeaple (2004), which illustrates the pattern of selection effects when firms can serve a foreign market either by exports (E) or FDI (F). (The derivations underlying the figure are standard, and are summarized in Appendix A.) The two schedules show how total profits Π under the two market-access modes vary across firms as a function of an appropriate transformation of productivity, equal to the inverse of marginal cost c. As drawn, these schedules embody three crucial assumptions. First, the fixed cost of FDI, f_F, is strictly greater than that of exporting, f_E. Second, the marginal access cost of FDI is strictly less than that of exporting. Third, firms face a CES demand function with elasticity of substitution σ, which implies that both schedules are linear in $c^{1-\sigma}$. The second and third assumptions together imply that the schedule representing total profits under FDI, Π^F, is always more steeply-sloped than that representing total profits under exports, Π^E. Adding the first assumption implies that the two schedules exhibit a single-crossing condition: they have one and only one intersection.

The lower row of labels in Figure 1 shows the now-familiar pattern of selection effects across firms that these assumptions imply. There are two threshold cost levels, c_E and c_F; firms with marginal costs above c_E, represented by points to the left of $c_E^{1-\sigma}$, choose not to serve the market (i.e., to "exit"); those with marginal costs between c_E and c_F choose to export (E); while those with very low marginal costs, below c_F, choose to serve the market by FDI (F).

A key insight of the present paper is a complementary view of selection effects given by the upper row of labels in Figure 1. This shifts attention from the ex post choices of firms to the ex ante decisions they face. Central to this alternative view is a different threshold

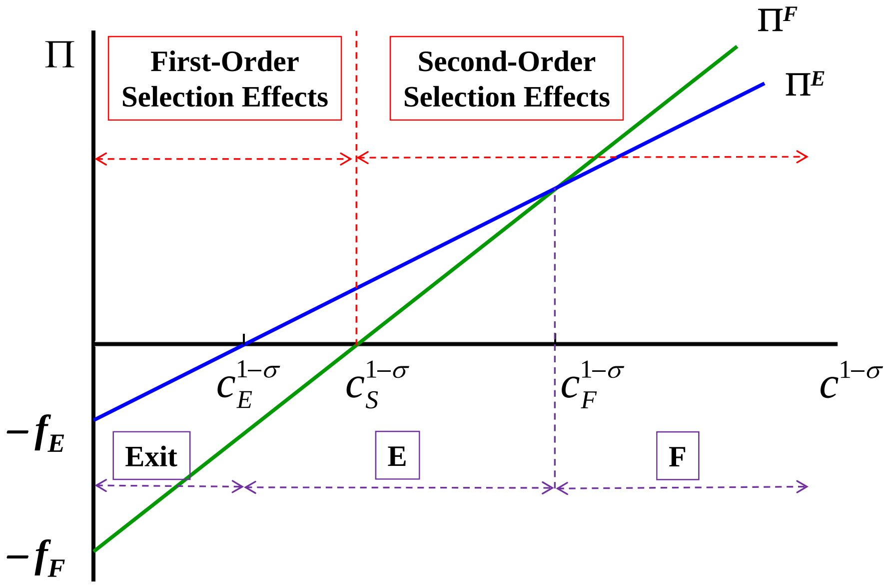


Figure 1: First- and Second-Order Selection Effects

level of marginal cost, denoted by c_S, which divides firms into two groups. To the left are low-productivity firms which never find it profitable to engage in FDI. We say that such firms exhibit a "first-order selection effect", since they face a choice between doing something or doing nothing.[4] By contrast, higher-productivity firms with marginal costs below c_S exhibit a "second-order selection effect", since they face a choice between two alternative ways of serving the foreign market, both of them profitable.

Our alternative perspective becomes most useful when we relax the assumption of CES preferences. Now there may be no transformation of productivity such that profits are a linear function of it. On the one hand, this does not affect the prediction of first-order selection effects: provided only that profits are increasing in productivity, the range of productivities to the left of $c_S^{1-\sigma}$ exhibits the single-crossing property; there is a single cut-off $c_E^{1-\sigma}$ separating firms that exit from those that export. By contrast, the prediction of a single clear ranking of second-order selection effects is less robust: depending on the microfoundations of the profit functions, total profits under FDI need not always increase with productivity more rapidly than under exports. Hence the Π^F and Π^E loci may intersect more than once, and the most efficient firms may earn higher profits from exporting than from FDI. (Figure 3 in Section 5 below gives a diagrammatic example.) In the remainder of the paper, we make these arguments more rigorously, and show

[4] We assume, as is natural, that "doing nothing" yields the same profit for all firms irrespective of their productivity.

how they apply in many contexts other than the choice between exports and FDI.

2.2 Monotone Comparative Statics

To formalize the arguments in the previous sub-section, we begin by summarizing some definitions and results from the literature on monotone comparative statics. Proofs can be found in Topkis (1978), Vives (1990), Milgrom and Shannon (1994) and Vives (1999).

Consider first a function $\pi(x, z)$ which depends on two scalar variables x and z; in many applications x is a choice variable and z is an exogenous parameter. We do not assume that π is differentiable or that x is continuous. We will make frequent use of the assumption that $\pi(x, z)$ is *supermodular*:

Definition 1. The function $\pi(x, z)$ is supermodular in (x, z) if:

$$\pi(x_1, z_1) - \pi(x_1, z_2) \geq \pi(x_2, z_1) - \pi(x_2, z_2) \qquad \text{when} \qquad x_1 \geq x_2,\ z_1 \geq z_2. \tag{1}$$

When the first inequality in (1) is reversed, the function is *submodular*; while a function that is both super- and submodular is called *modular*. The usefulness of this assumption is that it provides a sufficient condition for comparative statics results, as the following result shows:[5]

Result 1. *If $\pi(x, z)$ is supermodular in (x, z), then the optimal choice $x^*(z)$ is increasing in z.*

To see the power of this result, it is helpful to relate it to the case where π is continuous and differentiable. In this case supermodularity is equivalent to the assumption that x and z are Edgeworth complements:[6]

Result 2. *If $\pi(x, z)$ is continuous and twice differentiable, then it is supermodular in (x, z) if and only if $\pi_{xz} \geq 0$.*

Many of the functions we will consider are differentiable, so this result allows us to test for supermodularity just by differentiating. To relate it to Result 1, recall the usual way of deriving

[5] This result can be strengthened to a necessary and sufficient condition in different ways. One route is to assume that $\pi(x, z)$ satisfies the single-crossing property in z rather than supermodularity in (x, z): see Theorem 4 of Milgrom and Shannon (1994). Another, used in earlier versions of this paper, is to focus on total profits $\pi(x, z) - f$, which are quasi-linear with respect to f, and to require the result to hold for all f: see Theorem 10 of Milgrom and Shannon (1994). In most applications we can restrict attention to functions that are either supermodular or submodular throughout the interval $x \in \{x_1, x_2\}, z \in \{z_1, z_2\}$. In such cases, supermodularity is necessary as well as sufficient for the optimal choice $x^*(z)$ to be increasing in z.

[6] We use subscripts to denote partial derivatives: $\pi_{xz} = \frac{\partial^2 \pi}{\partial x \partial z}$ etc.

comparative statics results: this applies the implicit function theorem to the first-order condition $\pi_x(x, z) = 0$ and invokes the second-order condition $\pi_{xx} < 0$:

$$\frac{dx^*}{dz} = -\frac{\pi_{xz}}{\pi_{xx}} \tag{2}$$

Hence x^* is increasing in z if and only if π_{xz} is positive. Result 1 generalizes this familiar result to cases where π need not be differentiable, x need not be continuous, and the second-order condition need not hold.

Definition 1 and Result 1 were stated in terms of scalar arguments. We can extend them to cases where x is a vector as follows.[7] First, the definition of supermodularity is extended to all pairs of scalar cross-differences: $\tilde{\pi}(x, y, z)$ is supermodular in (x, y, z) if Definition 1 holds for each of the pairs (x, y), (x, z), and (y, z). Next, we can invoke the property that supermodularity continues to hold when some arguments of a function are chosen optimally:[8]

Result 3. *If $\tilde{\pi}(x, y, z)$ is supermodular in (x, y, z), then $\pi(x, z) \equiv \max_y \tilde{\pi}(x, y, z)$ is supermodular in (x, z).*

So, supermodularity is preserved when a subset of the endogenous variables is optimally chosen. Combining this with Result 1 gives a further result:

Result 4. *If $\tilde{\pi}(x, y, z)$ is supermodular in (x, y, z), then the optimal choice $x^*(z)$ is increasing in z.*

As we will see, supermodularity of the profit function turns out to be a crucial sufficient condition for selection effects to have the conventional sign in models of monopolistic competition with heterogeneous firms. We will therefore want on many occasions to test whether a particular profit function is supermodular or not. Doing so in practice is greatly helped by the following results:

Result 5. *The function $\pi(x, z) = \gamma\theta(x, z) + \nu\zeta(x, z)$ is supermodular in (x, z) if $\theta(x, z)$ and $\zeta(x, z)$ are supermodular in (x, z).*[9]

[7] More generally, following Topkis (1978), 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 $x \vee y \equiv \inf\{z \mid z \geq x, z \geq y\}$ is the least upper bound of x and y, and $x \wedge y \equiv \sup\{z \mid z \leq x, z \leq y\}$ is their greatest lower bound. This is equivalent to Definition 1 when we set: $x = \{x_1, z_2\}$ and $y = \{x_2, z_1\}$.

[8] See Proposition 4.3 of Topkis (1978)

[9] This is Theorem 7(i) in Milgrom and Shannon (1994).

This is often convenient, since it allows us to determine whether a function is supermodular by inspecting its component functions. A corollary will prove particularly useful:

Result 6. *The function* $\pi(x,z) = \gamma\theta(x,z) + \nu(x) + \zeta(z)$ *is supermodular in* (x,z) *if and only if* $\theta(x,z)$ *is supermodular in* (x,z)*.*

This allows us to ignore multiplicative constants and additive functions that depend on only one of x or z in determining whether a function is supermodular. The final result is a useful decomposition of the second cross-partial derivative:

Result 7. *If* $\tilde{\pi}(x,y,z)$ *is continuous and twice differentiable and* $\pi(x,z) \equiv \max_y \tilde{\pi}(x,y,z)$*, then:*

$$\pi_{xz} = \tilde{\pi}_{xz} + \tilde{\pi}_{xy}\frac{\partial y^*(x,z)}{\partial z} = \tilde{\pi}_{xz} - \tilde{\pi}_{xy}\tilde{\pi}_{yy}^{-1}\tilde{\pi}_{yz} \tag{3}$$

This expression decomposes the second derivative π_{xz} into a direct effect (holding y constant) and an indirect one (allowing for induced changes in y), and will prove useful in providing intuition for some of our results.[10]

3 First-Order Selection Effects

Armed with the results in the previous section, we are now ready to consider their implications for selection effects by heterogeneous firms. Consider a profit-maximizing firm which contemplates serving a particular market. Doing so incurs a fixed cost $f > 0$, 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.[11] Operating profits also depend on a range of decisions taken by the firm in this and all other markets, including prices and sales 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

[10] This result has been used, for example, in the theory of household behavior under rationing: see Neary and Roberts (1980).

[11] We discuss access costs in Sections 5 and 6.

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.[12]

To apply Result 1, we use two stratagems from the monotone comparative statics toolkit. First, note that the choice variable x need not be continuous. So we can reinterpret it as a discrete choice variable X, taking the value one if the firm enters the market and zero otherwise. Second, we write total profits as a function of $-c$ rather than c, so it is presumptively supermodular in its arguments.[13] The total profit function can therefore be written as:

$$\Pi(X, -c) \equiv X\left[\pi(c) - f\right] \tag{4}$$

From Result 1, supermodularity of $\Pi(X, -c)$ in $(X, -c)$ implies that the optimal choice of access mode, $X^*(-c)$, is increasing in $-c$; in other words, that more productive firms choose to serve the market and less productive ones do not. Writing out the condition for supermodularity gives:

$$\Pi(1, -c_2) - \Pi(0, -c_2) \geq \Pi(1, -c_1) - \Pi(0, -c_1) \tag{5}$$

Here and subsequently, without loss of generality, we adopt the convention that $c_1 > c_2$. Since profits are zero if the firm does not enter ($X = 0$), and fixed costs are the same for all firms, (5) is equivalent in turn to:

$$\pi(c_2) \geq \pi(c_1), \quad \forall\ c_1 > c_2 \tag{6}$$

We can therefore conclude the following:

Proposition 1. *Suppose there exists a $c' \in [0, \infty)$ such that $\pi(c') > f$, and a $c'' \in [0, \infty)$ such that $\pi(c'') < f$. Then, if and only if $\Pi(X, -c)$ is supermodular in $(X, -c)$, there exists a $c^* \in [c', c''] \subset [0, \infty)$ such that: for any $c \geq c^*$, we have $\pi(c) \leq f$; and for any $\{c_1, c_2\}$ with $c^* \geq c_1 > c_2$, we have $\pi(c_2) \geq \pi(c_1) \geq f$.*

In words, provided there is at least one firm that makes strictly positive profits and at least one that makes strictly negative profits, and provided (6) holds for all $c_1 > c_2$, then selection takes place, i.e., there is a threshold cost level c^* such that all firms with lower costs enter the market

[12] Heterogeneous firm models with multiple firm characteristics have been considered by Antràs and Helpman (2004), Hallak and Sivadasan (2013), and Harrigan and Reshef (2015).

[13] Milgrom and Shannon (1994) call this kind of sign change, "selective ordering".

and earn positive profits, while those with higher costs do not enter.

Proposition 1 holds whether or not the profit function is differentiable, and whether or not the firm's choice variables are continuous. However, in many practical applications, we can go further if we assume differentiability and continuity. With these extra restrictions, the condition for first-order selection effects is simply that profits are strictly decreasing in marginal cost: $\pi_c < 0$. Moreover, checking whether this condition holds in practice is greatly simplified by invoking the envelope theorem. In most models of heterogeneous firms, the maximum operating profits of a firm in a particular market can be written as the outcome of choosing the optimal values of one or more choice variables:

$$\pi(c) \equiv \max_{x,y} \tilde{\pi}(x, y; c) \tag{7}$$

Here $\tilde{\pi}$ denotes the ex ante operating profit function, maximization of which yields the ex post function π.[14] Differentiating (7) yields the envelope result:

$$\pi_c(c) = \frac{d\tilde{\pi}[x(c), y(c), c]}{dc} = \tilde{\pi}_c[x(c), y(c), c] \tag{8}$$

Hence, provided we can write the firm's objective as an unconstrained maximization problem, it is straightforward to check if Proposition 1 applies.

We show in Appendix B how this approach can be applied to the canonical Melitz (2003) problem of selection into exporting, as well as to selection into marketing as in Arkolakis (2010a), and selection into worker screening as in Helpman, Itskhoki, and Redding (2010). In the first two of these, equation (8) can be immediately applied. By contrast, the firm's problem in the third model is one of maximizing profits subject to constraints, and we need to convert this to an unconstrained maximization problem in order to invoke Proposition 1.

The point of these examples is not just that they extend the original models to arbitrary demand functions; 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 answer is that we focus on cross-section selection effects, comparing the actions of different firms in a

[14] There is no time dimension in the model. What we call the "ex post" profit function could also be described as the profit function conditional on the optimal choice of the endogenous variable, x.

particular industry equilibrium. We do not directly address the much more complex issue of time-series selection effects, comparing the responses of different firms when an equilibrium is perturbed by a shock, though as we show below the former results are an essential prerequisite for signing the latter. By contrast, most models of monopolistic competition with heterogeneous firms compute the industry equilibrium, while at the same time demonstrating that it exhibits both cross-section and time-series selection effects. Our approach in effect *assumes* that an equilibrium exists, and then shows that π decreasing in c is sufficient for the conventional cross-section selection effects to emerge. This is similar to the approach taken by Maskin and Roberts (2008), who show that all the central theorems of normative general equilibrium theory can be proved using elementary methods provided an equilibrium is assumed to exist. Our approach does not address the existence of an equilibrium.[15] Instead, by dispensing with computing one explicitly, it applies without specific restrictions on the functional forms of preferences, technology, or the distribution of costs; and it avoids the need to assume that countries are symmetric.

A different perspective on our results comes from noting that, while they apply to cross-section comparisons in a given equilibrium, they are also a prerequisite for time-series comparisons between equilibria. Such comparisons have been considered in a number of recent papers, but without relating them to the cross-section properties as we do here.[16] To illustrate the issues, we consider only the simplest type of model that exhibits first-order selection effects. Even so, we need to expand the framework already introduced. Denote the ex post operating profits of a firm by $\pi(c, \lambda, \psi)$:

$$\pi(c, \lambda, \psi) \equiv \max_{x,y} \tilde{\pi}(x, y; c, \lambda, \psi) \tag{9}$$

Here, c is marginal cost as before, and ψ is an exogenous variable whose effect on selection we wish to determine. As for λ, it is a demand parameter, exogenous to firms but endogenous

[15] Though this is not a major limitation of our analysis. Existence of 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.

[16] Zhelobodko, Kokovin, Parenti, and Thisse (2012), Bertoletti and Epifani (2014) and Mrázová and Neary (2013) consider time-series selection effects in extensions of the Melitz model to more general demands, while Costinot (2009) and Bache and Laugesen (2015) examine them using the methods of monotone comparative statics.

in industry equilibrium. This specification is consistent with a very broad class of demands which Pollak (1972) calls "generalized additive separability", such that the demand for each good depends on its own price (or quantity in the case of inverse demands) and on a single aggregate.[17] We assume that profits are decreasing in λ, $\pi_\lambda < 0$, so we can interpret λ as a measure of the degree of competition faced by an individual firm.

With firm behavior summarized by the operating profit function, industry equilibrium is determined by two conditions. First is the break-even condition for marginal firms, which requires that their operating profits equal the common fixed cost f:

$$\pi(c_0, \lambda, \psi) = f \tag{10}$$

This implicitly determines the threshold cost c_0 as a function of λ and ψ. Second is the zero-expected-profit condition, which requires that the expected profits of a potential entrant, i.e., the expected value of a firm $\bar{v}(\lambda, \psi)$, should equal the sunk cost of entering the industry f_e. The profits of each firm in turn equal its operating profits net of fixed costs, subject to a lower bound of zero:

$$\bar{v}(\lambda, \psi) \equiv \int_{\underline{c}}^{\bar{c}} v(c, \lambda, \psi) g(c)\, \mathrm{d}c = f_e, \qquad \text{where:} \quad v(c, \lambda, \psi) \equiv \max\left[0, \pi(c, \lambda, \psi) - f\right] \tag{11}$$

and where $g(c)$, with $c \in [\underline{c}, \bar{c}]$, is the density of the distribution of marginal costs. Equation (11) determines the level of competition as a function of the exogenous parameter ψ.

Totally differentiating equations (10) and (11), we can solve for the effect of a change in the exogenous variable ψ on the productivity cutoff c_0:[18]

$$\left(-\pi_c^0\right) dc_0 = \left[\underbrace{\pi_\psi^0}_{D} - \underbrace{\pi_\lambda^0 \left(\bar{v}_\lambda\right)^{-1} \bar{v}_\psi}_{C}\right] d\psi \tag{12}$$

This equation determines time-series selection effects: the comparative statics of a change in any exogenous variable ψ on the threshold marginal cost c_0. It exhibits two key features.

[17] This class includes directly and indirectly additive preferences as in Dixit and Stiglitz (1977) and Bertoletti and Etro (2017) respectively, where λ equals the marginal utility of income; quasi-linear quadratic preferences as in Melitz and Ottaviano (2008), where λ equals the total sales of all firms; and the demands assumed by Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012) where λ is an aggregate price index.

[18] We use π_c^0 to denote $\frac{\partial \pi(c_0, \lambda, \psi)}{\partial c}$, etc.

First, the expression in parentheses on the right-hand side is in general ambiguous in sign. It consists of a direct effect on the profits of a marginal firm, denoted by "D", and an indirect competition effect, denoted by "C". Which of these effects dominates depends on the nature of the shock and on the details of the model.[19] Second, the coefficient on the left-hand side is exactly the expression in the continuous corollary of Proposition 1. This is the sense in which the cross-section selection effects highlighted in Proposition 1 are an essential prerequisite for their time-series counterparts, even in cases where the sign of the right-hand side of (12) is unambiguous.

4 Second-Order Selection Effects

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 present a general sufficiency condition for such selection effects. To fix ideas, we present the condition in the context of the "proximity-concentration trade-off", already illustrated in the CES case in Figure 1: a firm located in one country contemplates serving consumers located in a foreign country, and can do so either by exports or FDI, where exports incur a higher access cost t.[20] However, as we will see in later sections, this is just a motivating example. The general result in Proposition 2 applies to any situation where firms choose between two mutually exclusive activities, both of which incur a marginal cost of c, while t represents the additional marginal cost of one of the activities, typically the one with higher fixed costs.

As in Section 3, we summarize firm behavior using the ex post profit function, $\pi(t, c)$, which equals the maximum operating profits the firm can earn in the foreign country. This function now depends on t, the access cost (tariffs and transport costs) the firm faces, as well as on its marginal production cost c. We assume that π is non-increasing (though not necessarily continuous) in both t and c. It turns out that the key criterion for second-order selection effects is

[19] For example, a reduction of trade costs has an infinitesimal direct effect (since the profits of a threshold firm are initially zero) but a finite competition effect (as inframarginal firms expand), so such a shock unambiguously lowers the threshold cost parameter for selection into exporting, as shown by Bertoletti and Epifani (2014). By contrast, an increase in the size of the world economy in the absence of trade costs has a negative competition effect which exactly equals the positive direct effect if demands are CES, as in Melitz (2003), but may be less than or greater than it depending on whether demands are more or less convex than CES, as shown by Zhelobodko, Kokovin, Parenti, and Thisse (2012) and Mrázová and Neary (2013).

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

whether $\pi(t, c)$ is supermodular in both exogenous cost parameters. Intuitively, supermodularity of $\pi(t, c)$ in (t, c) 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 from (1) we can see that supermodularity in this case is equivalent to:

$$\pi(t_2, c_2) - \pi(t_1, c_2) \geq \pi(t_2, c_1) - \pi(t_1, c_1) \geq 0 \quad \text{when} \quad t_2 \leq t_1 \;\text{ and }\; c_2 \leq c_1 \tag{13}$$

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 in later sections, this is often the case, but there are important counter-examples.

We now compare the relative profitability of different modes of serving the foreign market. Exporting faces a higher access cost, t, so FDI has the advantage of proximity: for simplicity, we assume that access costs conditional on FDI are zero. However, FDI 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: it equals f_E if the firm exports but f_F if the firm engages in FDI and builds a plant in the foreign market, with f_F strictly greater than f_E. Finally, as in Section 3, we introduce a dummy variable to indicate the mode of market access: X equals one if the firm serves the market via FDI, zero if it exports. With these assumptions, ex post profits can be written as follows:

$$\Pi(X, -c) \equiv \begin{cases} \Pi^F(c) = \pi(0, c) - f_F & \text{when} \quad X = 1 \\ \Pi^E(c) = \pi(t, c) - f_E & \text{when} \quad X = 0 \end{cases} \tag{14}$$

Supermodularity of $\Pi(X, -c)$ in $(X, -c)$ is equivalent to:

$$\Pi(1, -c_2) - \Pi(1, -c_1) \geq \Pi(0, -c_2) - \Pi(0, -c_1) \tag{15}$$

which can be rewritten as:

$$\pi(0, c_2) - \pi(0, c_1) \geq \pi(t, c_2) - \pi(t, c_1) \tag{16}$$

This in turn from (13) is implied by supermodularity of $\pi(t, c)$ in (t, c). Recalling Result 1 gives one of the key results of our paper:[21]

Proposition 2. *If $\pi(t, c)$ is supermodular in (t, c), then the optimal choice $X^*(c)$ is decreasing in c; i.e., higher-cost firms have a higher incentive to select into exports, while lower-cost firms have a higher incentive to select into FDI.*

This result motivates our focus in the remainder of the paper on establishing micro-foundations for supermodularity of $\pi(t, c)$.

Proposition 2 implies only that supermodularity is a sufficient condition for conventional selection effects to hold. A violation of supermodularity does not in itself imply that selection effects are reversed. Nevertheless, we will see many examples later where submodularity of the profit function is consistent with such a reversal. Indeed, just as we have saw in Figure 1 that, in the canonical CES case, the profit function is supermodular at all cost levels consistent with second-order selection effects, so it is also possible for it to be submodular at all such cost levels, implying that standard selection effects are reversed for all firms.[22]

A striking feature of Proposition 2 is that it does not depend directly on fixed costs. While fixed costs affect the difference in profits between exporting and FDI, they vanish when we compare this difference across firms. 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 may not hold. What matters for the direction of second-order selection effects is not a trade-off between fixed and variable costs, but whether there is a complementarity between variable costs of production and of trade. Putting this differently, for FDI to be the preferred mode of market access, a firm must be able to afford the additional fixed costs of FDI, but whether it can afford them or not depends on the cross-effect on profits of tariffs and production costs. When supermodularity prevails, a more efficient firm has relatively higher operating profits in the FDI case, but when submodularity holds, the opposite may hold.

[21] Propositions 1 and 2 could be combined as a single mathematical result. However, it seems economically more insightful to highlight the distinction between first-order and second-order selection effects by writing them as distinct propositions.

[22] A simple example of global submodularity is where the Π^E and Π^F loci intersect twice, but only the second intersection occurs at a productivity level that is high enough for both exports and FDI to be profitable.

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 t and c, as we shall see in Section 7. First, we turn in Sections 5 and 6 to consider two specializations of the profit function $\pi(t, c)$, where the production and access costs c and t enter multiplicatively and separably respectively.

5 Multiplicative Production and Access Costs

In an important class of models, the two cost parameters t and c enter the ex ante operating profit function multiplicatively, so (7) takes the special form:

$$\pi (t, c) \equiv \max_x \tilde{\pi} (x; t, c), \quad \tilde{\pi} (x; t, c) = \{p (x) - tc\} x, \quad t \geq 1 \tag{17}$$

Here $p(x)$ is the inverse demand function taken as given, i.e., "perceived", by the firm: we impose no restrictions other than that it is downward-sloping: $p' < 0$. This specification of $\tilde{\pi}$ can be given at least three interpretations. First is the proximity-concentration trade-off model of horizontal FDI as in Helpman, Melitz, and Yeaple (2004), which is the benchmark case we have considered so far. In this case, t is an iceberg transport cost: for each unit of exports sold abroad, t units must be produced at home. Second, following Antràs and Helpman (2004), equation (17) can represent a model of vertical FDI, where the firm has to choose the location of production in order to serve its home market, rather than the mode of accessing a foreign one. In this case c is the production cost in wage units that must be paid in the low-cost "South", and t is the proportional wage premium that must be paid in the high-cost "North", the firm's home market. Third, following Lileeva and Trefler (2010) and Bustos (2011), equation (17) can represent a model in which each firm chooses between two techniques. The "high" technique has a lower variable cost, given by c, whereas the "low" technique has a higher variable cost, equal to tc. In this interpretation, t is the premium on variable cost which a firm must pay if it does not invest in improving its technology.

These three interpretations of equation (17) have very different economic implications. However, they have the same formal structure, and so Proposition 2 applies in the same way to each. The form that selection takes depends only on whether profits are supermodular in the two components of variable costs. The new feature of the parameterization in (17) is that explicit expressions for supermodularity can be derived, and they depend only on the shape of the demand function $p(x)$.

To see this, 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 $\log x$.[23] 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 Dixit and Stiglitz (1977) and Krugman (1979). It implies plausibly that consumers are more responsive to price changes the greater their consumption; and it encompasses many of the most widely-used non-CES specifications of preferences, including quadratic (to be considered further below), Stone-Geary, and additive exponential or "CARA" preferences.[24] 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.[25] Lemma 7 in Appendix C shows 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 multiplicative costs as in* (17)*,* $\pi(t, c)$ *is supermodular in* (t, c) *at all levels of output if the demand function is weakly superconvex; i.e., if the elasticity of demand is non-decreasing in output,* $\varepsilon_x \geq 0$*.*

[23] For a formal definition, and proofs of the statements that follow, see Appendix C. The term "superconvexity" seems to be used, if at all, as a synonym for log-convexity, i.e., $p(x)$ is log-convex when $\log p$ is convex in x. (See Kingman (1961).) For related discussions, see Neary (2009), Zhelobodko, Kokovin, Parenti, and Thisse (2012), and Bertoletti and Epifani (2014). For the most part, these papers assume that preferences are additively separable, though this is not necessary for cross-section results using our approach, since we only consider the demand function from the firm's perspective.

[24] See Bertoletti (2006) and Behrens and Murata (2007) on the latter.

[25] This corresponds to the superconvex subset of the family of demand functions introduced by Pollak (1971).

To prove Proposition 3, we express the cross-partial derivative of the profit function in terms of the elasticity of demand and its responsiveness to output (see Appendix D):

$$\pi_{tc} = \frac{(\varepsilon - 1)^2 + x\varepsilon_x}{\varepsilon - 1 - x\varepsilon_x}\, x \tag{18}$$

The denominator $\varepsilon - 1 - x\varepsilon_x$ must be positive from the second-order condition. The numerator shows that, whenever ε_x is strictly negative, submodularity may hold for sufficiently high x, which proves the proposition. To see this from a different perspective, we can decompose the right-hand side of (18) into its value in the CES case (when ε equals σ) and a term whose sign depends only on ε_x:

$$\pi_{tc} = (\varepsilon - 1)x + \frac{\varepsilon\varepsilon_x}{\varepsilon - 1 - x\varepsilon_x}\, x^2 \tag{19}$$

Hence, we can be sure that supermodularity holds for all output levels only in the CES and strictly superconvex cases.

Intuitively, the result follows from another implication of superconvexity. A positive 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, $\mathcal{L} \equiv \frac{p-tc}{p}$, since $\mathcal{L} = \frac{1}{\varepsilon}$. 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 additional cost t by engaging in the higher-fixed-cost activity.

Proposition 3 is important in 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:[26]

Proposition 4. *With multiplicative costs as in* (17)*,* $\pi(t, c)$ *is supermodular in* (t, c) *if and only if either of the following equivalent conditions holds:*

[26] Formal proofs are in Appendix E. Spearot (2012), p. 40, derives Condition 2 of Proposition 4 in a model where firms choose between investing in new capital and acquiring a rival. The Online Appendix to Spearot (2013) derives a related result in a model where firms choose between investing in capital at home and abroad.

1. *The elasticity of output with respect to marginal cost is greater than one in absolute value.*

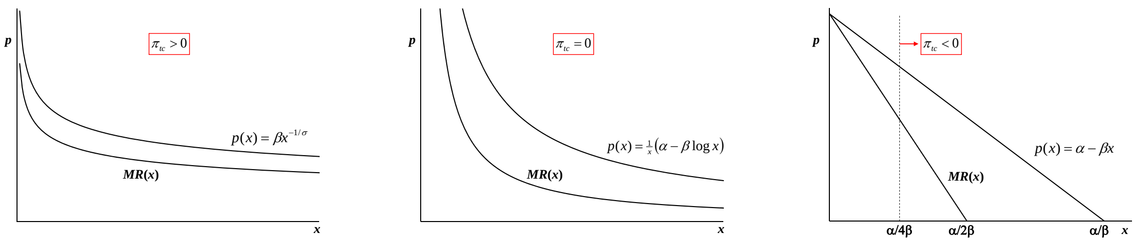
2. *The elasticity of marginal revenue with respect to output is less than one in absolute value.*

3. *The sum of the elasticity,* $\varepsilon \equiv -\frac{p}{xp'}$*, and convexity,* $\rho \equiv -\frac{xp''}{p'}$*, of demand is greater than three.*

Condition 1 of the proposition follows immediately by differentiating the profit function (17) twice, invoking the envelope condition:

$$\pi_t = -cx \quad \Rightarrow \quad \pi_{tc} = -x\left(1 + \frac{c}{x}\frac{dx}{dc}\right) \tag{20}$$

Condition 2 also follows easily by recalling that marginal cost is equal to marginal revenue, so the elasticity of output with respect to marginal cost, $-\frac{c}{x}\frac{dx}{dc}$, is the inverse of the elasticity of marginal revenue with respect to output. Both Conditions 1 and 2 require that a more productive firm raises its output by more, and its marginal revenue falls by less, as costs fall, so the firm enjoys a greater Matthew Effect. As for Condition 3, it is less intuitive, but provides an easy way of testing whether a particular demand function exhibits super- or submodularity at a point. It shows that 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. (Recall that ρ must be less than 2 from the firm's second-order condition.)



(a) CES (b) Inverse Translog (c) Linear

Figure 2: Examples of Demand and Marginal Revenue Functions

To give further intuition for Propositions 3 and 4, consider in turn the three demand and marginal revenue functions shown in Figure 2; detailed derivations are given in Appendix F.

Figure 2(a) shows the CES case. Marginal revenue as a function of sales is highly inelastic: its elasticity is the same as that of the inverse demand function, $1/\sigma$ (equal to $1/4$ in the figure). This implies a very large response of sales to costs: more productive firms sell a lot more than less productive ones, and so gain more from a reduction in trade costs, enjoying a strong Matthew Effect. Hence the profit function is always *supermodular*, as we have already seen.

Figure 2(b) illustrates an inverse translog demand function which is equivalent to a profit function that is always *modular* (i.e., on the boundary between strictly super- and strictly sub-modular):[27]

Lemma 1. *With multiplicative costs as in* (17)*,* $\pi(t, c)$ *is modular in* (t, c) *at all levels of output if and only if the demand function takes an inverse translog form:* $p(x) = \frac{1}{x}(\alpha + \beta \log x)$*.*

In this case, the marginal revenue function is a rectangular hyperbola, $MR(x) = \frac{\beta}{x}$, implying that total variable costs are the same for all firms, irrespective of the tariff they face. Hence the choice between exports and FDI depends on fixed costs only, and since these are assumed to be the same for all firms, no selection will be observed.

Finally, Figure 2(c) illustrates a linear demand function, $p(x) = \alpha - \beta x$, which from the firm's perspective is consistent with any specification of quadratic preferences.[28] The convexity of this demand function is zero, so it follows from Condition 3 of Proposition 4 that there must be a range of costs for which the profit function is *submodular*, corresponding to output levels where the elasticity of demand is less than three. Figure 2(c) illustrates this from a different perspective. Like the inverse demand function, the marginal revenue function is linear, and its elasticity with respect to output rises steadily (.e., its elasticity with respect to price falls, $\varepsilon_x < 0$), as output increases. We show in Appendix F that the elasticity of marginal revenue equals one, and so profits switch from super- to sub-modularity, at half the maximum level of output $x = \frac{\alpha}{4\beta}$ (indicated by the dashed line in Figure 2(c)). Unlike the CES case, the profit function is therefore submodular for low-cost exporters. Hence, provided both exporting and FDI are profitable in the relevant range, it is possible to have 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. Figure 3 illustrates this configuration, which contrasts

[27] This function is implied by a translog distance function, and so from Diewert (1976) it is consistent with a Törnqvist quantity index. It has been used empirically in agricultural economics and other fields where the assumption that prices respond endogenously to pre-determined quantities is plausible.

[28] As always in monopolistic competition, the demand parameters α and β are taken as given by firms, but are endogenous in general equilibrium. For example, in the model of Melitz and Ottaviano (2008), they depend on market size, the mass of firms, and the aggregate price index.

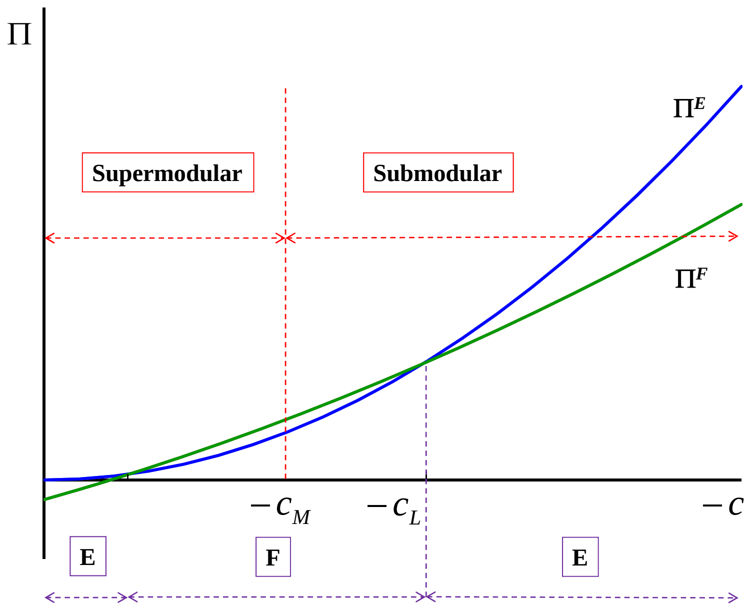


Figure 3: Selection Effects with Linear Demands and Iceberg Transport Costs

with the conventional sorting that we saw in Figure 1 in Section 2.1.[29] The profit function switches from super- to submodularity at a marginal cost of c_M, where the slopes of the two profit schedules are the same: $\Pi_c^E = \Pi_c^F$, implying that $\pi_c(t, c) = \pi_c(1, c)$; while the preferred mode of serving the foreign market switches from FDI to exports at a marginal cost of c_L.

The fact that the point c_M lies to the left of c_L in Figure 3 (so c_M is strictly greater than c_L) illustrates a feature we have already highlighted: submodularity is a necessary condition for the standard selection effects to be reversed, not a sufficient one. However, with linear demands we can say more than that: provided marginal costs can be arbitrarily low, there must be some firms that exhibit unconventional selection effects; in the limit, if marginal production cost is zero, then the marginal cost of producing at home and selling in the export market, tc, is also zero, and the firm gains nothing from engaging in FDI. That is, for linear demand there is a maximum quantity that the consumer will buy, so variable profits for FDI and exports are identical as productivity become infinite. Thus, provided the fixed cost of FDI relative to exports is strictly positive, the most productive firms always engage in exports. This is an important result, as the most productive firms are the largest and typically account for the bulk of exports. Moreover, this pattern is not confined to linear demands; the case of linear demands points to a general class of demand functions that guarantee not just submodularity

[29] To facilitate comparison with Figure 1, c is measured from right to left in Figure 3, starting at zero. See Appendix F for further details.

of the profit function for some firms, but non-conventional selection effects for the most efficient firms. A maximum or satiation level of demand is clearly sufficient; this can be weakened to requiring satiation of the marginal revenue function rather than the demand function:

Lemma 2. *With multiplicative costs as in* (17)*, conventional selection effects must be reversed for firms with unbounded productivity if marginal revenue as a function of price exhibits a satiation level: i.e.,* $\lim_{p \to 0} MR(p)$ *is finite.*

There are many plausible demand functions that exhibit this property.[30] Hence unconventional selection effects are ensured for a broad class of demand functions.

To summarize, the three examples in Figure 2 show that different plausible and widely-used demand functions have very different implications for selection effects. The inverse translog case (b) is a threshold case of mainly theoretical interest, but it is not obvious whether case (a) or (c) is closer to reality. Direct empirical evidence on whether demands are super- or sub-convex is not available, and the indirect evidence is ambiguous: see for example the empirical findings in favor of each cited by Zhelobodko, Kokovin, Parenti, and Thisse (2012). However, there is some direct evidence on selection effects which suggests that the most efficient firms engage less in low-marginal-cost activities than a CES model predicts: see the results of Yeaple (2009) for FDI and Spearot (2013) for investment. The results of this section provide a framework for thinking about these findings and suggest that further empirical work on this question is needed.

6 Separable Production and Access 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."[31] 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

[30] For example, the Pollak family of demand functions mentioned in footnote 25 has a satiation point for all values of σ less than zero; i.e., for any Pollak demand function less convex than the CARA (constant absolute risk aversion) limiting case of $\sigma \to 0$. The linear demand function is the special case of this family where $\sigma = -1$.

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

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 tx units, so the technology of transportation is identical to that of production: $(p - tc)x = px - c(tx)$.

To see how alternative specifications of transport costs affect the outcome, we use Result 7 to decompose the cross-partial derivative of π into direct and indirect effects:

$$\pi_{tc} = \tilde{\pi}_{tc} + \tilde{\pi}_{tx} \frac{\partial x^*(t, c)}{\partial c} = \tilde{\pi}_{tc} - \tilde{\pi}_{tx} \left(\tilde{\pi}_{xx}\right)^{-1} \tilde{\pi}_{xc} \tag{21}$$

The direct effect given by $\tilde{\pi}_{tc}$ is the effect of a difference in production costs on the profit disadvantage of higher transport costs at a *given* level of output; while the indirect effect is 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 $\tilde{\pi}_{tx}\tilde{\pi}_{xc}$. This is presumptively positive; for example it must be so in the case of constant production costs and iceberg transport costs, when $\tilde{\pi}_{tx} = -c$ and $\tilde{\pi}_{xc} = -t$. This is the Matthew Effect from Section 4: it arises because a higher-cost firm is less vulnerable to a rise in transport costs since it has presumptively lower sales: both $\tilde{\pi}_{tx}$ and $\frac{\partial x^*}{\partial c}$ 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 $\tilde{\pi}_{tc} = -x$ 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 ($\tilde{\pi}_t$ is more negative) since its cost of shipping one unit of exports is $(t - 1)\, c$.

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:

$$\tilde{\pi}\,(x; t, c) = r\,(x, t) - cx \tag{22}$$

Here sales revenue accruing to the firm, r, depends in a 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 (21): total transport costs and hence $\tilde{\pi}_t$ do not depend directly on 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, where $r(x,t) = x\,[p(x) - t]$, provide one example of (22). Another is where transport costs are ad valorem or proportional to price, so sales revenue becomes: $r(x,t) = \frac{xp(x)}{t}$. Relative to the case of iceberg transport costs, the firm's first-order condition is unchanged, but profits are deflated by t: $\tilde{\pi}\,(x;t,c) = \left[\frac{p(x)}{t} - c\right] x$. Similar derivations to those already given shows that equation (21) 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. Hence, with quadratic preferences and proportional transport costs, the conventional sorting is fully restored: qualitatively, the configuration of the two profit functions is the same as in Figure 1. Hence the model predicts that the most efficient firms will always engage in FDI rather than exporting. (See Head and Ries (2003) and Mukherjee (2010).)

Note finally that separability of production and access costs is unlikely to hold other than in the case of horizontal FDI with non-iceberg transport costs. In particular, it does not apply in the other two cases considered in Section 5, vertical FDI and choice of technique. In both of these cases the access cost parameter t relates to the supply side, measuring either the wage premium of producing in the North, or the variable-cost premium of not investing in new technology. Hence separability as in (22) does not apply, and the standard selection effects can be reversed when the condition in Proposition 4 does not hold.

7 Second-Order Selection Arising from Fixed Costs

So far, our focus has been on the choice between exports and FDI. However, as we show in this section, the same approach applies to a wide range of other firm choices. We first show how our approach extends to the case where fixed costs differ between locations and between firms. Next, we consider how selection effects can 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 above apply: supermodularity between the firm's own cost parameter and a parameter representing the marginal cost of the mode of accessing a market guarantees the standard selection effect, whereby more productive firms select into the access mode with lower marginal cost, whereas this effect is likely to be reversed for some firms if supermodularity does not hold.

7.1 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. 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. To fix ideas, we illustrate this in the horizontal FDI context, but the issues clearly apply more widely.

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, it continues to vanish when we take the difference across firms of the difference in profits between exports and FDI (or between FDI in two different locations). 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 production costs c as well as with the mode of access. Our approach can still be applied, but some care is needed. Suppose for example that only the fixed cost of FDI varies with c. Then the total profit function from equation (14) becomes:

$$\Pi(X, -c) \equiv \begin{cases} \pi(0, c) - f_F(c) & \text{when } X = 1 \\ \pi(t, c) - f_E & \text{when } X = 0 \end{cases} \tag{23}$$

where $\pi(t, c)$ is the operating profit function from earlier sections. Now, there is an additional reason why supermodularity of $\Pi(X, -c)$ may not hold, depending on how fixed costs vary with productivity. Supermodularity of $\Pi(X, -c)$ in $(X, -c)$ depends not just on supermodularity of $\pi(t, c)$ in (t, c) but also on the difference in fixed costs between a low- and a high-cost firm $f_F(c_1) - f_F(c_2)$. Applying the finite difference operator $\Delta_c f_F(c) \equiv f_F(c_1) - f_F(c_2)$ to total profits (23) gives:

$$\Delta_c \Pi(1, -c) - \Delta_c \Pi(0, -c) = [\Delta_c \pi(t, c) - \Delta_c \pi(0, c)] + \Delta_c f_F(c) \tag{24}$$

The first term in parentheses on the right-hand side is the same as in previous sections, and is positive if and only if $\pi(t, c)$ is supermodular in (t, c). The second term is new, and shows that supermodularity of $\Pi(X, -c)$ is more likely to hold if fixed 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 difference in fixed costs $f_F(c_1) - f_F(c_2)$ becomes $(c_1 - c_2)f$, which is strictly positive for $c_1 > c_2$. Hence, supermodularity of $\Pi(X, -c)$ 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 difference in fixed costs $f_F(c_1) - f_F(c_2)$ is strictly negative for $c_1 > c_2$, so $\Pi(X, -c)$ may be submodular even when $\pi(t, c)$ is supermodular (e.g., even when preferences are CES). 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.

7.2 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. It is desirable to explore whether our approach extends to these cases, where firms face more complex trade-offs. For simplicity, we focus on the case of R&D in what follows. We have already seen in Section 5 that the decision to engage in R&D, conditional on serving a market, illustrates the case of multiplicative production and access costs. In this section we want to explore the more complex decision where, conditional on engaging in R&D, assumed to be specific to a particular foreign market, the firm faces the choice of how much to invest and whether to locate its investment at home or in the target market. As before, we want to understand how differences in productivity between firms affect their choices.

Now we need to define 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 forms of R&D 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, an inverse measure of efficiency, though it no longer corresponds to marginal cost. The average cost function $C(c,k)$ 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:

$$\pi(t,c) \equiv \max_{x,k} \tilde{\pi}(x,k;t,c), \quad \tilde{\pi}(x,k;t,c) = [p(x) - C(c,k) - t]\,x - F(k) \tag{25}$$

To highlight the new issues that arise from investment in R&D, we assume that transport costs are specific, so, unlike in earlier sections, sub-convexity of demand cannot be a source of submodularity here.[32] What matters is the specification of the average cost function $C(c,k)$. To understand the implications of this, we proceed as in Section 5: we derive some general conditions for supermodularity of π in (t,c), and then illustrate with three specific examples.

We first derive a sufficient condition for the profit function (25) to exhibit supermodularity in (t,c), proceeding as before by invoking Results 3 and 6:

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

This sufficient condition could be described as an alignment between differences across firms in static and dynamic efficiency: a positive value of C_{ck} implies that investment raises the cost

[32] If instead we assume iceberg transport costs, then the ex ante variable profit function becomes: $\tilde{\pi}(x,k;t,c) = [p(x) - tC(c,k)]\,x - F(k)$. Supermodularity of the ex post profit function now depends on $\pi_{tc} = \tilde{\pi}_{tc} + \tilde{\pi}_{t\nu}\tilde{\pi}_{\nu\nu}^{-1}\tilde{\pi}_{\nu c}$, where $\nu = [x\ k]'$, so submodularity can arise if *either* the demand function or the average cost function exhibits "too little" convexity.

advantage of a higher-productivity firm; equivalently, a positive value of C_{kc} implies that a lower-productivity firm benefits less from investment (C_k is less negative). Note an immediate corollary: if firms differ only in the level of the average cost function, denoted by c_0 in what follows, then the conventional sorting will always be observed, since $C(c_0, c, k) = c_0 + \tilde{C}(c, k)$ is weakly supermodular in (c_0, k), which from Proposition 5 is sufficient for $\pi(t, c_0, c)$ to be supermodular in (t, c_0).

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

$$C\left(c,k\right) = c_0 + c\phi(k), \ \phi' < 0, \quad \text{and} \quad F'' = 0 \tag{26}$$

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

Proposition 6. *$\pi\left(t,c\right)$ in (26) is supermodular in $\left(t,c\right)$ if and only if $\phi\left(k\right)$ is log-convex in k.*

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

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

Example 1. [Linear Returns to R&D] d'Aspremont and Jacquemin (1988) assume that the average cost function is linear in k, as illustrated in Panel (a) of Figure 4, while fixed costs are quadratic in k:[33]

$$C\left(c,k\right) = c_0 - c^{-1}k \qquad F(k) = \tfrac{1}{2}\gamma k^2 \tag{27}$$

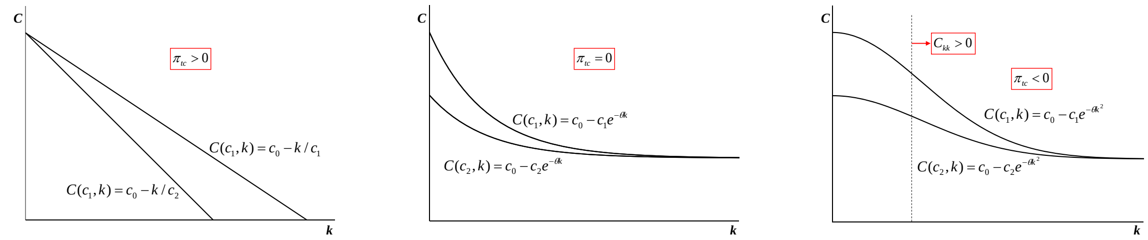
Since $C_{ck} = c^{-2}$, which is positive, supermodularity is assured for this specification of R&D costs.

The remaining two examples illustrate the applicability of Proposition 6:

Example 2. [Exponential Returns to R&D] Constant returns to investing in R&D as in the d'Aspremont-Jacquemin specification is an implausible feature.[34] A more attractive and

[33] 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.

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



(a) Linear Returns to R&D (b) Exponential Returns (c) Threshold Effects in R&D

Figure 4: Examples of Average Production Costs as a Function of Investment in R&D

only slightly less tractable alternative due to Spence (1984) is also widely used:[35]

$$C\left(c,k\right)=c_0+ce^{-\theta k}\qquad F(k)=k \tag{28}$$

In this case investment lowers average production costs ($C_k=-\theta ce^{-\theta k}<0$) but at a diminishing rate ($C_{kk}=\theta^2 ce^{-\theta k}>0$), as illustrated in Panel (b) of Figure 4; while fixed costs increase linearly in k ($F''=0$). 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.[36] Expressed in terms of Proposition 6, equation (28) is a special case of (26), with $\phi(k)=e^{-\theta k}$. Hence ϕ is log-linear in k (since $\frac{d\log\phi}{dk}=-\theta$), implying that $\pi\left(t,c\right)$ is *modular*, i.e., both supermodular and submodular. 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.[37] 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, all firms will adopt the same mode of serving the foreign market, so no

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

[36] 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 $-\theta$.

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

selection effects will be observed.

Example 3. **[R&D with Threshold Effects]** The fact that the specification due to Spence is just on the threshold between super- and submodularity has implausible implications as we have seen. It also implies from Proposition 6 that a less convex average cost function would yield submodularity. Such a specification is found by generalizing that of Spence as follows:

$$C(c,k) = c_0 + ce^{-\theta k^a},\ a > 0 \qquad F(k) = k \tag{29}$$

In this case $\phi(k) = e^{-\theta k^a}$ and so $\frac{d^2 \log \phi}{dk^2} = -\theta a(a-1)k^{a-2}$. This is negative for $a > 1$, implying that profits are submodular in (t, c). This case is illustrated in Panel (c) of Figure 4.[38] As shown, the average cost function is initially concave in k and then becomes convex.[39] This justifies calling this specification one of *threshold effects in R&D*: low levels of investment have a relatively small effect on production costs whereas higher levels yield a larger payoff. In the FDI context this implies that firms will select into different modes of market access in exactly the opposite way to Proposition 2. Since profits are submodular in t and c, less efficient firms have a greater incentive to establish a 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.

[38] This is drawn for the Gaussian special case of $a = 2$.

[39] From (29), $C_{kk} = -\theta ack^{a-2}e^{-\theta k^a}(a - 1 - \theta ak^a)$. 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. It is illustrated by the vertical dashed line in the figure. 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.

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 each profitable in itself. 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 a high-cost and a low-cost firm depend only on the *difference* in their ex post profits. This difference is presumptively negative in all models (though proving this 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 (2010a); 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).

Our second main contribution is to show that second-order selection effects are much less robust, and depend on the *difference in differences* of ex post profits with respect to marginal production costs and the marginal access cost of the two modes of serving the market. 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. By contrast, if profits are submodular, the reverse sorting pattern may occur.

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.

The third main contribution of our paper is to identify microfoundations for supermodularity in a range of canonical settings. 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 producing at home versus offshoring to a lower-wage location, or between adopting high- or low-tech modes of production, our result continues to hold, even in the absence of transport costs. It implies that, for most non-CES preferences, the most productive firms will select into producing at home, where their greater efficiency offsets the higher wage penalty they incur, or will refrain from investing in cost-reducing technology since their production costs are so low to begin with. 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 addition to the cases considered in the paper, 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.[40]

[40] For a recent example, see Bernard, Blanchard, Van Beveren, and Vandenbussche (2016).

Appendices

A Derivations underlying Figure 1

Firms face a CES demand function, which we write in inverse form as $p(x) = \beta x^{-1/\sigma}$. Here, σ is the elasticity of substitution in demand, which must be greater than one; and β is a catch-all term, common to all firms, which summarizes how each firm's perceived demand depends on market-level variables such as total expenditure and the prices of all other goods. The value of β is determined endogenously in industry equilibrium, but is taken as given by each firm in choosing its optimal production and sales. Standard derivations show that the maximum value of variable profits is:

$$\pi(t, c) = \max_x (\beta x^{-1/\sigma} - tc)x = B(tc)^{1-\sigma} \tag{30}$$

where $t > 1$ is an iceberg transport cost, and $B = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}\beta^\sigma$. Consider now the properties of Figure 1, where total profits from exports and FDI are $\Pi^E = \pi(t, c) - f_E$ and $\Pi^F = \pi(1, c) - f_F$ respectively. The assumption that $f_E > 0$ ensures that $c_E^{1-\sigma} > 0$. Assuming in addition that $t^{1-\sigma} f_F > f_E$ (which is stronger than $f_F > f_E$), implies the configuration $c_E^{1-\sigma} < c_S^{1-\sigma} < c_F^{1-\sigma}$ as shown in Figure 1.

B Derivations for Section 3

In this section, we show how Proposition 1 can be used in some well-known models of heterogeneous firms to extend the results on first-order selection effects to general demands.

B.1 *Example 1*: Selection into Exports

Relaxing the assumption of CES preferences in the Melitz (2003) model of selection into exporting, each firm's maximum operating profits from (7) can be written as follows:

$$\pi(c) \equiv \max_x \tilde{\pi}(x; c), \qquad \tilde{\pi}(x; c) = \{p(x) - tc\} x \tag{31}$$

Here $p(x)$ is the inverse demand function taken as given, i.e., "perceived", by the firm: we impose no restrictions other than $p' < 0$; and $t \geq 1$ is an iceberg transport cost, representing the number of units which must be produced in order to deliver one unit to consumers. It is straightforward to check that equation (31) satisfies the conditions for Proposition 1. Differentiating (31) and invoking the envelope theorem as in (8), shows that profits are decreasing in c: $\pi_c = \tilde{\pi}_c = -\tau x < 0$. It follows that first-order selection effects always hold in the Melitz model, irrespective of the form of the demand function.[41]

As a corollary, we note from Result 1 that the optimal level of output x^* is decreasing in c, and hence so are profits, since they are increasing in $(x, -c)$. For completeness, we derive explicit expressions for these responses. The first-order condition sets marginal revenue equal to marginal cost: $p + xp' = tc$. Totally differentiating gives: $(2p' + xp'')\, dx = tdc$. Reexpressing in terms of proportional changes (denoted by a "hat" over a variable, e.g., $\hat{x} \equiv d \log x, x > 0$), and using the first-order condition to eliminate c, gives $\hat{x}$. A similar elimination of tc yields $\hat{\pi}$:

$$\hat{x} = -\frac{\varepsilon - 1}{2 - \rho}\hat{c}, \qquad \hat{\pi} = -(\varepsilon - 1)\hat{c} \tag{32}$$

Here $\varepsilon \equiv -\frac{p}{xp'}$ and $\rho \equiv -\frac{xp''}{p'}$ denote the elasticity and the convexity of the demand function respectively. 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 (32) simplifies to $\hat{x} = -\sigma\hat{c}$. Equation (32) confirms that both output and profits are strictly decreasing in c for $c > 0$, since the firm's first-order condition requires that $\varepsilon \geq 1$, with $\varepsilon > 1$ for $c > 0$, and the second-order condition requires that $\rho < 2$.

B.2 *Example 2*: Selection into Marketing

To see how our approach can be applied more widely, consider the case of a firm that must engage in marketing expenditure in order to reach consumers. Following Arkolakis (2010a), it faces two decisions: what price p to charge, or, equivalently, how much to sell per consumer, x;

[41] To see this explicitly, $\pi_c = \tilde{\pi}_c = -\tau x < 0$ implies from Result 3 that the ex post profit function $\Pi(X, -c) \equiv X\,[\pi(c) - f]$ is supermodular in $(X, -c)$. This implies from Result 1 that the optimal choice X is decreasing in c and so Proposition 1 holds: more productive firms select into serving a market, whether domestic ($t = 1$) or foreign ($t > 1$), for all downward-sloping demand functions.

and what proportion n of consumers to target:[42]

$$\pi(c) \equiv \max_{x,n} [\tilde{\pi}(x, n; c)], \qquad \tilde{\pi}(x, n; c) = \{p(x) - c\} nx - f(n; c) \tag{33}$$

Increased spending on marketing targets a higher proportion of 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 (both in total and per marginal consumer) is weakly higher for less productive firms, $f_c \geq 0$ and $f_{nc} \geq 0$, and is convex in the number of consumers targeted, $f_{nn} > 0$. The latter assumption is necessary for an interior solution: the second-order condition for choice of n is $f_{nn} > 0$.

Arkolakis (2010a) assumes that preferences are CES, and that the marketing cost function takes a particular parametric 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 (2010b) allows for a more general marketing cost function similar to that in (33), though retaining CES preferences.

Now consider the effects of costs on profits. Using the envelope theorem as before we obtain:

$$\pi_c = \tilde{\pi}_c = -nx - f_c < 0 \tag{34}$$

Profits are strictly decreasing in c, the inverse of firm productivity. Hence, we can invoke Proposition 1. The Arkolakis model, extended to general functional forms as here, exhibits unambiguous first-order selection effects: more efficient firms select into exporting.

We can also show that more productive firms have higher sales and profits, and also engage in more marketing. 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: $(p - tc)\,x = f_n$. Totally differentiating gives: $-txdc = f_{nn}dn + f_{nc}dc$. Collecting terms confirms that the effect of costs

[42] To reduce inessential notation, we normalize to unity the levels of market size, iceberg transport costs, and wages, and we assume that consumers are homogeneous. We also assume that n is both the proportion of consumers who purchase the good and the proportion of consumers targeted. The two are the same if preferences are CES, but may not be if preferences are such that the demand function exhibits a choke price, in which case the details of the $f(n, c)$ function are more complex.

on marketing is unambiguously negative:

$$\hat{n} = -\frac{\varepsilon - 1 + \frac{cf_{nc}}{f_n}}{\frac{nf_{nn}}{f_n}}\hat{c} \tag{35}$$

irrespective of the convexity of the demand function. With the Arkolakis specification of the marketing-cost function given above, this simplifies to $\hat{n} = -\frac{1-n}{n}\frac{\varepsilon-1}{\beta}\hat{c}$. A different special case is to assume that marketing costs are log-linear in c: $f(n;c) = \tilde{f}(n)c^{\alpha}, \alpha \geq 0$; in this case (35) simplifies to $\hat{n} = -\tilde{f}_n\frac{\varepsilon-1+\alpha}{n\tilde{f}_{nn}}\hat{c}$. Finally, higher production costs at the margin are also associated with lower sales and profits: (32) continues to hold, irrespective of the shape of the marketing-cost function.[43]

B.3 *Example 3*: Selection into Worker Screening

Our final example is the model of Helpman, Itskhoki, and Redding (2010), in which heterogeneous firms invest in screening prospective workers who have unobservable heterogeneous abilities. Post-hiring, firms and workers engage in multilateral bargaining in the manner of Stole and Zwiebel (1996) to set the firm-specific wage. This model is more complex than the others considered so far. However, we show that the approach of Proposition 1 can still be applied, and in the process we extend the model to general as opposed to CES demands. The key step in applying our approach is to show that a complex problem of maximization subject to constraints as in equation (36) below can be reexpressed as one of unconstrained maximization as in equation (45).

Except for the specification of demand, the details of the model are as in Helpman, Itskhoki, and Redding (2010). A firm must choose the number of workers it screens for their ability, denoted n, as well as the threshold ability level it will accept, denoted a. Each of these incurs direct costs: first, search costs bn in the case of workers sampled (where the search cost b depends on the tightness of the labor market and so is endogenous in general equilibrium but is taken as given by firms); and second, screening costs $\frac{c_0}{\delta}a^{\delta}$ in the case of the hiring threshold. These two variables in turn 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

[43] The expression for $\hat{\pi}$ in equation (32) holds with π interpreted as operating profits before the endogenous fixed costs $f(n;c)$ are paid.

examples.[44] Subtracting all of these costs from sales revenue $r(x) = p(x)x$ yields the ex post profit function:

$$\begin{aligned} \pi(c) \; &\equiv \; \max_{n,a} \left[\tilde{\pi}(n,a;c): \; x = c^{-1} h^{\gamma} \tfrac{k}{k-1} a, \; h = n \left(\tfrac{a}{\underline{a}} \right)^{-k} \right], \\ \tilde{\pi}(n,a;c) &= p(x)\, x - w(h,a;c)h - bn - \tfrac{c_0}{\delta} a^{\delta} - f(c) \end{aligned} \qquad (36)$$

Maximization of profits is subject to two kinds of constraints: technological constraints on production and hiring, and a constraint on the wage schedule arising from the bargaining process. We consider these in turn.

The technological constraints on production and hiring are indicated in the square brackets in (36): 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. The functional forms of these 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 $\frac{k}{k-1}a$.

The second constraint arises from the wage bargaining process. As in Acemoglu, Antràs, and Helpman (2007) and Helpman, Itskhoki, and Redding (2010), we follow Stole and Zwiebel (1996) and assume that an individual worker's ability is unobservable, and that the firm cannot write binding contracts with its workers, who are risk-neutral, face an outside option of zero, and have equal bargaining power with the firm. To incentivize its workers to stay, the firm engages 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. Since the firm's surplus equals its revenue less its wage costs, this implies a differential equation in the wage, the solution to which (to be derived below), denoted $w(h, a; c)$ in (36), is the final constraint on the firm's profit maximization.

To see the implications of the Stole-Zwiebel bargaining rule more formally, let $r(x)$ denote revenue as a function of sales; and let $R(h, a; c)$ denote revenue as a function of hires, the

[44] 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.

screening threshold, and the firm's cost, via the production function:

$$r(x) \equiv xp(x), \qquad R(h,a;c) \equiv r\left[x(h,a;c)\right] \tag{37}$$

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

$$S(h,a;c) \equiv R(h,a;c) - w(h,a;c)h \tag{38}$$

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 (38) yields a differential equation in $w(h,a;c)$:

$$2w(h,a;c) = R_h(h,a;c) - w_h(h,a;c)h \tag{39}$$

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

$$w(h,a;c) = \frac{1}{h^2}\int_0^h R_\xi(\xi,a;c)\xi d\xi \tag{40}$$

In the CES case, R is a power function of h and so (40) can be integrated directly. With general demands, integrate by parts and rearrange to obtain Proposition 3 in Stole and Zwiebel (1996):

$$S\left(h,a;c\right) = \frac{1}{h}\int_0^h R\left(\xi,a;c\right) d\xi \tag{41}$$

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 (41) cannot be expressed in closed form. However, all we need are the partial derivatives of the surplus function:

$$\begin{gathered} S_h = -\tfrac{S}{h} + \tfrac{1}{h}R\left(h,a;c\right) = w \\ S_a = \tfrac{1}{h}\int_0^h R_a\left(\xi,a;c\right)d\xi = \tfrac{1}{h}\int_0^h r'\left[x\left(\xi,a;c\right)\right]x_a\left(\xi,a;c\right)d\xi = \tfrac{1}{ha}\int_0^h r'\left[x\left(\xi,a;c\right)\right]x\left(\xi,a;c\right)d\xi = \tfrac{wh}{a\gamma} \\ S_c = \tfrac{1}{h}\int_0^h R_c\left(\xi,a;c\right)d\xi = \tfrac{1}{h}\int_0^h r'\left[x\left(\xi,a;c\right)\right]x_c\left(\xi,a;c\right)d\xi = -\tfrac{1}{hc}\int_0^h r'\left[x\left(\xi,a;c\right)\right]x\left(\xi,a;c\right)d\xi = -\tfrac{wh}{c\gamma} \end{gathered} \tag{42}$$

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 (40):

$$\int_0^h r'\left[x\left(\xi,a;c\right)\right]x\left(\xi,a;c\right)d\xi = \frac{1}{\gamma}\int_0^h r'\left[x\left(\xi,a;c\right)\right]x_\xi\left(\xi,a;c\right)\xi d\xi = \frac{1}{\gamma}\int_0^h R_\xi(\xi,a;c)\xi d\xi = \frac{wh^2}{\gamma} \tag{43}$$

Rewriting in terms of proportional changes, the total derivative of the surplus function is:

$$\hat{S} = \frac{\omega}{1-\omega}\left[\hat{h} + \frac{1}{\gamma}\left(\hat{a}-\hat{c}\right)\right] \tag{44}$$

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

We can now restate the firm's problem from (36) as an unconstrained maximization problem:

$$\underset{n,a}{Max}\left[\tilde{\pi}\left(n,a;c\right)\right], \quad \tilde{\pi}\left(n,a;c\right) = S\left[h\left(n,a\right),a;c\right] - bn - \frac{c_0}{\delta}a^\delta - f\left(c\right), \quad h\left(n,a\right) = n\left(\frac{a}{\underline{a}}\right)^{-k} \tag{45}$$

By contrast with the constrained maximization problem in (36), the effects of costs on profits can now be established by inspection. Applying the envelope theorem to (45) yields a simple expression:

$$\pi_c = \tilde{\pi}_c = S_c - f_c = -\frac{wh}{c\gamma} - f_c < 0 \tag{46}$$

Therefore, as in the previous two examples, the assumption in (6) holds: profits are strictly decreasing in costs and so ex post profits are supermodular and the model exhibits unambiguous first-order selection effects, irrespective of the form of the demand function.

We can also derive the responses of the firm's choice variables to differences in costs across firms. We begin by differentiating (45) to obtain the first-order conditions for the number of workers screened n and the threshold ability level a:

$$\begin{array}{rcl} \tilde{\pi}_n\left(n,a;c\right) = S_h h_n - b = 0 & \Rightarrow & wh = bn \\ \tilde{\pi}_a\left(n,a;c\right) = S_h h_a + S_a - c_0 a^{\delta-1} = 0 & \Rightarrow & \frac{1-\gamma k}{\gamma}wh = c_0 a^\delta \end{array} \tag{47}$$

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:

$$\hat{w} + \hat{h} = \hat{n} \quad \text{and} \quad \delta\hat{a} = \hat{w} + \hat{h} \tag{48}$$

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

$$\hat{r} = \omega\left(\hat{w} + \hat{h}\right) + (1-\omega)\hat{S} = \omega\left[\hat{w} + 2\hat{h} + \frac{1}{\gamma}(\hat{a} - \hat{c})\right] \tag{49}$$

making use of (44). The remaining equations comes from totally differentiating the revenue function and the production and hiring constraints:

$$\hat{r} = \theta\hat{x}, \quad \hat{x} = \gamma\hat{h} + \hat{a} - \hat{c}, \quad \text{and} \quad \hat{h} = \hat{n} - k\hat{a} \tag{50}$$

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 (48), (49), and (50) 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 (48) with the total derivative of the hiring constraint from (50) to solve for changes in h, a and n as functions of $\hat{w}$:

$$\hat{h} = \frac{\delta - k}{k}\hat{w}, \quad \hat{a} = \frac{1}{k}\hat{w}, \quad \text{and} \quad \hat{n} = \frac{\delta}{k}\hat{w} \tag{51}$$

Substituting for $\hat{h}$ and $\hat{a}$ into the total derivative of the production function in (50) gives:

$$\hat{x} = \frac{1 + \gamma(\delta - k)}{k}\hat{w} - \hat{c} \tag{52}$$

Substituting from this and the expression for $\hat{h}$ from (51) into the total derivative of the bargaining rule (49) yields:

$$\hat{r} = \omega\left[\hat{w} + \hat{h} + \frac{1}{\gamma}\hat{x}\right] = \omega\left[\frac{\delta}{k}\hat{w} + \frac{1}{\gamma}\hat{x}\right] \tag{53}$$

Equating this to the total derivative of the revenue function from (50) shows that sales and wages are monotonically related:

$$\hat{x} = \frac{\gamma}{\gamma\theta - \omega}\frac{\omega\delta}{k}\hat{w} \tag{54}$$

Using this we can eliminate $\hat{w}$ from (52) to get an expression for $\hat{x}$ in terms of $\hat{c}$ only, and we can write the changes in the other choice variables as monotonic functions of $\hat{x}$:

$$\hat{x} = -\Gamma^{-1}\hat{c}, \;\; \hat{r} = \theta\hat{x}, \;\; \hat{n} = \frac{\gamma\theta - \omega}{\gamma\omega}\hat{x}, \;\; \hat{w} = \frac{k}{\delta}\hat{n}, \;\; \hat{h} = \frac{\delta - k}{\delta}\hat{n}, \;\; \hat{a} = \frac{1}{\delta}\hat{n} \tag{55}$$

Here Γ is the inverse elasticity of sales with respect to costs: $\Gamma \equiv 1 - \frac{\gamma\theta-\omega}{\gamma\omega}\frac{1+\gamma(\delta-k)}{\delta}$.[45] Equation (55) shows that, if and only if Γ is positive, all variables are monotonically decreasing in c.

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: $\frac{\pi+f}{S} = \Gamma + \left(\theta - \frac{1}{\gamma}\frac{\omega}{1-\omega}\right)\frac{1+\gamma(\delta-k)}{\delta\omega}$.

Summing up, we have shown that, under suitable regularity conditions, all the firm's choice 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.

C Superconvexity

Our formal definition of superconvexity is as follows:

Definition 2. A function $p(x)$ is superconvex at a point (p_0, x_0) if and only if $\log p$ is convex in $\log x$ at (p_0, x_0).

This can be compared with log-convexity:

Definition 3. The inverse demand function $p(x)$ is log-convex at a point (p_0, x_0) if and only if $\log p$ is convex in x at (p_0, x_0). Analogously, the direct demand function $x(p)$ is log-convex at a point (p_0, x_0) if and only if $\log x$ is convex in p at (p_0, x_0).

[45] 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 (55), $\delta - k$ must be positive if the model is to exhibit an employer-size wage premium. From equations (51) and (54), $\gamma\theta - \omega$ 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.

Some implications of superconvexity are easily established:

Lemma 3. *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 1, expressed in terms of $\varepsilon \equiv -\frac{p}{xp'}$ and $\rho \equiv -\frac{xp''}{p'}$. 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 $\varepsilon > 0$. □

	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 1: Criteria for Convexity of Direct and Inverse Demands

Lemma 4. *A demand function is superconvex at a point (p_0, x_0) if and only if it is more convex than a CES demand function with the same elasticity at that point.*

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 1, 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 5. *A demand function is superconvex at a point (p_0, x_0) if and only if its elasticity is increasing in sales at (p_0, x_0).*

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

$$\varepsilon_x = -\frac{1}{x} + \frac{p(p' + xp'')}{(xp')^2} = -\frac{1}{x}(1 + \varepsilon - \varepsilon\rho) \quad (56)$$

Comparison with the final row of Table 1 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 6. *A demand function is superconvex at a point* (p_0, x_0) *if and only if its elasticity is decreasing in price at* (p_0, x_0)*.*

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 5. □

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

Lemma 7. *Provided marginal cost is strictly positive, a demand function is superconvex at a point* (p_0, x_0) *if and only if the elasticity of marginal revenue is less than the inverse of the elasticity of demand in absolute value at* (p_0, x_0)*.*

Proof. Define revenue r as $r(x) \equiv xp(x)$. Clearly, $r' = xp' + p = xp'(1 - \varepsilon)$, which is nonnegative by assumption; and $r'' = 2p' + xp'' = p'(2 - \rho)$, which must be negative from the second-order condition. Hence the elasticity of marginal revenue equals:

$$\varepsilon_{MR,x} \equiv -\frac{xr''}{r'} = \frac{2-\rho}{\varepsilon - 1} \tag{57}$$

Recalling the final row of Table 1, it follows that, when $c > 0$, so $\varepsilon > 1$, superconvexity of the demand function is equivalent to $\varepsilon_{MR,x} < \frac{1}{\varepsilon}$:

$$\rho > \frac{\varepsilon + 1}{\varepsilon} \quad \Leftrightarrow \quad 2 - \rho < \frac{\varepsilon - 1}{\varepsilon} \quad \Leftrightarrow \quad \varepsilon_{MR,x} < \frac{1}{\varepsilon} \tag{58}$$

□

When marginal cost is constant and strictly positive, the second-order condition requires that the profit function be strictly concave: $2p' + xp'' < 0 \Rightarrow \rho < 2 \Rightarrow \varepsilon_{MR,x} > 0$. Hence Lemma 7 formalizes the notion that superconvex demands come "closer" than subconvex demands to violating the second-order condition.

D Proof of Proposition 3

Differentiating the profit function $\pi(t,c) = \max_x [p(x) - tc]\,x$ gives: $\pi_t = -cx$; and

$$\pi_{tc} = -x - c\frac{dx}{dc} = -x - \frac{tc}{2p' + xp''} \tag{59}$$

where the expression for $\frac{dx}{dc}$ comes from differentiating the first-order condition, $p + xp' = c$. We want to express the right-hand side in terms of ε and ε_x. First, solve (56) 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:

$$2p' + xp'' = -\frac{p}{x\varepsilon^2}\left(\varepsilon - 1 - x\varepsilon_x\right) \tag{60}$$

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 $p - tc + xp' = 0$ to express tc in terms of p and ε: $tc = p + xp' = p - \frac{p}{\varepsilon} = \frac{\varepsilon-1}{\varepsilon}p$. (This is very familiar in the CES case.) Finally, substitute these results into (59):

$$\pi_{tc} = -x + \frac{\varepsilon - 1}{\varepsilon - 1 - x\varepsilon_x}\varepsilon x \tag{61}$$

Collecting terms gives the desired expression in (18). □

E Proof of Proposition 4

The proof of Condition 1 is in the text. Differentiating the first-order condition $p + xp' = c$ as in (59) above and reexpressing in terms of ε and ρ allows us to calculate the cost elasticity of output as: $\frac{dx}{dc} = -\frac{\varepsilon-1}{2-\rho}$. Comparing this with (57) shows that, in absolute value, this is the inverse of the output elasticity of marginal revenue, which proves Condition 2. Finally, the proof of Condition 3 follows from equation (57) in Lemma 7:

$$\varepsilon_{MR,x} \equiv -\frac{xr''}{r'} = \frac{2-\rho}{\varepsilon - 1} = 1 - \frac{\varepsilon + \rho - 3}{\varepsilon - 1} \tag{62}$$

Hence the criterion for supermodularity can be written as follows:

$$\pi_{tc} = \frac{\varepsilon + \rho - 3}{2 - \rho} x \tag{63}$$

F Figures 2 and 3 and Lemma 1

Table 2 summarizes some key properties of demand and marginal revenue functions in general, and of the demand functions in Figure 2 in particular.

	General	CES	Inverse Translog	Linear
$p(x)$	$p(x)$	$\beta x^{-1/\sigma}$	$\frac{1}{x}(\alpha + \beta \log x)$	$\alpha - \beta x$
ε	$-\frac{p}{xp'}$	σ	$\frac{\alpha+\beta \log x}{\alpha-\beta+\beta \log x}$	$\frac{\alpha-\beta x}{\beta x}$
ρ	$-\frac{xp''}{p'}$	$\frac{\sigma+1}{\sigma}$	$\frac{2(\alpha+\beta \log x)-3\beta}{\alpha-\beta+\beta \log x}$	0
$r'(x)$	$p\frac{\varepsilon-1}{\varepsilon}$	$\frac{\sigma-1}{\sigma}\beta x^{-1/\sigma}$	$\frac{\beta}{x}$	$\alpha - 2\beta x$
$r''(x)$	$p'(2-\rho)$	$-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}\beta x^{-\frac{\sigma+1}{\sigma}}$	$-\frac{\beta}{x^2}$	-2β
$\varepsilon_{MR,x}$	$-\frac{xr''}{r'} = \frac{2-\rho}{\varepsilon-1}$	$\frac{1}{\sigma}$	1	$\frac{2\beta x}{\alpha-2\beta x}$
$1 - \varepsilon_{MR,x}$	$\frac{\varepsilon+\rho-3}{\varepsilon-1}$	$\frac{\sigma-1}{\sigma}$	0	$\frac{\alpha-4\beta x}{\alpha-2\beta x}$

Table 2: Properties of Some Demand and Marginal Revenue Functions

For the inverse translog demand function and Lemma 1, sufficiency follows from the derivation of the elasticity of the marginal revenue curve, $\varepsilon_{MR,x}$, in the final column of Table 2. Necessity follows by writing the condition $\varepsilon + \rho = 3$ in terms of the derivatives of the demand function and integrating.

As for the linear demand function, maximizing operating profits, $\pi = (p - tc)x$, yields the first-order condition, $\alpha - 2\beta x = tc$, which can be solved for optimal output: $x = \frac{1}{2\beta}(\alpha - tc)$. Substituting back into the expression for profits gives the maximized operating profit function:

$$\pi(t, c) = \beta x^2 = \frac{1}{4\beta}(\alpha - tc)^2 \tag{64}$$

Hence the second cross-derivative is:

$$\pi_{tc} = -x + \frac{tc}{2\beta} = -\frac{1}{2\beta}(\alpha - 2tc) \tag{65}$$

This is clearly positive, so π is supermodular in $\{t, c\}$, for $c \geq \frac{\alpha}{2t}$, equivalent to $x \leq \frac{\alpha}{4\beta}$; while it is negative for $c \leq \frac{\alpha}{2t}$, equivalent to $x \geq \frac{\alpha}{4\beta}$. Note that $\frac{\alpha}{4\beta}$ is half the maximum level of output (in the absence of fixed costs of exporting), $\frac{\alpha}{2\beta}$. The condition for supermodularity of π when we make the finite comparison between exporting and FDI is even more stringent. This requires that the slope of Π^F in Figure 3 exceed that of Π^E, i.e., that $\pi_c(c,1) \geq \pi_c(c,t)$. This implies that $c \geq \frac{\alpha}{t+1}$, which is strictly greater than $\frac{\alpha}{2t}$ for $t > 1$. At the threshold marginal cost, $c = \frac{\alpha}{t+1}$, output equals $\frac{1}{t+1}\frac{\alpha}{2\beta}$ in the case of exports and $\frac{t}{t+1}\frac{\alpha}{2\beta}$ in the case of FDI.

The case illustrated in Figure 3 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: $c < \frac{1}{t}\left(\alpha - 2\sqrt{\beta f_X}\right)$; (ii) FDI must be profitable, $\Pi^F \equiv \pi(0,c) - f_F > 0$, which requires: $c < \alpha - 2\sqrt{\beta f_F}$; 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: $(t-1)\alpha^2 > 4(t+1)\beta\,(f_F - f_X)$. Note that, in general (though not in Figure 3), 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 α 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.

G Proof of Proposition 6

By the envelope theorem, the derivative of maximum profits with respect to the tariff equals minus the level of output: $\pi_t = -x\,(t,c)$. Hence it follows that: $\pi_{tc} = -x_c$. So, to check for supermodularity, we need only establish the sign of the derivative of output with respect to the cost parameter c.

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

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

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

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

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

Solving for the effect of the cost parameter on output and substituting into π_{tc} gives:

$$\pi_{tc} = -x_c = \underbrace{D^{-1}}_{+} \left[\underbrace{C_c(xC_{kk} + F'')}_{+} - \underbrace{xC_k}_{-} C_{kc} \right] \tag{69}$$

The second-order conditions imply that the determinant $D = -(2p' + xp'')(xC_{kk} + F'') - C_k^2$ and the first term inside the brackets are positive, as indicated, which gives a calculus proof of Proposition 5: a positive value of C_{kc} is sufficient but not necessary for a positive value of π_{tc}. Specializing equation (69) to the investment cost functions in (26) gives:

$$\pi_{tc} = D^{-1} cx \left[\phi\phi'' - (\phi')^2 \right] \tag{70}$$

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 (70) is equivalent to ϕ being log-convex, which proves Proposition 6. □

H Selection Effects in Oligopoly

In the text of the paper, as in almost all the recent literature on trade with heterogeneous firms, markets are assumed to be 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. 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 4 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 3. 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 3: 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, there is even less presumption than in earlier cases that a departure from supermodularity may lead to a reversal of the conventional sorting. This is because supermodularity of the profit function conditional on rivals' responses is only a local condition, and is relevant to the conventional sorting only at those points which are relevant to a particular Nash equilibrium. Thus, supermodularity might not hold over a range of the profit function; but if that range was never relevant given rivals' responses, then the conventional sorting would still apply.

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

——— (2010b): "Market Penetration Costs and the New Consumers Margin in International Trade: Appendix," web appendix, Yale University.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2012): "The Elusive Pro-Competitive Effects of Trade," working paper, MIT, revised July 2014.

Arrow, K. J., and F. Hahn (1971): *General Competitive Analysis*. San Francisco: Holden-Day.

Bache, P. A., and A. Laugesen (2015): "Monotone Comparative Statics for the Industry Composition," Working Paper, Aarhus University.

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.

Behrens, K., and Y. Murata (2007): "General Equilibrium Models of Monopolistic Competition: A New Approach," *Journal of Economic Theory*, 136(1), 776–787.

Bernard, A. B., E. J. Blanchard, I. Van Beveren, and H. Y. Vandenbussche (2016): "Carry-Along Trade," Working Paper, Tuck School of Business at Dartmouth.

Bernard, A. B., and J. B. Jensen (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 52(1), 1–25.

Bertoletti, P. (2006): "Logarithmic Quasi-Homothetic Preferences," *Economics Letters*, 90(3), 433–439.

Bertoletti, P., and P. Epifani (2014): "Monopolistic Competition: CES Redux?," *Journal of International Economics*, 93(2), 227–238.

Bertoletti, P., and F. Etro (2017): "Monopolistic Competition when Income Matters," *Economic Journal (forthcoming).*

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.

Clerides, S. K., S. Lach, and J. R. Tybout (1998): "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico and Morocco," *Quarterly Journal of Economics*, 113(3), 903–947.

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.

Diewert, W. E. (1976): "Exact and Superlative Index Numbers," *Journal of Econometrics*, 4(2), 115–145.

Dixit, A. K., and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297–308.

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

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 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.

Hallak, J. C., and J. Sivadasan (2013): “Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia,” *Journal of International Economics*, 91(1), 53–67.

Harrigan, J., and A. Reshef (2015): “Skill Biased Heterogeneous Firms, Trade Liberalization, and the Skill Premium,” *Canadian Journal of Economics*, 48(3).

Head, K., and J. Ries (2003): “Heterogeneity and the FDI versus export decision of Japanese manufacturers,” *Journal of the Japanese and International Economies*, 17, 448–467.

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.

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.

Lileeva, A., and D. Trefler (2010): "Improved Access to Foreign Markets Raises Plant-level Productivity ... For Some Plants," *Quarterly Journal of Economics*, 125(3), 1051–1099.

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*, 62(1), 157–180.

Mrázová, M., and J. P. Neary (2013): "Not so Demanding: Preference Structure, Firm Behavior, and Welfare," Economics Discussion Paper No. 691, University of Oxford.

Mukherjee, A. (2010): "A Note on Firm Productivity and Foreign Direct Investment," *Economics Bulletin*, 30(3), 2107–2111.

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.

Neary, J. P., and K. W. Roberts (1980): "The Theory of Household Behaviour under Rationing," *European Economic Review*, 13(1), 25–42.

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.

Pollak, R. A. (1971): "Additive Utility Functions and Linear Engel Curves," *Review of Economic Studies*, 38(4), 401–414.

——— (1972): "Generalized Separability," *Econometrica*, 40(3), pp. 431–453.

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

Spearot, A. C. (2012): "Firm Heterogeneity, New Investment and Acquisitions," *Journal of Industrial Economics*, 60(1), 1–45.

——— (2013): "Market Access, Investment, and Heterogeneous Firms," *International Economic Review*, 54(2), 601–627.

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.

Topkis, D. M. (1978): "Minimizing a Submodular Function on a Lattice," *Operations Research*, 26(2), 305–321.

Vives, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19(3), 305–321.

——— (1999): *Oligopoly Pricing: Old Ideas and New Tools*. Cambridge, Mass.: MIT Press.

Yeaple, S. R. (2009): "Firm Heterogeneity and the Structure of US Multinational Activity," *Journal of International Economics*, 78(2), 206–215.

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