ECONOMIES OF SCALE AND THE SIZE OF EXPORTERS

Roc Armenter
Federal Reserve Bank of Philadelphia

Miklós Koren
Central European University

Abstract
How important are economies of scale in exporting? We argue that firm size cannot be the main determinant of export status if a model is to be consistent with the observed number and size of exporters. Instead, we need a lot of variation independent of firm size to reconcile the model with the data. We show that the augmented model also has markedly different implications regarding the margin of adjustment in the event of a trade liberalization: most of the adjustment is through the intensive margin and productivity gains due to reallocation are halved. (JEL: F12)

1. Introduction

Exporters represent less than 20% of all US manufacturing firms and are larger than non-exporters—total sales per firm are about 4–5 times higher for exporters.[1] These stylized facts are often cited as support for models with economies of scale and firm heterogeneity. In Melitz (2003), only larger, more productive firms generate enough net revenues abroad to cover the fixed costs associated with exporting. A key implication is that, in the event of a trade liberalization, resources will be reallocated toward the more productive firms, raising the average productivity in the industry.

In this paper we evaluate the *quantitative* importance of economies of scale for understanding the export behavior of firms. We start by exploring the stark hypothesis that economies of scale are the only determinant of the export status of the firm.

---

1. Both facts are documented in Bernard et al. (2007) for the universe of US manufacturing firms operating in 2002.

Firms differ only in productivity, as in Melitz (2003). We show how to derive the model's predictions just by focusing on a single equilibrium condition, the export entry condition. The economies of scale imply that we should observe exporters and non-exporters to be strictly sorted by sales; that is, the smallest exporter should be larger than the largest non-exporter. The set of exporters is thus easily characterized with a cutoff rule in terms of total sales. We use the cutoff property to derive the implications for exporter size given the share of exporters and the distribution of total sales, without having to specify the full model.[2] Given that exporters are roughly one-fifth of all firms, strict sorting implies that exporters should be concentrated in the top quintile of the firm size distribution. We find that the firms below the top quintile are quite small (an average of $740,000 total sales), while the firms in the top quintile are much larger ($70 million on average)—reflecting the huge skewness in the firm size distribution. Hence, the model greatly overpredicts the exporter size premium: exporters should have between 90 and 100 times more total sales than non-exporters, instead of just the ratio of 4 to 5 observed in the data.[3]

We proceed to reconcile the economies-of-scale model with the data by introducing additional firm-level heterogeneity. We take a latent variable approach so there is no need to specify which are the additional sources of variation in the export decision—only that these factors are independent of firm size. We find that a lot of size-independent variation is needed to match the observed frequency and size of exporters. In other words, total sales are a poor predictor of the export status of a firm.[4]

We then explore whether introducing the necessary size-independent variation alters the model's predictions in the event of a trade liberalization. We set up a partial equilibrium model based on Melitz (2003). We then explore two versions of the model. In the first version—which we call the strict sorting model—productivity is the only source of variation across firms. The model displays the strict sorting property and thus exporters are counter-factually very large. In the second version we assume that firms face heterogeneous fixed costs—replicating the latent heterogeneity model. We calibrate the distribution of fixed costs to match the share and size of exporters. We then compare the models' predictions in response to a fall in trade costs.

We find that, in the aggregate, both models are indistinguishable from a representative-firm model. Free entry amplifies the response of exports to a trade liberalization through the love-of-variety effect. However, firm-level heterogeneity virtually cancels all the amplification introduced by free entry. This holds for both the strict-sorting and the latent-heterogeneity models, suggesting that the introduction

---

2.   We use the distribution of total sales for manufacturing firms (NAICS code 31–33) as given by the 2002 Statistics of US Businesses of the Census.

3.   Of course, we did not expect the strict sorting of exporters to hold *exactly* in the data. Melitz (2003) certainly does not intend to preclude the importance of other idiosyncratic factors in the export decision.

4.   We also note that the introduction of the latent heterogeneity helps explain the large fraction of exporters with little foreign sales, even if the foreign to domestic sales ratio is taken to be constant across exporters.

of heterogeneity simply does not affect the response of aggregate trade and export prices.[5]

Behind the similarities in the aggregate there are large differences in the margins of adjustment. Critically, the latent heterogeneity model has only a minor role for the extensive margin. First, there is much less entry: we find that the growth rate of the number of exporters is more than 60% in the strict sorting version, but only 15% in the latent heterogeneity version. Second, new exporters are about one third of the size of existing exporters in the latent heterogeneity model. This stands in marked contrast with the standard Melitz model in which new exporters are 10 to 12 times smaller than the average exporter prior to trade liberalization. We decompose total export growth in an intensive and extensive margin and find that the strict-sorting model the extensive margin accounts for more than 60% of the trade growth. In contrast, the extensive margin accounts for less than 20% once we match the size of exporters.

Finally, we look at the aggregate productivity gains due to the reallocation effect across both models. We find that the productivity gains due to trade liberalization are halved in the augmented model. In short, the large amount of size-independent variation "waters down" the core mechanism of reallocation from non-exporters to exporters.[6]

We are not the first to introduce additional heterogeneity in a trade model with economies of scale in order to fit firm- or plant-level data. Two well-known contributions are Das, Roberts, and Tybout (2007), who estimate a model with fixed and sunk costs using Colombian plant data; and Eaton, Kortum, and Kramarz (2011), who estimate a model of firm heterogeneity and export participation using data from France. Both papers aim for a near-complete fit of the data and correspondingly require rich and complex models with many sources of heterogeneity—correlated firm-market demand shocks and fixed costs, for example. Their success is a testament to how far trade models have progressed in quantitative analysis, but does not offer a clear guidance regarding which model features are indispensable to explore the implications of a trade liberalization.

We view our paper as a complement to their approach, providing a transparent evaluation of the key mechanism in the Melitz model. First, we focus on only two empirical moments, namely, the frequency of exporters and their size premium relative to non-exporters, to be combined with the distribution of total sales.[7] Second, we impose as little structure as possible when introducing the necessary heterogeneity to

---

5.   Atkeson and Burstein (2010) find that the firms' responses to a trade liberalization do not quantitatively impact the implications for aggregate productivity and welfare in a general equilibrium model. We must emphasize that ours is a partial equilibrium model, so the offsetting effects are not due to factor prices or aggregate demand adjusting in general equilibrium. Arkolakis, Costinot, and Rodriguez-Clare (2012) document conditions such that heterogeneity is irrelevant for the welfare gains from a trade liberalization.

6.   However, it must be noted that we miss on the productivity gains due to exit in the domestic market, since ours is partial equilibrium model and the wage rate—or more generally, unit costs—is taken as given.

7.   To the best of our knowledge, we are the first to highlight the key role played by these two facts. Eaton, Kortum, and Kramarz (2011), for example, exclude non-exporters from the model estimation.

match the data. Our results also highlight the implications of the dispersion in fixed costs, while the literature has focused on the average fixed cost.

Some researchers have also extended the Melitz model seeking to explain some particular facts of interest. Johnson (2012) and Kugler and Verhoogen (2012) introduce firm-level variation in output quality in order to explain several facts regarding unit values and firm size, respectively. In a similar setting, Hallak and Sivadasan (2011) talk of differences in "caliber" among firms and note that such differences are necessary to break the monotonic relationship between size and export status. Nocke and Yeaple (2012) explore firm heterogeneity in managerial resources and talent: firms with high labor productivity export, but not all exporters are large and not all large firms export. Halpern, Koren, and Szeidl (2011) structurally estimate a firm-level model of *imports*, using our method to calibrate the dispersion in the fixed costs of importing. Dynamic models naturally generate heterogeneity across firms through past investment decisions and exporter hysteresis due to sunk costs: see for example Ruhl (2008), Aw, Roberts, and Xu (2011), and Alessandria and Choi (2012).

The empirical literature has not come to a consensus on the quantitative importance of the extensive margin for aggregate trade patterns. Two papers are well known for arguing that the extensive margin is an important dimension of aggregate data. Hummels and Klenow (2005) find that the extensive margin accounts for 60% of the cross-country differences in trade. Broda and Weinstein (2006) find large welfare gains associated with the expansion in import variety for the United States over the last three decades. Both papers build upon the analysis in Feenstra (1994) and share a focus on the long run.

Recently, several papers have taken the position that the extensive margin contributes little to aggregate trade patterns. Besedes and Prusa (2011) argue that new exporting relationships have little impact on long-run export growth because they tend to be very short-lived. Arkolakis et al. (2008) document a sizable increase in variety in Costa Rica from 1986 to 1992. They argue, though, that the increase did not translate into large welfare gains because new varieties were imported in small quantities. Armenter and Koren (2014) show that several well-known facts about the extensive margin are the result of the sparsity of the data.

2. On the Size of Exporters

We start by deriving the quantitative predictions when economies of scale are the *only* determinant of the export status of a firm. As in Melitz (2003), firm i will export only if its foreign sales, net of the associated variable costs, would cover the fixed costs associated with exporting,

$$r_i^* - c_i^* \geq f. \tag{1}$$

Firms differ in their labor productivity or in the quality of goods produced. Either way, more efficient firms can generate more net income abroad and are thus more likely to be exporters. Hence, the model predicts that exporters are more productive

and/or sell higher quality goods than non-exporters. Incidentally, more efficient firms also sell more in the domestic market, so exporters are unambiguously larger than non-exporters in terms of total sales. These qualitative predictions are borne out in the data and often cited as support for economies of scale in exporting.

We seek a *quantitative* evaluation of economies of scale and how they shape the set of exporters. We show here how to do so only on the basis of export entry condition (1), without having to specify or calibrate the rest of the model. The first step is to rewrite the entry condition (1) in terms of total sales. In Melitz (2003), firms compare their potential export sales with the fixed cost, and decide whether to export or not. Of course, we only observe foreign sales for firms that actually export. Fortunately, there is a tight relationship between a firm's total sales and its underlying efficiency parameter in the model: a firm with higher productivity will always have more total sales in equilibrium. Since net income abroad is also strictly increasing in productivity, we have an increasing, monotonic relationship between total sales and net income abroad—even if the latter is counterfactual because the firm does not export. We can then use data on total sales, which are easily accessible and observed independently of whether the firm exports.

We thus summarize the set of exporters with a simple threshold rule in place of equation (1): firm i will export only if its total sales r_i are above some level t,

$$r_i \geq t. \tag{2}$$

It is immediate that exporters and non-exporters are strictly sorted by size; that is, the smallest exporting firm has more total sales than the largest non-exporting firm. In a fully developed structural model, the value of the threshold level t is determined in equilibrium and is bound to depend on the parameters as well as the choice of several specifications. We show in what follows how we can derive the quantitative predictions straight from the data on the basis of (2) alone: what we call the *strict-sorting* model.

The threshold condition (2) implies that the share of exporters is equal to the fraction of firms with total sales equal or larger than t. We have thus that $s_x = 1 - \Psi_r(t)$, where Ψ_r is the empirical cumulative distribution function of total sales, and s_x is the fraction of firms with positive export sales. In other words, we can solve for the $(1 - s_x)$th percentile in the distribution of total sales and obtain the threshold t consistent with the observed share of exporters. We can then easily compute the truncated mean,

$$E\{r_i | r_i \geq t\} = \int_t^\infty r_i \frac{d\Psi_r(r_i)}{s_x}.$$

The truncated mean is the strict-sorting model prediction for the average total sales for exporters.

Bernard et al. (2007) report that only 18% of the US manufacturing firms had positive sales abroad in 2002.[8] Exporters, thus, are relatively rare. For the distribution

---

8. Exporters are similarly scarce if we look at plants or establishments rather than firms—see Bernard et al. (2003), for example. The scarcity of exporters has also been confirmed in a variety of countries.

TABLE 1. The distribution of firm sales in manufacturing—*Census*.

Size bin	Frequency	Cumulative frequency	Average sales
0–$100,000	0.145	0.145	$55,600
$100,000–$500,000	0.305	0.450	$257,000
$500,000–$1 million	0.144	0.594	$718,000
$1–5 million	0.257	0.851	$2.26 million
$5–10 million	0.060	0.911	$6.84 million
$10–50 million	0.063	0.974	$19.3 million
$50–100 million	0.010	0.984	$56.4 million
over $100 million	0.015	1.000	$670 million
All Firms			$13.2 million

of total sales we look at the 2002 Statistics of US Businesses published by the Census. Table 1 summarizes the data for manufacturing firms. As is well known, there is an enormous amount of skewness in the firm size distribution. The average firm sells $13.2 million, and yet 45% of the firms sell less than $500, 000. In short, there are many, many small firms and a few very, very large ones.

From Table 1 we see that the 82nd percentile falls somewhere between $1 and $5 million sales, definitively closer to the latter. This already suggests that non-exporters are expected to be quite small under strict sorting: firms below $1 million sales average less than $320, 000 in sales and would constitute over 70% of all non-exporters. At the same time, exporters will be quite large. Firms above $1 million average $82 million in total sales and would represent over 80% of all exporters. We want to be more precise than this, though. To do so we only need to specify the distribution of firms within the range of $1–$5 million total sales.[9] We assume firms in this particular bin follow a two-side truncated Pareto distribution, parameterized to match the average total sales in the range ($2.26 million).

We find that strict sorting by size implies that exporters should sell, on average, between 90 and 100 times more than non-exporters. The exporter size premium remains very large for whatever distribution one assumes for the firms in the $1–$5 million range. As a check we computed the *lower* bound on the exporter's size implied by Table 1 by taking all firms with sales between $1 million and $5 million to be identical, with total sales equal to the bin's average ($2.26 million). Even in this case we find that exporters are predicted to be more than 80 times larger than non-exporters.

How does the implied exporter size premium compare with the data? Bernard et al. (2007) report that exporters are four to five times larger than non-exporters.[10]

---

9.   There is no need to assume anything about the distribution of firms outside this size bin as they are all squarely in the exporter or non-exporter category.

10.   Table 3 in Bernard et al. (2007) states that the difference in average log shipments between exporters and non-exporters is 1.48 for the same set of firms we used. The finding that exporters are larger than non-exporters has also been confirmed for plants and establishments, as well as for other countries. The size differences are all in the range between 4 and 6.

TABLE 2. Exporters and non-exporters in the model and the data.

	Data	Strict sorting
Average sales—non-exporters	$8.1 million	$740,000
Average sales—exporters	$36.4 million	$70.1 million
Exporter size premium	4.5	95

Strict sorting thus greatly overpredicts the size differences between exporters and non-exporters. Table 2 compares the predicted size of exporters and non-exporters with the data for US manufacturing firms in 2002. Under the hypothesis of strict sorting, exporters should have, on average, $70.1 million in total sales—double what we actually observe. In the model non-exporters are expected to be very small, just $740,000 in total sales. In the data, though, there are non-exporters that are large enough to bring their average total sales above $8 million. Clearly, larger firms are indeed more likely to export in the data, but only a little bit more. In the model, though, the conditional probability of exporting rises very sharply, taking value zero if r_i is below the cutoff t, and one otherwise.[11]

It is possible that some exporters were mistakenly recorded as non-exporters in the data. We redo the previous exercise assuming that the actual share of exporters is $s_x = 0.25$—which implies than close to 30% of non-exporters were actually misclassified. We still find that strict sorting implies an exporter size premium of almost 90.[12] As we increase the share of exporters further, the predicted size premium declines but only slowly, since non-exporters are predicted to be even smaller than before. For example, say we take exporters to be twice as frequent as recorded in the data, $s_x = 0.35$. The average total sales of exporters is then very close to the data. Non-exporters, though, are then very small (less than $500,000) and the implied size premium is about 75.

Is our result unique to the United States? We redo the exercise using data for French firms as reported in Eaton, Kortum, and Kramarz (2010).[13] We find that the implied size premium is above 54, and possibly around 60—well above its empirical counterpart.

Due to its ubiquitous nature in the literature, we have also experimented with a parameterized Pareto distribution for the total sales distribution. The Pareto distribution is not a good approximation of the *overall* firm size distribution—only its upper tail. We can thus view this exercise as ignoring the data regarding smaller firms which, after all, add to a small fraction of the overall economy. To pin down the exporter size

---

11. The strict sorting model *maximizes* the correlation between exporting and firm size: the predicted joint distribution of sales and export status is actually the Fréchet maximal distribution given the respective marginal distributions.

12. In the Online Appendix we show that the gap between data and model remains very large across a range of specifications and values for measurement error.

13. See Table 5. The data include around 60% of manufacturing firms in 1986. Only 14% of them export.

premium, we only need the slope parameter for the Pareto distribution that fits the data best. Axtell (2001) estimates a Pareto parameter of 1.024, which implies an exporter size premium of 110 under strict sorting.[14] On the upper end of the estimates, a slope parameter $k = 1.065$ brings the exporter size premium down to 41. Note the latter number is actually half the *lower bound* on the size premium implied by the actual distribution, clearly indicating that the Pareto distribution is ignoring the left tail of the firm size distribution. In order to match the share and size of exporters, we would require to set $k = 1.65$, well above any estimate.

3. Introducing Latent Heterogeneity

Of course, we did not expect the strict sorting of exporters to hold *exactly* in the data. Melitz (2003) certainly does not intend to preclude the importance of other idiosyncratic factors, unrelated to size, in the firm's decision to export. In this section we introduce latent heterogeneity, unrelated to size and thus to the economies of scale, in the decision to export. We then evaluate the role of the economies of scale against the latent heterogeneity.

3.1. Fitting the Data

We proceed to fit the data by introducing the necessary firm-level heterogeneity in the export decision. We rewrite the threshold condition (2) in terms of a latent variable, t_i. A firm i exports if its total sales satisfy

$$r_i \geq t_i. \tag{3}$$

The latent variable is identically and independently distributed across firms, with a c.d.f. Ψ_t over support $\Re_+$. Thus the realization t_i is independent of the size of the firm i.

With the latent variable condition (3) we can capture all of the underlying heterogeneity without having to specify the sources of variation. Indeed, the only structure imposed on the latent variable is that it is independent of firm size.

It is now useful to fit the empirical distribution of total sales with a parametric distribution. We use a lognormal distribution with mean $\mu_r = 6.3$ and standard deviation $\sigma_r = 2.6$—so we reproduce the average total sales (in thousands of dollars) as well as the approximated location of the 82th percentile.

We also assume that the latent variable t_i follows a lognormal distribution. This is mainly a choice of convenience: we want a flexible two-parameter distribution defined over positive numbers. We pick the mean μ_t and standard deviation σ_t such that the

---

14.   Estimates in Luttmer (2007) and elsewhere are also very close to one. See Online Appendix for a derivation of the size premium under a Pareto distribution.

model reproduces the share and size of exporters. That is, we solve for μ_t and σ_t such that the equations

$$s_x = \int \Psi_t(r) \, d\Psi_r(r), \tag{4}$$

$$E\{\log(r_i)|r_i \geq t_i\} = \int \log(r) \, \Psi_t(r) \, d\Psi_r(r)/s_x, \tag{5}$$

where Ψ_t denotes the c.d.f. of the latent variable distribution, reproduce the observed values for the share of exporters, $s_x = 0.18$, and the average log total sales of exporters, $E\{\log(r_i) \,|\, r_i \geq t_i\} = 5.66$.[15]

The latent variable is very dispersed: we find that the parameter values matching moments (4)–(5) are $\mu_t = 13.73$ and $\sigma_t = 7.7$. These parameters imply that the coefficient of variation of the latent variable is many orders of magnitude larger than for total sales! We explored an array of parameters for the distribution of total sales and found always that we need a huge dispersion for the latent variable in order to reproduce the share and size of exporters. More precisely, we consider values μ_r in the range 5.5–7.5 and σ_r in the range 2–3. The resulting parameters for μ_t and σ_t were always above 10 and 5, respectively.

The results are perhaps not that surprising: after all, we have to make up for a large gap between the strict-sorting model and the data in the size premium of exporters. In order to reduce the size of exporters we need the latent variable to take very large values with high probability, so some large firms do not export. Simultaneously, some other firms must draw a low realization of the latent variable and export independently of their size.[16]

Moreover, there are also good reasons to view our findings as a lower bound on the dispersion of the latent variable. Firms often carry multiple product lines, and there are many foreign markets to serve. However, it only takes one product to be exported to one destination for a firm to be called an exporter. Hence, if we think that each market/product offers an independent opportunity to export, we have to see the latent variable t_i as the *minimum* realization among the ensemble of product- and destination-specific thresholds. Thus, the underlying distribution from which the product- and destination-specific latent variables are drawn would feature much more dispersion.

---

15.   Here we use the log total sales instead of total sales, so the exporter size premium is now given by the difference in average log total sales between exporters and non-exporters. The change of units has virtually no implication for the parameters of the latent variable distribution—but it turns out to be very convenient for the calibration of the model later.

16.   It is also necessary to introduce latent heterogeneity in order to match *dynamic* facts. Entry and exit rates in foreign markets are relatively high, as documented in Bernard, Jensen, and Schott (2007). The volatility of firm employment in the data is clearly too low to explain these high rates in a basic Melitz model: firms rarely grow or contract enough to start and stop exporting so frequently. See Atkeson and Burstein (2010) for a discussion.

TABLE 3. Conditional probability of exporting.

	Percentile		
Ordered	25th	50th	75th
By total sales	0.1179	0.1674	0.2286
By latent variable	10^{-7}	0.0016	0.1881

3.2. The Role of Firm Size

The huge dispersion of the latent variable indicates that firm size plays only a small role in the determination of the export status of a firm. We can illustrate this point with a simple exercise. Note that given a firm size r_i, the probability that firm i exports is $\Pr(r_i \geq t_i | r_i)$ or simply $\Psi_t(r_i)$. Given our estimates, a firm of median size exports with probability 0.167, very close to the unconditional probability of 0.18.[17] In other words, a firm of median size could be taken as representative of the industry as a whole. In contrast, a firm with the median latent variable will export only with probability 0.0016. Table 3 repeats the exercise with the 25th and 75th percentiles for total sales and the latent variable.

Next we set to measure the explanatory power of a firm's total sales on its export status. Following Eaton, Kortum, and Kramarz (2011), we compute a pseudo-R^2 as

$$R = 1 - \frac{E[V^C(r)]}{V^U}$$

where $V^C(r)$ is the variance of an exporter dummy, conditional on a sales level r; and V^U is the unconditional variance of the exporter dummy.[18] The unconditional variance is pinned down by the share of exporters, $V^U = s_x(1 - s_x)$, since the exporter dummy follows a Bernoulli distribution with success probability s_x. The conditional variance is again easily computed using our latent-variable distribution, as the conditional export probability is $\Psi_t(r)$ and thus $V^C(r) = \Psi_t(r)(1 - \Psi_t(r))$. We then integrate across sales levels, $E[V^C(r)] = \int V^C(r) d\Psi_r(r)$.

We find that firm size explains only 5% of the overall variation in export participation. This result contrasts with Eaton, Kortum, and Kramarz (2011), who report that over half of the variation in market access is attributed to the firm's productivity. However, the two results cannot be directly compared for several reasons. In particular, Eaton et al. analyze "market entry" for each of 112 possible foreign countries, and thus their result is mainly driven by the number of countries exporters

---

17.   The median size is given by $\exp(\mu_r)$ or approximately \$600,000.

18.   See Section 4.7.1 in Eaton et al. (2011) for further discussion.

TABLE 4. Exporter size premium in terms of domestic sales.

	Data	Model
Average domestic sales—non-exporters	$8.1 million	$740,000
Average domestic sales—exporters	$31.3 million	$60.3 million
Exporter size premium	3.8	81.5

sell to.[19] Our calculation is instead for export participation. As a firm is classified as an exporter as long as it sells to one foreign country, it only takes a single low realization of the fixed cost (or a high one of the demand shock) among the 112 independent draws for a firm to export. Thus the explanatory power of the common component may be substantial for a typical, single foreign market yet small for the overall export status of the firm.

3.3. Small Exporters

Finally, we note that matching the size premium of exporters helps explain why a large fraction of exporters actually sell very little abroad.[20] Small exporters are clearly at odds with economies of scale, as it is hard to reconcile exporters with very little foreign revenues with the overall low number of exporters.

Arkolakis (2010) develops a theory of market penetration in which export intensity is increasing with size. Our first observation is that, as long as the strict property is preserved, variation in export intensity cannot reconcile the model with the data. The simplest way to see this is to compute the exporter size premium in terms of domestic sales under strict sorting, following the same procedure we did with total sales. We report the results in Table 4. Excluding foreign sales does not reduce the disparity between the data and model. This is perhaps not surprising given that foreign sales are a small fraction of total sales for the United States.

Once we introduce latent heterogeneity small exporters arise naturally without variation in export intensity. A small firm may export despite its low productivity if it draws a small fixed cost. Such a firm will export little, since it will not be competitive abroad. Similarly some very productive firms do not export despite their large potential foreign sales because they draw a very high fixed cost.

---

19.   Eaton et al. (2011) assume that each market–firm pair has an independent realization of a fixed cost and a demand shock. The latter also disconnects domestic sales from productivity, which is another possible reason why our results are so different.

20.   See Bernard, Jensen, and Schott (2007), and Eaton, Kortum, and Kramarz (2004). Bernard, Jensen, and Schott (2007) report that more than 75% of all exporters sell less than $1 million abroad, with plenty of firms selling a very small amount abroad—for example, less than $20,000.

4. A Simple Framework of Exports and Exporters

In this section we set up a model with economies of scale and firm heterogeneity.[21] The model is simpler than that of Melitz (2003) in that it is a partial-equilibrium model, taking the wage rate as given. As a result the model abstracts from entry and exit in the domestic market.

    We explore the implications of the additional heterogeneity for an episode of trade liberalization. To this end we consider two versions of the model. In the first, there is strict sorting, so the model does not match the size of exporters. In the second version, we introduce the needed variation as a random fixed cost. The rest of the parameters are common across models.

4.1. Framework

There is a set Ω of firms that produce and sell in the home country. Firms are heterogeneous in their productivity, denoted φ, and the fixed costs they face if they start exporting, denoted f. Productivity and fixed costs are independently distributed over $\Re_+$ with c.d.f. G and H, respectively. We summarize a firm by its type $\omega = \{\varphi, f\}$. The set of firms Ω (and their distribution) is taken as a given.

    Each firm is the single producer of a differentiated good, operating the production function $y(\omega) = \varphi(\omega) l(\omega)$, where $l(\omega)$ is the labor demanded by firm ω. Consumers combine the differentiated goods according to

$$Y^d = \left[\int_\Omega (y^d(\omega))^\rho d\omega \right]^{1/\rho},$$

where $\rho \in (0, 1)$ and $y^d(\omega)$ denote the output of firm ω sold in the home country.

    Firms are monopolistic competitors and set the price

$$p^d(\omega) = \frac{1}{\rho} \frac{w}{\varphi(\omega)},$$

where w is the wage rate and $\theta = (1 - \rho)^{-1}$ is the price elasticity. We take the wage rate as exogenously given, so ours is a partial equilibrium model.

    It is clear that only the productivity parameter φ will determine domestic sales. We can thus write $p^d(\varphi)$ and $y^d(\varphi)$. The c.e.s. demand structure also implies that firm φ revenues from domestic sales can be expressed as

$$r^d(\varphi) = \left(\frac{\varphi}{\bar{\varphi}} \right)^{\theta-1} R^d, \tag{6}$$

---

21.   The Online Appendix contains a detailed description of the model.

where R^d are total sales revenues in the domestic market, and

$$\tilde{\varphi} = \left[\int_0^\infty \varphi^{\theta-1} dG(\varphi)\right]^{\frac{1}{\theta-1}}$$

is the average productivity defined as in Melitz (2003). Equation (6) delivers the key relationship between productivity and domestic sales.

We now move to the determination of exports and exporters. Not all firms export: let Ω_x denote the set of firms that do and M_x its measure. We normalize the measure of all firms to one, so M_x is also the share of exporters. Consumers in the foreign country combine the subset of exported goods according to a CES aggregate, with identical parameter ρ. The export price index is

$$P^f = \left[\int_{\Omega_x} (p^f(\omega))^{1-\theta} d\omega\right]^{\frac{1}{1-\theta}}. \tag{7}$$

To close the model, we require an aggregate demand for exports, given by

$$Y^f = Y^* (P^f)^{-\nu}, \tag{8}$$

where Y^* is the (exogenously given) income of the foreign country, and ν is the price elasticity of aggregate exports of the home country. We assume that $\nu < \theta$; that is, exports of the home country are closer substitutes of each other than they are of a good produced elsewhere.

Let us first solve for the export revenues of a firm, taking as given the set of exporters Ω_x. Profit maximization implies that

$$p^f(\omega) = \frac{1}{\rho} \frac{\tau w}{\varphi(\omega)},$$

where $\tau > 1$ is an iceberg trade cost associated with exports. It is clear again that, conditional on exporting, only the productivity parameter φ determines sales. The demand system allows us to write a firm's export revenues as a function of the average export revenues within exporters,

$$r^f(\varphi) = \left(\frac{\varphi}{\tilde{\varphi}_x}\right)^{\theta-1} \frac{R^f}{M_x}, \tag{9}$$

where $R^f = P^f Y^f$ is total export sales and

$$\tilde{\varphi}_x = \left[\frac{1}{M_x} \int_{\Omega_x} (\varphi(\omega))^{\theta-1} d\omega\right]^{\frac{1}{\theta-1}}$$

is the average productivity among exporters. Note that the set of exporters Ω_x affects the export revenues of each firm (9), both through the share of exporters M_x and the productivity distribution within the set.

Finally, we solve for the set of exporters Ω_x. A firm that exports incurs in a per period fixed cost. As a result, a firm ω will find it profitable to export only if its net income abroad would cover the fixed expenses,

$$\frac{1}{\theta} r^f (\varphi(\omega)) \geq f(\omega), \tag{10}$$

where export net income—that is, export revenues minus costs—is expressed as $r^f (\varphi(\omega))/\theta$. Thus the set of exporters Ω_x is the set of firms $\omega \in \Omega$ such that the entry condition (10) holds.

4.2. Strict Sorting and Latent Heterogeneity

We now consider two versions of the model. In the first we stick to the original Melitz model and have a single fixed cost, common to all firms. As a result exporters and non-exporters are strictly sorted by size and the model inherits the inability to match the size of exporters as documented in Section 2. We name this first version of the model after the strict-sorting property. In the second calibration we use the dispersion on fixed costs to reproduce the latent variable distribution we worked out in Section 3. By construction the model then matches the share and size of exporters. We label this calibration as the "latent heterogeneity" model.

It must be emphasized that the *only* difference between the two models is the distribution of fixed costs. The models will share the same parameter values for the elasticities, trade costs, and the distribution of productivity. However, we will set the foreign income Y^f parameter differently in each model to ensure that both models match the amount of exports observed in the data.

We start with the common parameters. We set the elasticity of substitution across exported goods to $\theta = 8$. This number is essentially in the middle of the range of estimates surveyed by Anderson and Van Wincoop (2004). For the aggregate demand for home exports (equation (8)) we set the price elasticity of $\nu = 6$. Our baseline trade costs are set at 50%, $\tau = 1.5$. This is the midpoint between the estimated trade costs in 1987 and 2002 reported by Alessandria and Choi (2013), and slightly below the findings in Anderson and Van Wincoop (2004).

The last common parameter is the distribution of productivity, G. For the calibration we want to use the observed distribution for total sales. Unfortunately the mapping from the productivity distribution to the total sales distribution is not the same for both versions of the model.[22] It is thus not possible to have a common parametrization for

---

22.   Both models share the mapping from G to the distribution of *domestic* total sales. The distribution of export sales, though, are different for each model.

TABLE 5. Calibration—common parameters.

Parameter		Value
Elasticity of substitution	θ	8
Price-elasticity exports	ν	6
Trade cost	τ	1.5
Std dev. log-productivity	σ_φ	0.37
Mean log-productivity	μ_φ	−0.48

G that matches the observed distribution of total sales in both models. However, we do not want any difference in the calibration of the models to govern the results—other than the size of exporters, that is. We thus adopt the following compromise: we set the productivity distribution to capture *all* of the variation in total sales. More precisely, we take G to be a lognormal distribution with standard deviation equal to $\sigma_\varphi = \sigma_r/(\theta - 1)$. The location parameter can be set such that the average productivity $\tilde{\varphi}$ among all producers is normalized to one. Table 5 summarizes all of the parameters common to both models.

Next we get to what makes the strict-sorting and latent-heterogeneity models different. For the strict sorting we shut down all of the variation in the fixed cost, so H is a degenerate distribution at some point f. The value of the fixed cost f is set such that the share of exporters is 18% as in the data. For the latent-heterogeneity model we calibrate the distribution of fixed costs H such that we capture all the variation in the latent variable t_i. For this it is very convenient to use a log-normal distribution for fixed costs so the mapping is simple. By construction the latent-heterogeneity model replicates the observed exporter size premium; by appropriate choice of the location parameter, we match the share of exporters as well. The resulting parameters are $\mu_f = 0.98$ and $\sigma_f = 7.69$.

Finally, we set foreign income parameter Y^f in each model to ensure that both models reproduce the ratio of total exports to total sales in the industry. The resulting parameter differs significantly across models: the predictions for trade levels are very different. To see this, we evaluate total exports over domestic sales in both models for a common value for the foreign demand Y^f set such that domestic income is roughly one fifth of the world income. For this baseline, both models overestimate the share of exports over total sales, but the latent heterogeneity only slightly so. More precisely, the strict-sorting model predicts 67% more exports than the latent-heterogeneity model, and an exports to total sales ratio of almost 0.28. The latent-heterogeneity predicts a ratio of 0.17, which is not too far from the observed ratio of 0.15. It is easy to see why the strict-sorting model overpredicts the ratio of exports to total sales. Strict sorting implies that exporters are very large, thus very productive. Hence, given the same

world demand, exports in the strict-sorting model are priced significantly lower than in the latent-heterogeneity model.[23]

5. Trade Liberalization

We compare the two models' predictions for export growth in response to a fall in trade costs. The exercise is labeled as a "trade liberalization", but there is no distinction in our model as to whether it is a tariff reduction or an improvement in the shipping technology.

5.1. Aggregate Response

We find that both models have virtually identical implications for trade volume. Figure 1 plots the growth rate of different variables (as a percentage rate) as a function of the fall in trade costs (in percentage points). The solid and dashed lines correspond to the strict sorting and latent heterogeneity models, respectively. The top left panel displays the growth rate of exports. Both models predict that trade will approximately triple once trade costs are cut in half. The models' predictions are so similar that the lines are on top of each other for most of the range of trade costs. The top right panel in Figure 1 displays the growth rate in the export price index, as given by equation (7). Again there are no differences between both models, and the price index falls with trade costs at the same rate for both models. Hence the profile for export revenues is very similar (bottom left panel). Finally, the bottom right panel in Figure 1 displays total employment in the industry. The trade liberalization results in a sizable expansion of employment, about 10% for the largest fall in trade costs considered.

Why do the two models predict such similar aggregate trade patterns? The reason is that in both models the aggregate variables are mainly driven by the demand for exports, equation (8). In the Online Appendix we show that the demand for exports in a representative-firm model can be simply written as

$$\log(R_f) = -(\nu - 1)\log(\tau) + \text{const.} \tag{11}$$

In Figure 2 we have added the representative-firm model, indicated with a dotted line. We plot export growth (top) and the export price growth (bottom) as a function of the fall in trade costs. Clearly, all three models are very similar in both prices and quantities.

It may be puzzling that firm-level heterogeneity and the extensive margin do not make a difference for aggregate variables. The reason is that firm-level heterogeneity

---

23.   The flip side of this very same exercise is that, for a common value for the foreign demand, the latent heterogeneity model would need lower trade costs to explain the observed export flows. In particular, the strict sorting model requires trade cost to be 10 percentage points higher than in the latent heterogeneity model.
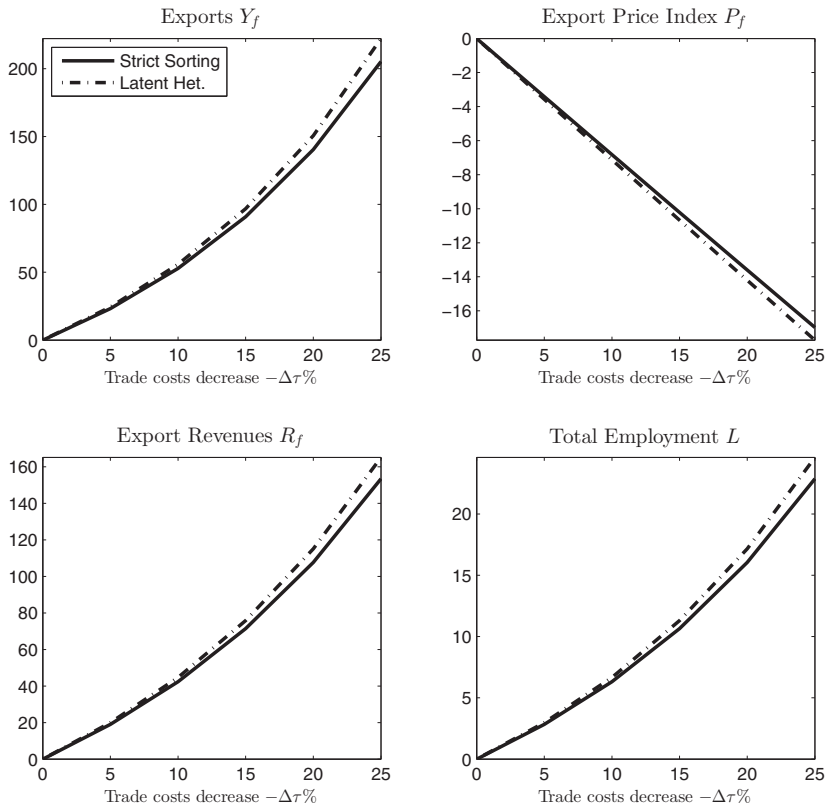
FIGURE 1. Aggregate exports; trade costs decrease in percentage points; all variables in growth rates.

essentially cancels the boost in trade due to the extensive margin. We can illustrate this by comparing the representative firm model with an homogeneous-firm model and a heterogeneous-productivity firm model with a Pareto distribution.[24]

Endogenous entry indeed amplifies the response of export revenues to a fall in trade costs. For an homogeneous firm model, export revenues are given by

$$\log(R_f) = -\frac{(\theta - 1)(\nu - 1)}{\theta - \nu} \log(\tau) + \text{const.} \qquad (12)$$

The elasticity of export revenues with respect to trade costs is augmented by $(\theta - 1)/(\theta - \nu)$ relative to the representative firm model. Everything else constant, more exporters bring the aggregate price of exports down through a love-of-variety effect. As long as there is entry, the price of exports will then fall by more than one-to-one with trade costs and sustain additional demand for exports. For the choice of

---

24.   These models are described in detail in the Online Appendix. The notation for parameters is the same as in the main framework presented earlier.
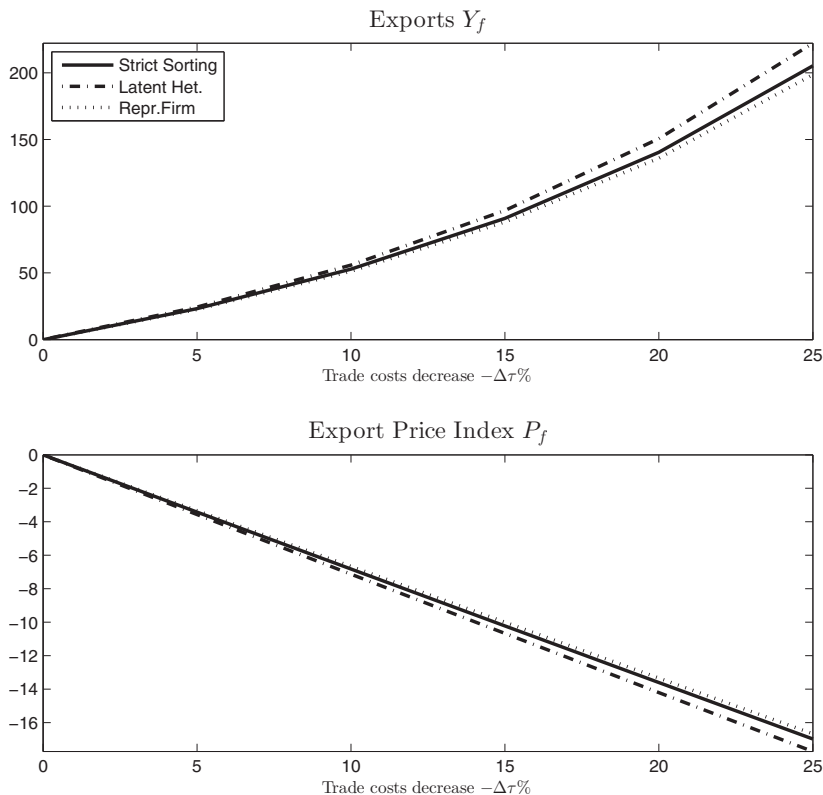
FIGURE 2. Aggregate exports—comparison with the representative-firm model; trade costs decrease in percentage points. All variables in growth rates.

parameters reported in Table 5, the response of trade volume is more than twice as large in the homogeneous firm model than in the representative firm model.

   We now look at a simple heterogeneous-productivity firm model with a Pareto distribution. Aggregate exports are given by

$$\log(R_f) = -\left[\frac{1}{\nu - 1} - \frac{1}{\theta - 1} + \frac{1}{k} \right]^{-1} \log(\tau) + \text{const.,} \tag{13}$$

where k is the slope of the Pareto distribution. Comparing (12) with (13) it is clear that the response in trade in the heterogeneous-firm model is between those in the representative-firm model and the homogeneous-firm model. As k grows large, the heterogeneous model approaches the homogeneous model and the export growth is the largest. This is not surprising, since the Pareto distribution becomes degenerate, with all of the mass at the location parameter.

   For low k, the Pareto distribution features a thick tail. In this case the response of exports to a fall in trade costs is quite muted. The firms that start exporting in response

to the fall in trade costs are less productive than incumbent exporters. Thus the average productivity among exporters falls rapidly, which in turn drives the aggregate price of exports up. The more skewed is the distribution of productivity, the faster the average productivity drops with entry.

Quantitatively, we find that the heterogeneity cancels virtually *all* of the amplification introduced by entry, rendering the heterogeneous and the representative firm models very similar in their implications for export growth. The right tail in the observed distribution of total sales is well approximated by a Pareto with a slope parameter ξ in the range 1.02–1.06. Setting $k = \xi(\theta - 1)$ we can rewrite the response of exports in the heterogeneous-firm model of equation (13) as

$$\log(R_f) = -\left[\frac{1}{\nu - 1} - \frac{1}{\theta - 1}\frac{\xi - 1}{\xi}\right]^{-1} \log(\tau) + \text{const.}$$

From this expression it is clear that if ξ is close to one, then the term $(\xi - 1)/\xi$ is approximately zero and the heterogeneous firm model is very close to the representative firm model (11).

5.2. Margins of Adjustment

Behind the similarities in the response of aggregate trade, the models display marked differences in the adjustment in the number of exporters. The top panel in Figure 3 plots the growth rate in the number of exporters for each model. In the strict sorting model the number of exporters grows very fast. Entry drives up the number of exporters by almost 60% when trade costs fall by 25 percentage points. Even with a small drop in trade costs, such as 5 percentage points, the number of exporters grows by more than 10%, suggesting that even at relatively high frequencies entry could play an important role.

In contrast, there is not much entry in the foreign market in the latent heterogeneity model. The growth rate of the number of exporters barely gets over 15%, one fourth of the growth in the strict sorting model. For smaller trade cost reductions such as 5%, the number of exporters is very close to flat. The similarities for total trade volume only make the differences in entry even more striking.[25]

We seek to quantify further the role of export entry in both models. We decompose the growth rate of exports in the change of export intensity and participation,

$$\Delta \log(Y^f) = \Delta \log(r^f) + \Delta \log(M_x). \tag{14}$$

The first term on the right-hand side is the intensive margin, that is, the growth rate on the average export revenues per exporter; and the second is the extensive margin, or the growth in the number of exporters. We normalize export growth by total sales

---

25. Differences in entry rates remain relevant even if both models are similar on the aggregate. First, it provides an additional set of predictions to contrast with the data. Second, entry is key to determine the impact of trade on aggregate productivity.

FIGURE 3. Entry and employment growth by exporters; trade costs decrease in percentage points; all variables in growth rates.

TABLE 6. The role of the extensive margin; trade costs decrease in percentage points.

Model	Trade cost decrease		
	5	15	25
Strict sorting	63.3%	63.3%	63.5%
Latent heterogeneity	19.2%	19.8%	20.4%

and express each margin as a percentage of the total.[26] Table 6 collects the results for trade cost reductions of 5, 15, and 25 percentage points. In the strict sorting model the extensive margin is more than 60% of the growth rate in exports. However, entry plays

_____

26. We are following Alessandria and Choi (2012) closely, although we do not distinguish between export intensity and premium as they do.

a much more minor role in the latent heterogeneity model, just below 20%.[27] These numbers are very similar across the range of trade cost cuts. They are also quite robust to alternative parameterizations for elasticities. This leads us to the conclusion that the Melitz model is at odds with a large role for the extensive margin, once it is augmented to account for the share and size of exporters.

The reader may be wondering how the two models can have such different entry rates and yet very similar export growth. The bottom panel of Figure 3 clues us in. Despite having much lower entry rates, total employment by exporters actually grows much faster in the latent heterogeneity model. The reason is that new exporters are very different in the two models. In the strict sorting model new exporters are very small in comparison with the incumbent exporters, about ten times smaller than the average exporter size prior to trade liberalization.[28] This is a direct consequence of the strict sorting: since exporters are almost 100 times larger than non-exporters, the firm at the threshold is still quite large compared with non-exporters but, more important, it is very small compared with exporters. Returning to Table 2 in Section 2, the *smallest* exporter has about five times more total sales than the average non-exporter, but 18 times less total sales than the average exporter.

In contrast, new exporters in the latent heterogeneity model are still smaller than existing exporters, but not by much. For a fall in trade costs of 5 percentage points, new exporters are just 40% smaller than the average exporter size prior to the trade liberalization. If we go all the way to a halving of trade costs, new exporters are actually 30% larger than the average exporter size prior to the trade liberalization. In short, new exporters are not very different from the average firm in the industry.

We now evaluate the main mechanism in Melitz (2003)—namely, the gains in average productivity in the industry due to the trade liberalization. We follow Melitz (2003) and define

$$\hat{\varphi} = \left(\tilde{\varphi}^{\theta-1} + M_x \left(\tilde{\varphi}_x / \tau \right)^{\theta-1} \right)^{\frac{1}{\theta-1}}$$

as the aggregate productivity.[29] It must be noted that our model can only do a partial evaluation. Since we take the wage rate as given, the set of non-exporters does not change so there are no productivity gains from the exit of the least productive firms as in Melitz (2003).

We find that aggregate productivity grows significantly less in the latent heterogeneity model than in the strict sorting model. Figure 4 plots the results. Aggregate productivity growth in the latent heterogeneity model is about half the one in the strict sorting model—a ratio that is approximately constant across the range of trade cost decreases. That productivity gains are smaller is not surprising given

---

27. Eaton, Kortum, and Kramarz (2011) also find that only a small fraction of trade growth is due to the extensive margin.

28. Because previous exporters grow rapidly with the trade liberalization, the new exporters are 50 times smaller when compared with the average size of the incumbent exporter after trade costs decrease.

29. Melitz (2003) refers to this as combined average productivity and shows that it completely summarizes the effects of the distribution of productivity levels on the aggregate outcome.
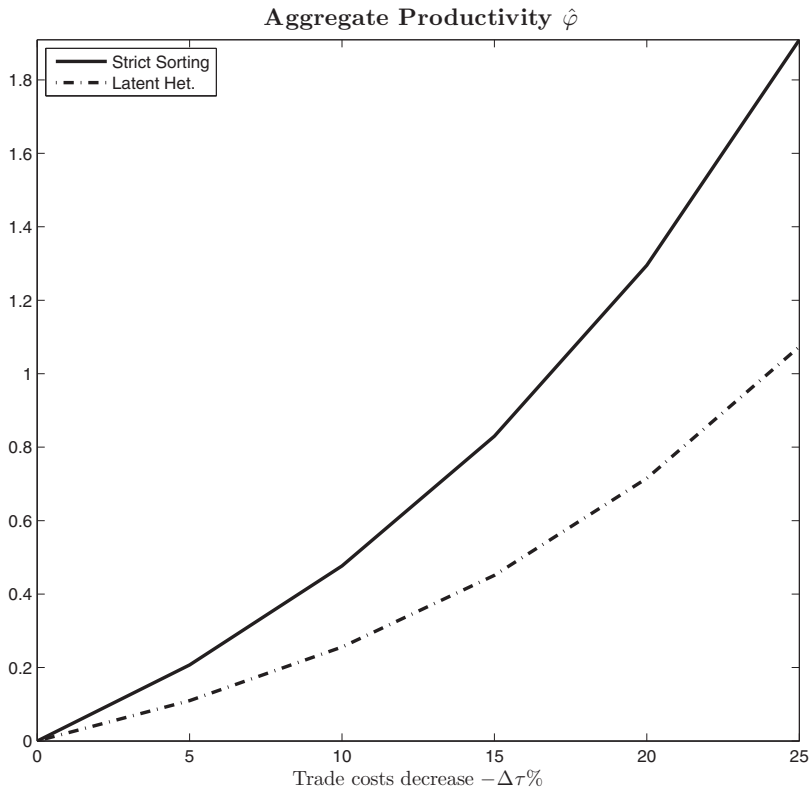
FIGURE 4. Aggregate productivity; trade costs decrease in percentage points; all variables in growth rates.

our previous results. Given that we abstract from exit in the domestic market, it can be said that the mechanism in Melitz (2003) works through the selection and entry of exporters. First, exporters are more productive than the average firm, so anything that expands the total employment of exporters will lead to productivity gains due to composition effects. Second, new exporters experience a large jump on their output, and since they are still more productive than most firms in the economy, also induce gains in aggregate productivity.

Both selection and entry are much weaker in the latent heterogeneity model than in the strict sorting model. We have seen that strict sorting greatly overpredicts the size of exporters—that is, their productivity. The latent heterogeneity reconciles the model and data: exporters are still more productive than non-exporters, but only modestly so. Moreover, there is much less entry in the latent heterogeneity model, so the second source of productivity gains is weaker too. The latent heterogeneity model makes up for the lower productivity growth with higher employment growth (see Figure 3), reflecting the differences in the intensive–extensive margin decomposition across models.
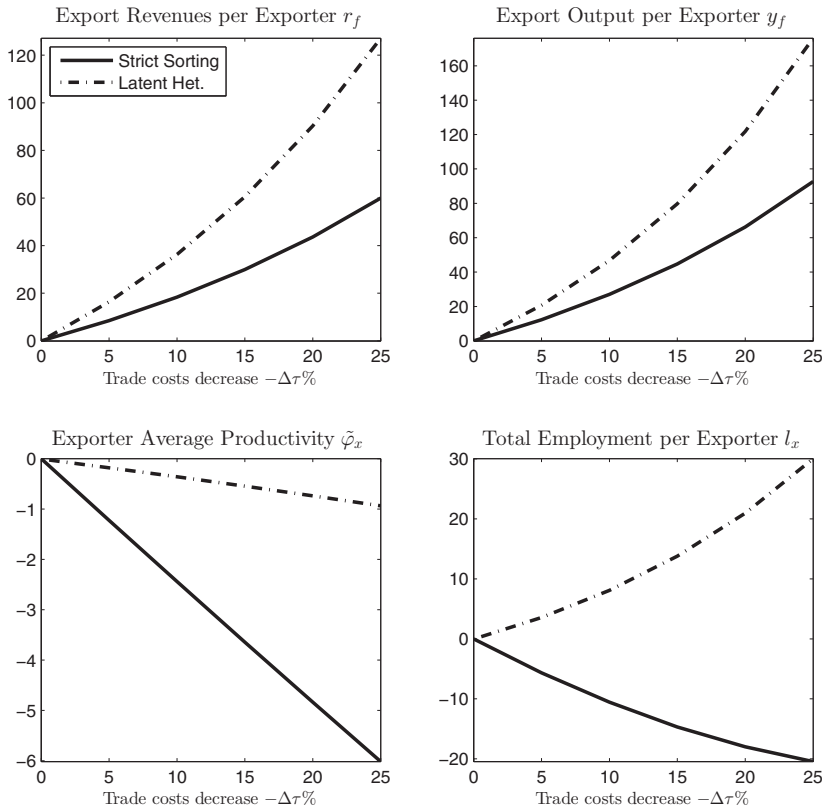
FIGURE 5. Exporters; trade costs decrease in percentage points; all variables in growth rates.

We conclude our analysis with a look at how the average exporter changes with trade liberalization. Figure 5 displays the average among exporters for export revenues, export output, average productivity, and total employment. It must be emphasized that the set of exporters changes as we cut the trade costs. First, we note that export revenues and output grow much faster in the latent heterogeneity model. This is, of course, the flip side of the results on the extensive margin documented in Table 6. Second, the average productivity for exporters falls in both models but by different amounts, as shown in the bottom left panel in Figure 5. In both models new exporters are less productive than incumbent exporters but in the strict sorting model, they are much less so. As a result the average productivity among exporters drops by a staggering 6% while it only does so a modest 1% in the latent heterogeneity model.

The differences in the average output and productivity of exporters combine for opposite predictions with respect to total employment for exporter. In the strict sorting model exporters, on average, employ fewer workers as trade costs fall. New exporters do employ more workers than they did before entering the foreign market. However,

TABLE 7. Robustness analysis. All variables are growth percentage rates after a fall of 10 percentage points in trade.

	Baseline		$\theta = 12$		$\nu = 3$	
	SS	LH	SS	LH	SS	LH
Total exports	52.8	55.8	52.3	54.1	23.3	23.7
Number of exporters	20.3	6.1	20.1	6.0	7.7	2.4
Agg. productivity	0.5	0.2	0.1	0	0.5	0.2

they are so small compared with the incumbent exporters that they bring the average down by an astounding 25%. Recall that the number of exporters grows by more than 50% so entry has a big impact on averages. In contrast, total employment for exporters grows in the latent heterogeneity model, as the weaker selection and entry effects cannot overturn the employment gains due to overall expansion of exports.

5.3. Robustness

We briefly discuss here alternative specifications for the common structure of both models. We start with our choice of the lognormal distribution for firm productivity. In particular, theory work has favored instead the Pareto distribution for its tractability. However, the Pareto distribution proves to be a very restrictive choice for our purposes. A Pareto distribution for firm productivity actually implies that average export sales per exporter are constant: while existing exporters sell more, new and thus smaller exporters enter until the average export sales returns to the previous value, pinned down by the entry cost. Hence the growth rate of trade will be equal to the growth rate of entry.[30] This exclusive role of the extensive margin is clearly at odds with the data.

Finally, we also explore some alternative parameter specifications. Table 7 reports the several moments for both the strict sorting (SS) and the latent heterogeneity (LH) model. We consider two alternative parameterizations. In the first, we set the elasticity of substitution between exports to 12. There is no significant change for both models' predictions for the growth rate of exports and exporters, but there is now virtually no aggregate productivity growth. This shows that the "love of variety" effect is the main driver of aggregate productivity. We also look at a second parameterization with a very low price elasticity of aggregate exports. Naturally, total trade growth is smaller as demand does not react to the fall in costs. Interestingly, aggregate productivity displays similar gains.

---

30.   It is important to recall that we have abstracted from general equilibrium effects on the wage rate, which, in turn, could have affected the fixed costs associated with exporting. This channel is operative in Melitz (2003) and leads to adjustment along the intensive margin.

6. Beyond Fixed Costs

6.1. Heterogeneity in Trade Costs and Foreign Demand

An obvious alternative to heterogeneous fixed costs is that firms may have idiosyncratic trade costs or perhaps foreign consumers do not value quality as domestic consumers do.[31] Since foreign sales are a small fraction of total sales for US firms, heterogeneity in trade costs or foreign demand would be only weakly correlated with total sales and uncorrelated with domestic sales.

We find that the model implications are virtually unchanged if the necessary variation to match the data is modeled as heterogeneous fixed costs, trade costs, or foreign demand.[32] As long as the additional heterogeneity enters the exporter revenues multiplicatively, then the resulting model is actually isomorphic to heterogeneity in fixed costs. The reason is that the model is essentially loglinear. Consider this simple specification: on top of the trade costs τ, the marginal cost of firm's ω exports is $(1 + \eta(\omega))$ times more than the marginal cost for output sold at the domestic market. That is, firms differ on the *variable* costs associated with exporting. We can rewrite equation (9) in this model as

$$r^f(\omega) = \left(\frac{\varphi(\omega)}{\tilde{\varphi}_x}\right)^{\theta-1} (1 + \eta(\omega))^{-\theta} \frac{R^f}{M_x}.$$

The entry condition (10) would now be

$$\left(\frac{\varphi(\omega)}{\tilde{\varphi}_x}\right)^{\theta-1} (1 + \eta(\omega))^{-\theta} \frac{R^f}{M_x} \geq \theta f(\omega).$$

We could, though, simply define $\tilde{f}(\omega) = f(\omega)(1 + \eta(\omega))^{\theta}$ and the equilibrium condition is identical. Quantitatively we would calibrate the distribution of both η and f to reproduce the variation in the latent variable—in other words, we would calibrate the variable $\tilde{f}(\omega)$ as we did with $f(\omega)$ in Section 4. So as long as η and f remain independent of φ, the results would be unchanged.

6.2. Sunk Costs

It is well known that the data display a large amount of exporter hysteresis. While we are not concerned with the dynamics of exporting per se, export hysteresis has important implications for the cross-section of exporters. Consider a simple model that combines a fixed per-period cost of exporting with a sunk cost of entry in the

---

31. Munch and Nguyen (2008) find that firm-specific factors such as productivity explain a very small fraction of the sales variation across destinations for Danish exporters.

32. As long, of course, as the full amount of latent heterogeneity is captured by the specification of choice.

foreign market.[33] In the presence of sunk costs, the exporter status of a firm is not determined by its current productivity alone. An exporter may have no reason to stop exporting for productivity level φ, since it can still cover per-period fixed costs. The same productivity level, though, may not be high enough to convince a firm to start exporting as it will not cover the sunk costs. It is thus possible that the history of each firm provides the necessary heterogeneity to match the exporter size premium.

While the sunk cost model breaks the strict sorting between exporters and non-exporters, it now features strict sorting between *new* exporters and non-exporters. Because the entry rate is so small, the threshold for entry must be way deep in the tail of the distribution. In the Online Appendix we provide a simple analysis of a sunk cost model, based on the approach developed in Section 3. We find that new exporters would be predicted to about 200 times larger than non-exporters, and several times bigger than existing exporters. All empirical evidence points to new exporters being significantly smaller than existing exporters—so it looks like to reconcile the sunk cost model with the size of new exporters we will need, again, a lot of idiosyncratic variation independent of firm size.

6.3. Relaxing Independence

We now drop the assumption that the latent variable t_i is independent of firm size. Some firm characteristics—for example, managerial quality, geographical location—may simultaneously affect total sales and the costs associated with exporting.

We now allow for total sales and the latent variable $\{r_i, t_i\}$ to be *jointly* distributed according to a log-normal distribution with mean μ and variance–covariance matrix Σ. We parameterize the joint distribution in terms of the latent variable as $\log(t_i) = \gamma \log(r_i) + \log(\varepsilon_i)$, where ε_i is independent of total sales r_i and distributed according to a log-normal distribution with mean μ_ε and standard deviation σ_ε. The parameter γ governs the correlation between total sales r_i and the latent variable t_i.

It is easy to rewrite the model in terms of total sales r_i and independent variation ε_i. Firm i exports if $r_i \geq t_i$, or, in logs and in terms of ε_i, $(1 - \gamma) \log(r_i) \geq \log(\varepsilon_i)$. Given a value for γ we can proceed as we did in Section 3 to find values for μ_ε and σ_ε such that the model matches the share and size of exporters observed in the data.

We find that total sales and the latent variable have to be positively correlated in order to reduce the amount of variation independent of firm size—that is, $\gamma > 0$. The reason is quite straightforward. Strict sorting overestimates the size of exporters. Having larger firms clear, on average, a higher hurdle to start exporting reduces the number of large firms that export, and allows the model to match the data with less size-independent variation.

---

33.   Baldwin (1988) and Baldwin and Krugman (1989) are credited with developing the first theories of export hysteresis. Roberts and Tybout (1997) evaluate empirically the role of sunk costs through reduced-form models. More recent work has estimated structural models with sunk costs; see Das, Roberts, and Tybout (2007) and Ruhl and Willis (2008), as well as Alessandria and Choi (2007, 2013).

A positive correlation between firm size and the latent variable questions the hypothesis that firms face *fixed* costs associated with exporting. A more natural explanation is that domestic and export production both require some input that is on a fixed supply at the firm level like, for example, a manager's limited span of control as in Lucas (1978). That said, the positive correlation does not rule out nonconvexities but rather suggests that they require a more nuanced approach. For example, Cooper and Haltiwanger (1993) assume that firms face some down time when replacing "machines" or, more generally, when pursuing some investment such as upgrading the production process. The nonconvexity arises because it is not possible to replace, say, a fraction of a machine with a fraction of the down time. The down time is more expensive for more productive firms, so while the investment technology is identical across firms, the *cost* of investment is higher for larger firms.

6.4. Zeros Without Economies of Scale

Economies of scale are not the only basis for non-exporters. An alternative is that firms compete head-to-head abroad, with some firms losing out to foreign producers and thus failing to export. This is, in a nutshell, the model in Bernard et al. (2003). More productive firms are more likely to export, but there is no strict sorting, as a small, less productive firm may be lucky enough to be competing with a very low-productivity foreign competitor while a large firm may instead face a juggernaut. However, Bernard et al. (2003) are only able to match the observed exporter size premium by overstating the share of exporters. Once the model is set to reproduce the observed share of exporters, it predicts an exporter size premium about 75.[34]

Yet another alternative put forward by Armenter and Koren (2014) is that all firms have a chance to export but, due to the sparsity of the data, not all of them do so during a given year. But, as Armenter and Koren report, sparsity cannot explain the low number of exporters. Interestingly, it also overpredicts the size of exporters despite that there is no strict sorting in their model.

6.5. Asking the Data

Ultimately, which are the key determinants of a firm's export status is a question that should be posed to the data. Perhaps the geographical location plays an important role? Do other international linkages—imported inputs or FDI—increase the likelihood that a firm exports? Unfortunately, our data are severely limited in this aspect and we can only address one hypothesis: that the latent heterogeneity stems from between-industry variation. Manufacturing includes goods as diverse as tobacco products and machinery. So it is quite reasonable to think that sectors face very different trade costs, both fixed and variable. In the Online Appendix we estimate the size premium implied by strict sorting for each three-digit NAICS industry code.[35] The procedure is the same

---

34.   See the Online Appendix for details.

35.   Unfortunately we do not have access to the same data for finer levels of disaggregation.

as the one we used in Section 2 for manufacturing as a whole. Sectors vary greatly regarding the share of firms exporting. In Printing only 5% of the firms exports, while in Computers and Electronic Products almost 40% of the firms do.

   We find that *for each sector* the predicted exporter size premium is very large, well above any estimate. The reason is that the firm size distribution within a sector remains very skewed, so any strict sorting exercise is bound to return large size premia. The differences on the share of exporters does create a lot of variation in the implied size premiums across sectors. However, it does not get them in the range observed in the data. Heavily traded sectors, like computer or electrical equipment, have implied size premiums of a magnitude comparable to the manufacturing sector as a whole. Machinery (another commonly traded) has a somewhat lower size premium, between 40 and 50, but it is still well above the differences documented in the data.

7. Conclusions

In this paper we have argued that a simple Melitz model cannot match *quantitatively* the basic cross-sectional facts about exporters—namely, their frequency and size. The model can be easily reconciled with the data by introducing enough additional sources of variation. We find, though, that a large amount of variation independent of firm size—and thus, unexplained—is needed. Moreover, introducing the additional variation can change the model's implications in the event of a trade liberalization. It is thus imperative for any quantitative model operating a selection mechanism à la Melitz to be able to match both the share of exporters and their size premium.

References

Alessandria, George and Horag Choi (2007). "Do Sunk Costs of Exporting Matter for Net Export Dynamics?"*Quarterly Journal of Economics*, 122, 289–336.

Alessandria, George and Horag Choi (2012). "Do Falling Iceberg Costs Account for Recent US Export Growth?" Working paper, FRB Philadelphia.

Alessandria, George and Horag Choi (2013). "Establishment Heterogeneity, Exporter Dynamics, and the Effects of Trade Liberalization." Working paper, FRB Philadelphia.

Anderson, James E and Eric van Wincoop (2004). "Trade Costs." *Journal of Economic Literature*, 42, 691–751.

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

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodrí-guez-Clare (2012). "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1), 94–130.

Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow, and Andrés Rodríguez-Clare (2008). "Endogenous Variety and the Gains from Trade." *American Economic Review*, 98(4), 444–450.

Armenter, Roc and Miklós Koren (2014). "A Balls-and-Bins Model of Trade." *American Economic Review*, 104(7), 2127–2151.

Atkeson, Andrew and Ariel Burstein (2010). "Innovation, Firm Dynamics, and International Trade." *Journal of Political Economy*, 118, 433–484.

Axtell, Robert (2001). "Zipf Distribution of U.S. Firm Sizes." *Science*, 293, 1818–1820.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu (2011). "R&D Investment,Exporting, and Productivity Dynamics." *American Economic Review*, 101, 1312–1344.

Baldwin, Richard (1988). "Hysteresis in Import Prices: The Beachhead Effect." *American Economic Review*, 78(4), 773–785.

Baldwin, Richard and Paul Krugman (1989). "Persistent Trade Effects of Large Exchange Rate Shocks." *Quarterly Journal of Economics*, 104, 635–654.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum (2003). "Plants and Productivity in International Trade." *American Economic Review*, 93(4), 1268–1290.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott (2007). "Firms in International Trade." *Journal of Economic Perspectives*, 21(3), 105–130.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott (2007). "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." In *Producer Dynamics: New Evidence from Micro Data*, edited by J. Dunne, J. B. Jensen and M. J. Roberts. University of Chicago Press pp. 513–556.

Besedes, Tibor and Thomas Prusa (2011). "The Role of Extensive and Intensive Margins and Export Growth." *Journal of Development Economics*, 96, 371–379.

Broda, Christian and David Weinstein (2006). "Globalization and the Gains from Variety." *Quarterly Journal of Economics*, 121, 541–585.

Cooper, Russell and John Haltiwanger (1993). "The Aggregate Implications of Machine Replacement: Theory and Evidence." *American Economic Review*, 83(3), 360–382.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout (2007). "Market Entry Costs, Producer Heterogeneity, and Export Dynamics." *Econometrica*, 75, 837–873.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2004). "Dissecting Trade: Firms, Industries, and Export Destinations." *American Economic Review*, 94(2), 150–154.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2010). "An Anatomy of International Trade: Evidence from French Firms." Penn State University Working paper.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2011). "An Anatomy of International Trade: Evidence from French Firms." *Econometrica*, 79, 1453–1498.

Feenstra, Robert (1994). "New Product Varieties and the Measurement of International Prices." *American Economic Review*, 84(1), 157–177.

Hallak, Juan Carlos and Jagadeesh Sivadasan (2011). "Firms' Exporting Behavior Under Quality Constraints." Research Seminar in International Economics No. 628.

Halpern, László, Miklós Koren, and Ãdám Szeidl (2011). "Imported Inputs and Productivity." Central European University Working paper.

Hummels, David and Peter J. Klenow (2005). "The Variety and Quality of a Nation's Exports." *American Economic Review*, 95(3), 704–723.

Johnson, Robert Chistopher (2012). "Trade and Prices with Heterogeneous Firms." *Journal of International Economics*, 86, 43–56.

Kugler, Maurice, and Eric Verhoogen (2012). "Prices, Plant Size, and Product Quality." *Review of Economic Studies*, 79, 307–339.

Lucas, Robert (1978). "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9, 508–523.

Luttmer, Erzo (2007). "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics*, 122, 1103–1144.

Melitz, Marc J. (2003). "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71, 1695–1725.

Munch, Jacob and Daniel Nguyen (2008). "Decomposing Firm-Level Sales." University of Copenhagen Working paper.

Nocke, Volker and Stephen Ross Yeaple (2012). "Globalization and Multiproduct Firms." CEPR Discussion Paper No. 9037.

Roberts, Mark and James Tybout (1997). "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review*, 87(4), 545–564.

Ruhl, Kim (2008). "The International Elasticity Puzzle." UT Austin Working paper.

Ruhl, Kim and Jonathan Willis (2008). "New Export Dynamics." NYU Working paper.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix for Economies of Scale and the Size of Exporters
Replication Codes