Multi-product Exporters: Facts and Fiction

JEL: D24, L11, F14

Lena Sheveleva[1]

Cardiff Business School, shevelevay1@cardiff.ac.uk

August 28, 2020

[1]The author acknowledges the useful discussions, advice and support from Milagro Rogriguez-Saborio, Kala Krishna, Peter Neary, Markus Eberhardt, James Tybout, David Collie and Sergey Popov.

Abstract

In this paper, we propose a statistical model of multi-product exporters where pure chance determines exporter outcomes. We show that some well-documented differences between large and small scope exporters that have motivated many theoretical models are also consistent with our statistical model. The statistical model provides a particularly good fit for patterns that involve order statistics. For example, it accounts for the differences between how much large and small scope exporters sell in their best/least selling products and variation in the ratio of sales between the best and the second best selling.

1 Introduction

Across the world, multi-product firms dominate international trade flows. Famously, Bernard et al. (2010) document that multi-product firms account for 98% of the U.S. manufacturing exports. In other countries multi-product exporters play an equally important role. In light of this empirical importance, researchers are interested in understanding why multi-product exporters arise and how they respond to changes in the international trade environment. In recent years there has been some progress in answering these questions both on the empirical and theoretical fronts. For example, Allanson and Montagna (2005), Nocke and Yeaple (2015), Bernard et al. (2011), Eckel and Neary (2010), Dhyne et al. (2017), Arkolakis et al. (2020), and Bernard et al. (2010) are just few of the papers that explore how to model multi-product exporters and their response to changes in the international trade environment. These papers feature product level heterogeneity either on cost or demand side but also emphasize firm level differences in productivity, brand-effects, as well as differences in the costs of introducing new products. In part, this views on multi-product exporters have been facilitated by a number of empirical regularities documented using the newly available firm-product level data.

In this paper we take a step away from the behavioral models of multi-product exporters, and instead focus on the empirical regularities that underpinned those models. We aim to understand which ones are informative

about the economic forces that drive multi-product exporters and which ones arise as a result of randomness and aggregation. The original motivation for this research question comes from the work of Ellison and Glaeser (1997), who highlight that data patterns and predictions of economic models should be compared to patterns that would arise if the outcomes of interest were randomly distributed rather than to a uniform pattern or an absence of a pattern.

This paper is also close in spirit to the work of Armenter and Koren (2014) who highlight that many facts about the extensive margin of trade are consistent with a surprisingly large class of models because of the sparse nature of trade data. In relation to multi-product firms they show that the "bins-and-balls" model quantitatively reproduces the frequency of single product exporters at a destination but not the right tail of the exporters' scope distribution. In this paper we take exporter scope as given and focus on the differences between large and small scope exporters on the intensive margin. We show that pure random chance accounts for some notable differences between large and small scope exporters because these differences involve aggregating data over the different number of products to the firm level. This is particularly the case for the patterns that involve order statistics, i.e., best, second best, and least selling products of multi-product exporters.

First, we address facts about multi-product exporters at a single destination. These have motivated several explanations as to why multi-product exporters dominate international trade. Then we discuss extensions to the

multi-country case and how multi-product exporters respond to differences in competitive environments across destinations.

Consider the following three facts that have been documented across a number of data-sets and countries[1].

1. Large scope exporters sell more in their best selling products than small scope exporters.

2. Large scope exporters sell less in their least selling products than small scope exporters.

3. Average sales per product are non-monotone with scope.

The first fact is consistent with many models of multi-product exporters where a single firm attribute, usually firm productivity, drives both scope (the ability to produce many products) and scale (the ability to produce at scale due to lower marginal cost). Examples include Bernard et al. (2011), Bernard et al. (2010), and Arkolakis et al. (2020). To set ideas straight, consider a Melitz's style model where firms face random product-specific shocks (on the cost or demand side) and a single firm productivity which translates into a marginal cost parameter common across all of its products. All else constant, lower marginal cost implies higher product sales and higher variable profits, which encourage high productivity firms to export more products. While not all products exported by high productivity exporters sell in large amounts

[1] Arkolakis et al. (2020) and Timoshenko (2015) document these patterns for Brazil; Bernard et al. (2011) and Bernard et al. (2010) document them for the U.S. In this paper we document them for Chinese exporters to the U.S.

due to product-specific shocks, the Melitz's style model implies a positive correlation between sales and scope. Indeed, the model is also consistent with the fact that large scope exporters sell more in their top-selling products than do small scope exporters.

The second fact is consistent with models where a product introduction fee (also known as a product specific market access cost or a fixed cost of exporting)[2] decreases with exporter scope as in Arkolakis et al. (2020) and Timoshenko (2015) [3]. A multi-product firm expands its scope as long as the profit from the marginal product exceeds the product introduction fee. Firms that face lower product introduction fees are able to profitably export less efficient products that sell in smaller amounts (Qiu and Zhou (2013)). To the extent that sales of the least selling product reflect the exporter's marginal product, the Fact 2 is consistent with a model where product introduction fees decrease with scope. With large scope exporters selling some of their products in extraordinarily large amounts, and other in tiny amounts, average sales per product are then indeterminate as a function of scope due to composition effects (Fact 3).

In this paper, we show that the facts above, while consistent with several models of multi-product exporters, can also arise due to random chance alone. In other words, there is no evidence in the data to support that large scope exporters sell their best (least) selling products in larger (smaller)

[2]In a static Melitz's style model a product introduction fee and the market access cost are equivalent to the fixed cost of exporting a product.

[3]Arkolakis et al. (2020) refer to the product introduction fee as a market access cost.

amounts than small scope exporters beyond what we would expect if sales were randomly distributed across firms and products.

To see the intuition, consider a two product firm and a ten product firm. Suppose that sales per product are just iid draws from the same distribution across firms and products, then the expected value of the maximum of ten products is higher than the maximum of two products. The reverse is true for the minimum. Hence larger scope exporters may have higher sales in their best selling products and lower in their least selling products relative to the smaller scope exporters simply because we aggregate data over a different number of products.

Formally, Facts 1 and 2 above are examples of order statistics, i.e., average sales across products at a given rank. Order statistics are systematically related to the number of products for which they are calculated. So mean sales of the best selling product increase with exporter scope while mean sales of the least selling product decrease with exporter scope even if all exporters drew sales from the same distribution. If more productive firms select into exporting more products then we should see that large scope exporters not only sell more than small scope exporters in their top-ranked products, but also that they sell more than if their sales came from the same distribution as single product exporters. Similarly, if large scope exporters faced lower product introduction costs, they would sell less in their least selling products compared to the small scope exporters, and less than if their sales were driven by chance.

While in theory chance and aggregation can generate the observed differences between large and small scope exporters, whether they are sufficient to replicate these quantitatively is an empirical question. To answer it we formalize chance in an intentionally stark statistical model of multi-product exporters. In the model the number of exported products per firm is treated as exogenous, and product revenues are iid realizations from some distribution F().

In our empirical implementation F() is either Pareto or log normal. We show that when $F()$ is log-normal, the statistical model replicates Facts 1 and 2 (as well as other order statistics) remarkably well. When $F()$ is Pareto, the statistical model tends to overestimate how quickly sales at a given rank increase/decrease with scope. The poor empirical performance of the Pareto stems from its failure to approximate the left tail of the sales distribution and echoes the findings of Head et al. (2014), Bee and Schiavo (2018), and Fernandes et al. (2015).

The fact that the statistical model can replicate Fact 1 highlights that one should be careful not to interpret it as evidence that large scope exporters are more productive. For example, in the data, Chinese exporters to the US with five or more products sell almost eight times as much as single product exporters in their top selling products.[4] Our results show that if large scope exporters drew sales from the same distribution as single product exporters the difference between how much large and small scope exporters sell in their

[4]See Table 1.

best selling products would be even greater. Similarly, Fact 2 on its own should not be interpreted as evidence of product introduction fees decreasing with scope. Variation in average sales per product, Fact 3, on the other hand cannot be replicated by the statistical model and provides a useful metric to compare large and small scope exporters.

Put another way, the firm scope appears to be a sufficient statistic for the differences between large and small scope exporters on the intensive margin. A model that can successfully replicate firm scope will also replicate observed differences between large and small scope exporters. Variance in firm-product shocks is enough to generate some of the observed differences between large and small scope exporters. In this regard, our work speaks to the literature on the role of firm productivity and product introduction costs in determining the scope and scale of multi-product firms. Our results echo Macedoni and Xu (2020), who showed that scope and measured productivity are only imperfectly correlated, and that variation in the ability to introduce new products plays a more important role in driving exporter scope. We contribute to this debate by highlighting that the notion of firm productivity being a key driver of both scope and scale should be treated with caution because the differences between large and small scope exporters seemingly caused by firm productivity can arise due to chance.

Finally, we extend the statistical model to the multi-destination setup to see if it can reproduce other empirical facts that rely on order statistics. We focus on a popular regularity that in more competitive markets multi-

product exporters tend to have their sales more concentrated in their top selling products. A measure of sales concentration that is often used in this context is the ratio of sales between the best selling and the second best selling product. One reason behind its widespread use is that some theoretical models predict the ratio as a function of variables that proxy market competitiveness at a destination. The best-known example of this is the seminal work by Mayer et al. (2014). They show that when product demand is linear, and firms face variable mark-ups, multi-product firms adjust their product mix by reallocating resources across products at a destination. Firms expand production of their core (low cost and high mark-up) varieties at the expense of the peripheral ones, thus increasing the concentration in the core (top selling) varieties. To test the prediction, Mayer et al. (2014) regress the log-ratio of sales between the best selling and the second best selling product on the destination market size, market supply potential, bilateral trade costs and other proxies of market competitiveness. Using the estimated impact of market size on the log-ratio together with their theoretical model they calculate that the within-firm resource reallocation contributes a nontrivial 19% to aggregate productivity.[5]

Our multi-destination extension of the statistical model highlights that the regression analysis where log-ratio is the dependent variable, as in Mayer

[5]Mayer et al. (2014) also use a global ratio and provide robustness checks in the appendix where they show that concentration increase is captured by the Theil and Herfindalh indices as well. However, their discussion of the economic implications and the main results in the paper are based on the log-ratio measure of concentration which is vulnerable to the bias due to the unaccounted number of products.

et al. (2014), may produce biased results. This is because unless the sales are distributed Pareto, the log-ratio, which is the function of two order statistics, systematically depends on the number of products. If sales are log-normally distributed, for example, the expected value of the ratio decreases with the number of products a firm exports to a destination. So, if firms export fewer products to a more distant destination or one with a higher foreign supply potential, the ratio would be higher there even if sales were randomly distributed as is in our statistical model. Our quantitative results show that the regression coefficient on the market size underestimates its effect on changes in the product mix. On the other hand, market supply potential or bilateral distance which are associated with smaller exporter scope will have coefficients biased upwards. Intuitively, the impact of factors that are associated with firms exporting more products to a destination will be underestimated, while the impact of factors associated with fewer exporters will be overestimated. Our results thus suggest that the within-firm resource reallocation may contribute more than the 19% to aggregate productivity that Mayer et al. (2014) calibrated based on the estimated coefficient of the market size.

In this paper we contribute to the empirical literature on multi-product exporters by highlighting that data regularities which involve order statistics, such as best/least selling products, systematically vary with the number of products for which they are calculated. Our insight is that instead of comparing sales at rank between exporters of different scope, one should compare sales that are observed to those that would arise if firms drew sales

from the same distribution. Some of the patterns used to characterize multi-product exporters turn out to be consistent with a wide variety of models, including the statistical model where ex-post differences between large and small scope exporters arise because they involve aggregating firm-product level outcomes driven by pure chance over a different number of products to the firm level.

One should, however, be careful not to interpret our results as evidence against behavioral models of multi-product exporters that feature firm productivity or resource reallocation in response to trade shocks. Beyond the well-documented patterns that we focus on here, there is work that provides direct evidence on these channels. For example, Dhingra (2013) find that firms invest in cost-cutting technology for their core products and drop peripheral products in response to trade liberalization in Malaysia . Dhyne et al. (2017) use a novel multi-product firm production function estimation approach to estimate technical efficiencies of individual products. They find that firms are more efficient at producing core products and respond to competition by focusing on them. Rather, one should see our results as insights into which data moments are informative about multi-product firm behavior and which should be considered with extra care.

The paper proceeds as follows. Section 2 lays out the statistical model and discusses the intuition behind it. Section 3 presents the calibrated results, Section 4 extends the baseline model to multi-country set-up and Section 5 concludes.

2 Statistical Model of multi-product Exporters

We now present a statistical model of multi-product exporter outcomes. We begin with a single destination version and later extend it to the multi-destination set-up in Section 4.

A multi-product exporter indexed by f is a collection of products K that a firm exports in a given year. Products within a firm are indexed by $k \in \{1, ..., K\}$. The number of products a firm exports (K) is treated as exogenous and in empirical applications will be taken from the data. Product sales S_{fk} are iid draws from some continuous distribution $F()$ with a non-negative support.

In this stark set-up, any differences between large and small scope exporters on the intensive margin are driven by aggregating random draws over the different number of exported products to the firm level. So by comparing the observed patterns to the patterns predicted by the statistical model, we will be able to separate the data patterns that are genuinely informative about the economic forces driving the intensive margin from the ones that reflect randomness and aggregation.

This statistical model has four main predictions about the distribution of product sales within a firm. When products within a firm are ranked by their contribution to total sales from the best selling to the least selling the model predicts:

Prediction 1. *Bigger scope exporters sell more in their best selling products*

than smaller scope exporters.

Prediction 2. *Bigger scope exporters sell less in the lowest ranked products than smaller scope exporters.*

In the model, the large scope exporters get to make more sales draws (one for each product) than smaller scope ones. As long as the draws are iid across firms and products, the maximum of the larger number of draws is bigger than the maximum of the smaller number of draws. The reverse is true for the minimum. In Section 3 we demonstrate that both predictions are independent from the assumptions about the distribution and find strong support in the data.[6]

While in the statistical model the two predictions arise due to randomness and aggregation, models where firm productivity and product introduction fees drive exporter scope and scale would generate similar predictions (at least qualitatively). Prediction 1 is consistent with the models of multi-product firms where firm productivity determines both scope and scale. Faced with random demand shocks as in Bernard et al. (2011) high productivity firms have higher expected profits from a new product and so select into more products. By virtue of higher productivity large scope exporters sell their best selling products in larger amounts than their small scope counterparts.

Prediction 2 is consistent with a model where a product introduction fee decreases with exporter scope. Exporters choose their scope by equating

[6]These patterns are also documented in Arkolakis et al. (2020), Timoshenko (2015), Bernard et al. (2011) and Bernard et al. (2010).

marginal profit from a product to the product introduction fee. When product introduction fees decrease with scope larger scope exporter are able to profitably sell less efficient products. Large scope exporters will then sell their least selling products in smaller amounts than the smaller scope exporters.

Prediction 3. *Average sales per product are constant with scope.*

Evidence on the relationship between average sales per product and firm scope is mixed. As a rule, average sales per product is not monotone with scope. See Bernard et al. (2011), Arkolakis et al. (2020). Many models of multi-product firms avoid making predictions about average sales per product because they reflect composition effects (i.e. large scope exporters sell their products in both extraordinarily large and small amounts). Yet, our results suggest that variation of average sales with scope cannot be replicated with randomness alone and so is a useful moment to differentiate between alternative models of multi-product exporters.

Prediction 4. *The expected value of the log ratio between sales of the best and the second best selling product is constant with scope when sales are distributed Pareto and decreases monotonically with scope when sales follow log-normal, Weibull or exponential distributions.*

In Appendix A, we present analytical proofs for Pareto, Weibull and exponential distributions. We use numerical integration to establish the result for the log-normal distribution.

The log ratio of the best selling to the second best selling product is often used as a measure of sales concentration within a firm. It is thought to capture how firms respond to changes in competition by adjusting quantities and prices across the product range. We highlight that there is also a mechanical relationship between the number of products a firm exports to a destination and the log-ratio, unless sales are Pareto distributed. This matters because within-firm reallocation of resources usually coincides with adjustments on the extensive margin. Faced with tougher competition, firms drop products which would have the effect of increasing the ratio even if firms did not respond optimally on the intensive margin. Hence, the effect of market competitiveness measures on the skewness of sales measured using the ratio is likely to be biased. We explore this issue in detail when we discuss the multi-destination extension of the model in Section 4.

3 Empirical Test of the Model

In this section, we test the predictions of the statistical model. When we compare predictions of the statistical model to the data we find that large scope exporters sell no more in their top selling products than if their sales were random. In fact they sell less in their top selling products than if they drew their sales from the same distribution as single product exporters (Prediction 1). Similarly, once we take into account randomness and aggregation, large scope exporters sell no less than small scope exporters in their least sell-

ing products (Prediction 2). This is in contrast to the conclusions one may make by comparing sales at rank between large and small scope exporters. However, the statistical model is unable to replicate the variation in average sales per product (Prediction 3). In terms of the log ratio of sales between the best selling and the second best selling product, the model replicates the rate at which the ratio decreases with scope but underestimates average ratio at a given scope (Prediction 4). In Section 4 we address the implications of Prediction 4 for measuring the exporters' response to changes in the competitive environment using a multi-country version of the statistical model.

3.1 Estimation Details

To test the predictions of the statistical model we use data on firm-product sales available at the HS-6 level for Chinese exporters to the US in 2003. In Table 1 we verify the multi-product exporter facts documented for other countries in the Chinese data. The table splits the sample based on the number of products each firm exported and for each group reports average total exporter sales, as well as average sales of the best and least selling products. While small in number, large scope exporters dominate the export market in terms of their sales. Single product exporters, for example, account for 41 percent of exporters while exporters with five or more products account for only twenty percent. Yet, exporters with five or more products sell ten times as much as single product exporters. Exporters with five or more

products sell almost eight times as much as single product exporters in their best selling products. They also sell much less in their least selling products than small scope exporters.

Table 1 – Summary Statistics for the Chinese Exporters to the US.

Scope	1	2	3 - 4	5+
Exporter Share with Scope	41	20	19	20
Mean Total Firm Sales	907	1,826	3,141	10,185
Mean Top Ranked Product	907	1,679	2,689	7,184
Mean Bottom Ranked Product	907	147	45	8

The table splits the sample of exporters from China to the US according to the number of products that they exported in 2003 into exporters with 1,2,3-4,and 5 or more products. For each group it reports the share of exporters, mean total exporter sales and mean sales of the best and least selling products for each group. Sales are in thousands of USD.

To test the predictions of the statistical model we first estimate the distribution of product sales $F()$ and then use the estimated distribution to simulate the moments about which the model makes predictions. We then compare the simulated moments to the data and infer which patterns can be explained with the statistical model alone, and which ones are informative about the economic forces that produced the data.

In the data, firm-product sales systematically vary with product categories (at the HS6 group level). To take this into account we adjust annual sales data (used in Table 1) relative to the average product sales across all Chinese exporters to the US selling in the same product category.

We use the log-normal and Pareto distributions to approximate the empirical distribution of product sales $F()$. To estimate the parameters of each distribution we use either the Maximum Likelihood (ML) or the Simulated

Method of Moments (SMM) approach. With the latter we target sales at rank statistics conditional on exporter scope. (The targeted moments are in Figures 2(a) and 2(b).) The ML approach is a more stringent test of the model since we do not directly target the moments we want to explain in the estimation. If the distribution $F()$ accurately describes the data on sales, both approaches should yield similar results. This is indeed the case for the log-normal distribution but not for the Pareto.

3.1.1 Log-normal Distribution

If sales S_{fk} follow the log-normal distribution with the location and shape parameters μ and variance τ, the natural logarithm of sales $ln(S_{fk})$ follows the normal distribution with mean μ and variance τ. The top panel of Table 2 reports the ML and SMM estimates for the shape and location parameters of the normal distribution fitted to the data on the natural logarithm of sales. The bootstrapped standard errors are in parentheses and indicate that estimates are statistically significant. Both the ML and SMM approaches yield very similar results and both closely track the actual distribution of the log sales in Figure 1. The ML approach yields only slightly thicker tails due to higher estimated variance. This is what one would expect if the estimated theoretical distribution approximates the empirical one reasonably well.

The predictions of the statistical model that we have set out to test involve order statistics conditional on scope, i.e., best, second best, and least selling products. The model will only be able to quantitatively match the data if

it can replicate the correlation between order statistics and exporter scope that we observe in the data. In Figure 2(a) each panel looks at average sales at rank for firms with the same scope. In panel three, for example, for each firm that exported three products in 2003 we rank the products by their contribution to its total sales at the destination. So the product with the highest contribution is assigned rank one. Along with average sales at rank and scope from the data we also report the corresponding moments simulated from the statistical model. The moments simulated at the ML and SMM estimates track the actual data remarkably well.

Table 2 – Distribution parameter estimates.

	MLE	SMM
	Log-normal	
μ	-1.6 ***	-1.6 ***
	(0.0155)	(0.00018)
σ	2.62 ***	2.12 ***
	(0.008)	(0.00024)
Log-likelihood	-104,570.4	
	Pareto	
ν	0.074 ***	0.13***
	(0.00043)	(0.0007)
α	0.000059	0.000003 ***
		(0.0000006)
Log-likelihood	-684,81.37	
	Log-normal	
	Single-product exporters sub-sample	
μ	-1.73 ***	
	(0.02)	
σ	2.62 ***	
	(0.014)	
Log-likelihood	-41,704	

Bootstrapped standard errors in the parentheses.
* p<0.10, ** p<0.05, *** p<0.01

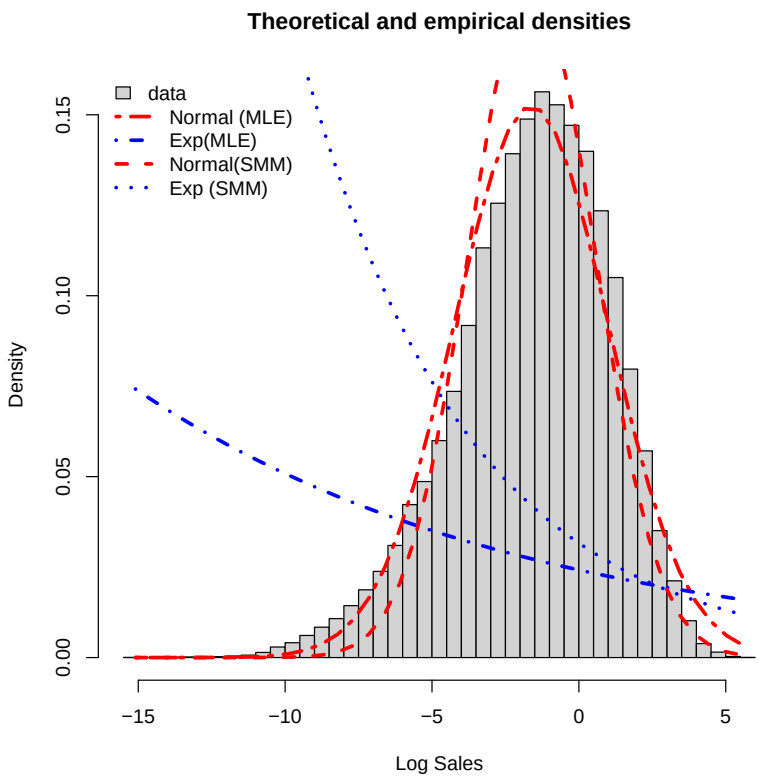
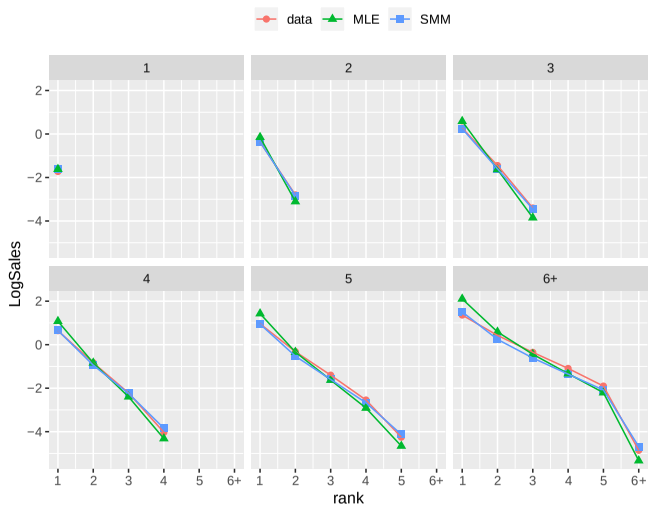
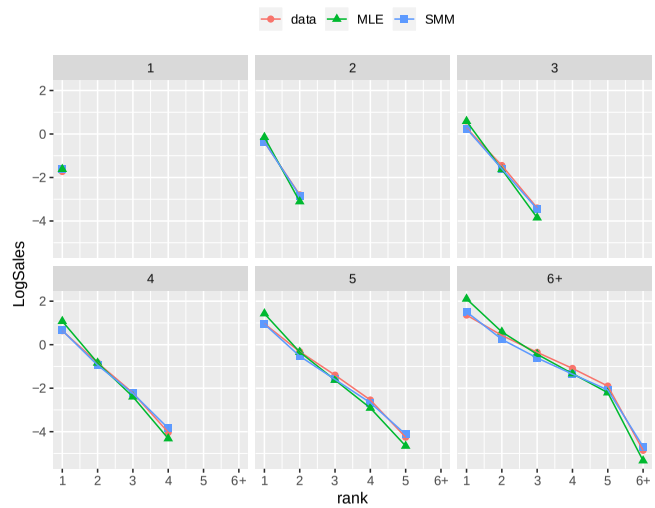


Figure 1 – The figure shows the estimated density of log-sales under the hypothesis that sales are Pareto or log-normally distributed using the Maximum Likelihood and the Simulated Method of Moments approaches against the empirical distribution of log-sales.



(a) Sales distributed log-normally.
SMM series corresponds to log-Normal distribution with mean -1.6 and variance 2.12 estimated by the SMM approach. MLE series corresponds to to log-Normal distribution with mean -1.6 and variance 2.64 estimated by the ML approach.



(b) Sales distributed Pareto.
SMM series corresponds to Pareto with location parameter 0.000059 and shape parameter 0.074 estimated by the SMM approach. MLE series corresponds to Pareto with location parameter 0.000003 and shape parameter 0.13 estimated by the ML approach.

Figure 2 – Each panel looks at average sales at rank for firms with the same scope (indicated at the top of each panel). The x-axis reports the rank of the product and the y-axis reports average sales of products conditional on product rank and exporter scope. Product rank is assigned for each product based on its contribution to total sales of the exporter at the destination. A product with the highest share of sales has rank 1. The last panel in Figures 2(a) and 2(b) shows sales at rank for firms with 6 or more products. The rank "6+" on the x-axis corresponds to the least selling product, i.e., the y-axis shows average sales in the least selling product for exporters with 6 or more products.

3.1.2 Pareto Distribution

If sales S_{fk} are Pareto distributed with a location parameter α and shape parameter ν then $ln(S_{fk}) - ln(\alpha)$ is distributed exponentially with the rate parameter ν. To estimate the parameters of the Pareto distribution with the Maximum Likelihood we first calibrate the location parameter α to the smallest sale value in the sample, $\hat{\alpha} = min(S_{fk})$, and then estimate the rate parameter ν by fitting the exponential distribution to data on $ln(S_{fk}) - ln(\hat{\alpha})$. With the SMM approach we jointly estimate α along with ν by targeting mean sales conditional on rank and scope. The second panel of Table 2 reports the estimates. The SMM and ML estimates are quite different, which indicates that the distribution is likely to be misspecified. Figure 1 compares the estimated exponential distribution with the actual distribution of log-sales and the fitted normal density. Regardless of how parameters are estimated, Pareto performs poorly in replicating the sales distribution. The poor empirical performance of the Pareto stems from its failure to approximate the left tail of the sales distribution and echoes the findings of Head et al. (2014), Bee and Schiavo (2018), and Fernandes et al. (2015) who show that the log-normal distribution provides a better fit for the entire distribution of the exporter sales than the Pareto. Similarly, in Figure 2(b) the order statistics simulated from the statistical model with sales distributed Pareto fit the data poorly. Whether we estimate the parameters of the distribution with ML or SMM, the statistical model tends to overestimate how quickly sales at rank decrease with rank at a given scope.

In the following subsection we show that the statistical model with sales-distributed log-normally quantitatively replicates the observed data patterns. If sales are distributed Pareto the statistical model predicts much bigger differences between the large and small scope exporters than we observe in the data.

3.2 Test of the Statistical Model

In this section, we compare the predictions of the statistical model to the patterns in the data and thus identify when differences between the large and small scope exporters arise due to aggregation and when these differences call for an economic explanation.

3.2.1 Prediction 1

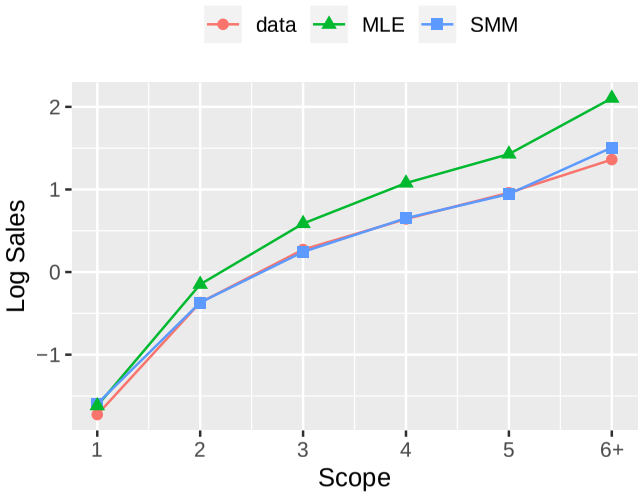
In Figures 3(a) and 3(b) we address Prediction 1 , i.e. large scope exporters sell more in their best selling products than small scope exporters. In Figure 3(a), we compare the mean sales of the best selling products as a function of exporter scope in the data to those simulated from the statistical model assuming sales are distributed log-normally. The parameters of the log-normal distribution were estimated using either the ML or the SMM approach and are given in the top panel of Table 2. In the data, exporters with six or more products sell more than six times as much as single product exporters in their top-selling products. If instead, we compare the data to the predictions of the model we see that large scope exporters sell no more in their top-selling

product than the statistical model predicts. The pattern simulated from the SMM estimates tracks the data closely[7] and shows that the differences between the large and small scope exporters can arise even if firm-product sales are iid random draws. The pattern simulated from the ML estimates, in fact, predicts that large scope exporters should sell more in their top-ranked products than we observe in the data.

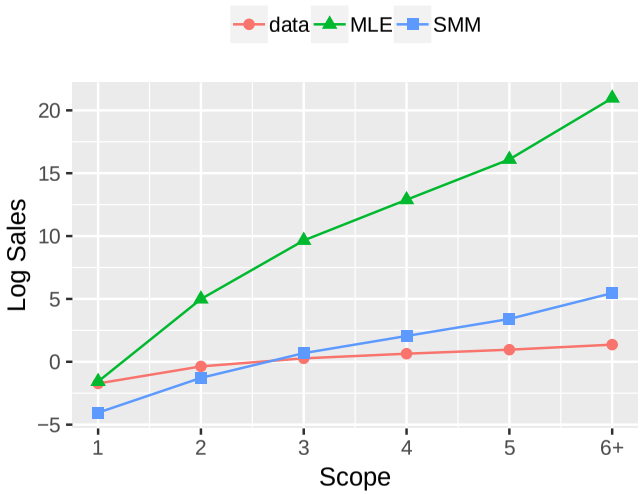
Comparing the predictions of the statistical model from the ML and SMM estimates highlights that differences between large and small scope exporters are driven by variance in sales. The SMM and ML approaches produce identical location parameter estimates, but the latter yields a bigger shape parameter σ, which implies a greater variance of sales. This illustrates that a higher variance of sales translates into bigger differences between large and small scope exporters.

Figure 3(b) is analogous to Figure 3(a) with the difference that the distribution of the firm-product sales $F()$ is set to be Pareto with parameter estimates given in the second panel of Table 2. While the moments simulated from the Pareto distribution poorly track the data, it's worth pointing out that the model still predicts that large scope exporters sell more in their top-selling products than small scope exporters. In fact, Pareto predicts much bigger differences between large and small scope exporters than we actually observe. The patterns simulated from the log-normal and Pareto both

[7] This is not surprising since we have explicitly targeted order statistics conditional on the firm scope.



(a) Sales distributed log-normally.
MLE series corresponds to the log-normal distribution with a mean of -1.6 and variance 2.12, and SMM series corresponds to the log-normal distribution with a mean of -1.6 and variance 2.62.



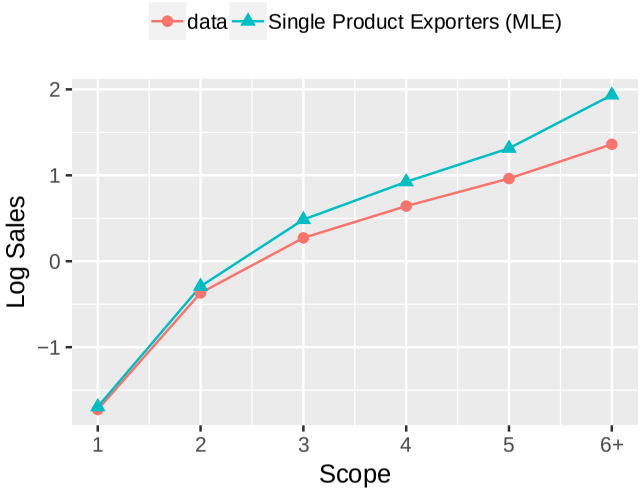
(b) Sales distributed Pareto.
SMM series corresponds to Pareto with location parameter 0.000059 and shape parameter 0.074 estimated by the SMM approach. MLE series corresponds to Pareto with location parameter 0.000003 and shape parameter 0.13 estimated with the ML approach.

Figure 3 – Average sales in the top-selling product for exporters of different scope.

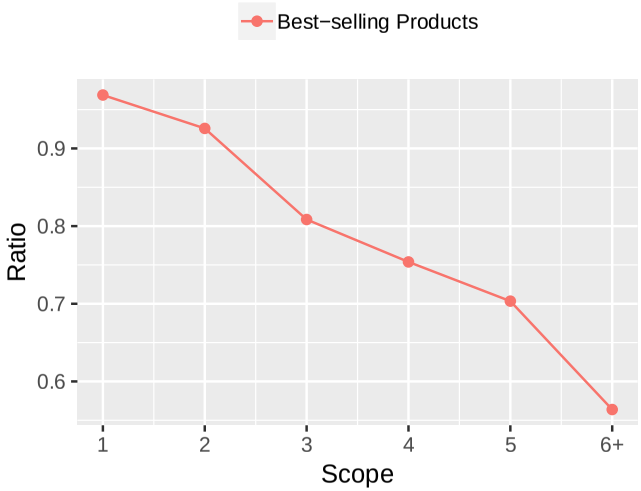
illustrate that one should be careful not to interpret the observation that large scope exporters sell more in their top-selling products than small scope exporters as evidence of large scope exporters being more productive.

To highlight this point further, we ask how much more large scope exporters would sell in their best selling products if their sales came from the same distribution as single scope exporters. To this end, using the ML estimator we estimate the location and scale parameters of the log-normal distribution from the data on single-product exporters only. The location and scale parameters are estimated to be -1.73 and 2.61, respectively. The details are in the bottom panel of Table 2. In Figure 4(a) we compare the simulated mean sales of the best selling products conditional on exporter scope to the data. Mean sales of the single product exporters in the data and the simulation are virtually identical by construction. The statistical model predicts that if larger scope exporters drew sales from the same distribution as single product exporters, they would sell more in their top selling products than we observe in the data.

To understand just how much more, in Figure 4(b) we plot the ratio of the simulated moments to the data moments. For single product exporters this ratio is one by construction but as scope increases the ratio decreases to 0.6 for exporters with six and more products. This means that if exporters with six and more products drew their sales from the same distribution as the single product exporters, they would sell almost 60 percent more in their



(a) Average sales in the top-selling product for exporters of different scope.



(b) Ratio of the mean best selling product in the data to the simulated analogue from the statistical model conditional on exporter scope.

Figure 4 – In the simulation, sales are distributed log-normally. The location and shape parameters are estimated from the data on single product exporters only.

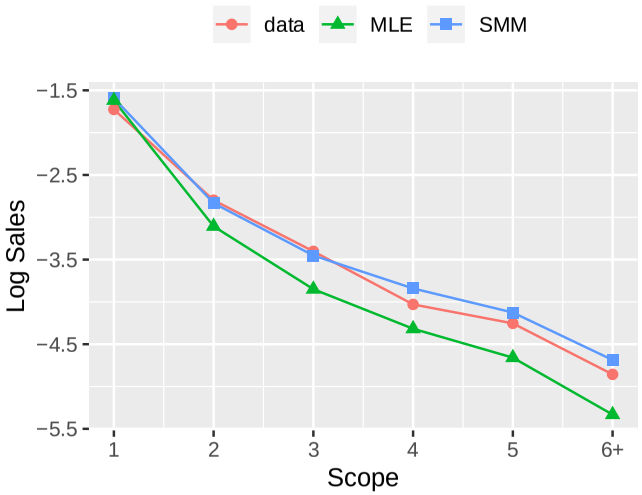
the top-selling products than we observe in the data. [8]

3.2.2 Prediction 2

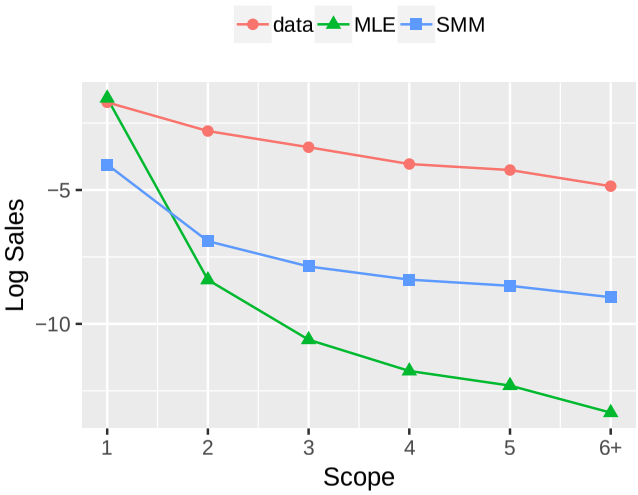
Figures 5(a) and 5(b) look into the mean sales of the least selling products by exporter scope to address Prediction 2. The two figures compare data to the moments simulated from the statistical model under the assumption that sales are log-normally (5(a)) or Pareto (5(b)) distributed. In 5(a) the moments based on the SMM estimates narrowly track the data and predict that large scope exporters sell only slightly more in their least selling products compared to the data. The ML estimates, on the other hand, predict that large scope exporters sell less in their least selling products relative to the data. Although Pareto in Figure 5(b) falls short of quantitatively replicating the data quite dramatically, it too predicts that average sales of the least selling product decrease with scope. This suggests the observation that large scope exporters sell less in their least selling products than the small scope exporters can arise because we aggregate sales across different number of products and should not be viewed as evidence that product introduction fees decrease with scope.

In Figure 6(a) we compare what we observe exporters to sell in their least selling products conditional on exporter scope to what they would sell if their sales were drawn from the log-normal distribution fitted to the data on single

[8] To be exact, the ratio between the observed and simulated average sales in the top-selling product for exporters with six or more products is 0.64. This implies that the difference between the simulated and observed sales constitutes 56 percent of the observed sales.



(a) Sales distributed log-normally.
MLE series corresponds to the log-normal distribution with a mean of -1.6 and variance 2.12, and SMM series corresponds to the log-normal distribution with a mean of -1.6 and variance 2.62.

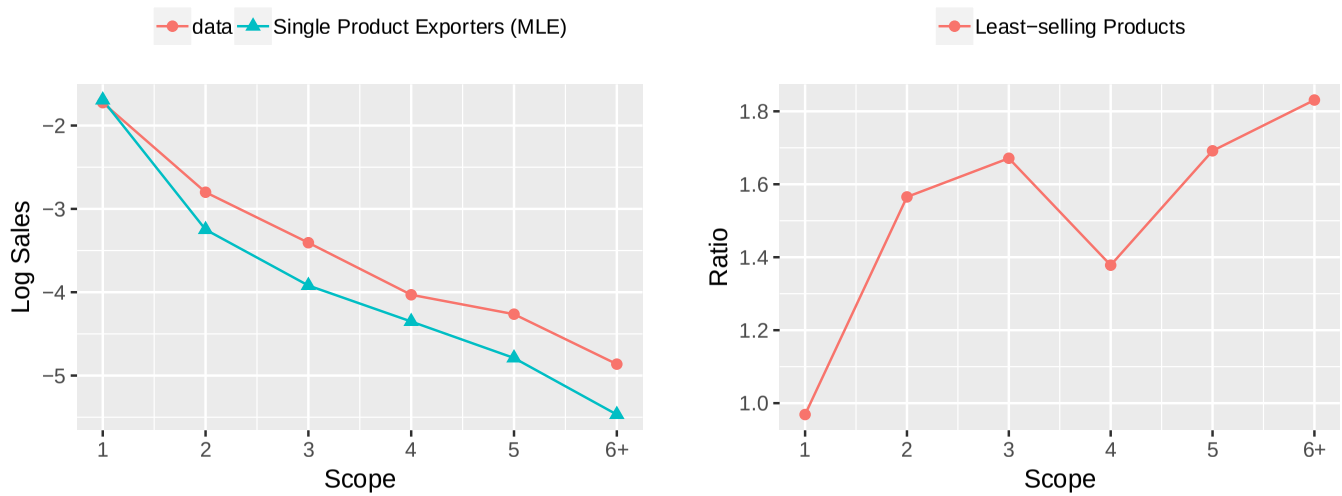


(b) Sales distributed Pareto.
SMM series corresponds to Pareto with location parameter 0.000059 and shape parameter 0.074 estimated by the SMM approach. MLE series corresponds to Pareto with location parameter 0.000003 and shape parameter 0.13 estimated with the ML approach.

Figure 5 – Average sales of the least selling products by exporter scope

product exporters only. (For the estimated parameters of the distribution see the bottom panel of Table 2). The figure shows that the differences between how much large and small scope exporters sell in their least selling products would have been even bigger if large scope exporters drew sales from the single product distribution. Figure 6(b) shows the ratio of the simulated and actual mean sales of the least selling products. In percentage terms this implies that exporters with six or more products would sell 11 percent less than in the data if their sales came from the same distribution as the sales of the single product exporters.

So far, we have shown that the relationship between the mean best/least selling products and exporter scope arises due to aggregation even in the absence of ex-ante firm heterogeneity. This ability of the statistical model to replicate the sales at rank pattern indicates that comparing sales of the best and least selling products between large and small scope exporters is misleading. In fact, comparing data to the benchmark of the statistical model highlights that large scope exporters sell no more than single product exporters in their top selling products but sell more than single product exporters in their least selling products. This is in contrast to the conclusion that one would reach by directly comparing mean sales at the highest/lowest rank for firms of different scope.



(a) Average sales in the least selling product for exporters of different scope.

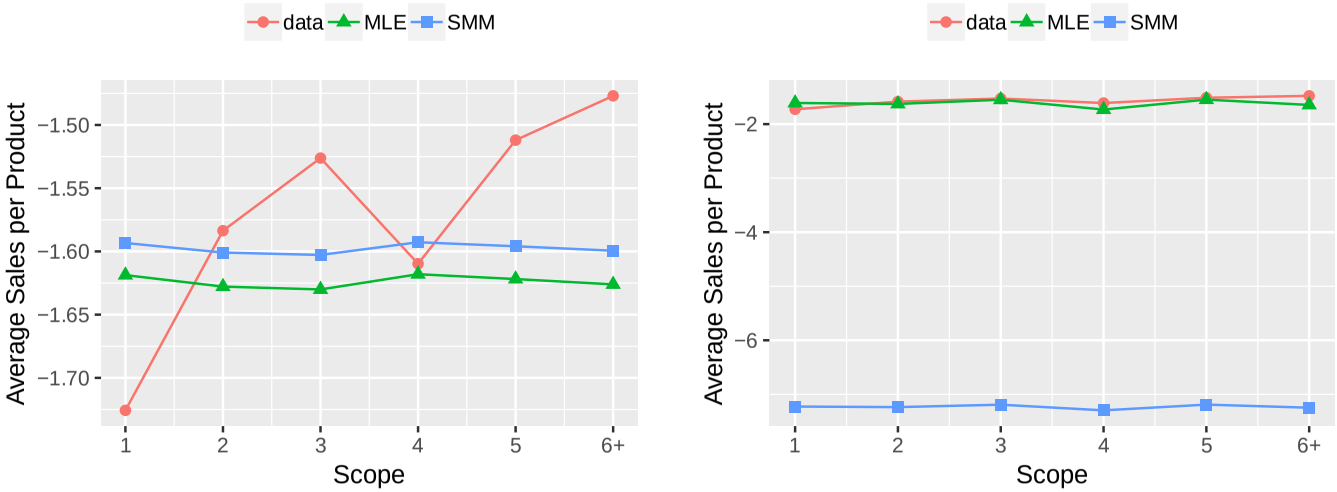
(b) Ratio of the mean least selling product in the data to the simulated analogue from the statistical model conditional on exporter scope.

Figure 6 – In the simulation, sales are distributed log-normally. The location and shape parameters are estimated from the data on single product exporters only.

3.2.3 Prediction 3

Next, we address Prediction 3 of the statistical model. In Figures 7(a) and 7(b) we compare average sales per product in the data and simulated from the statistical model with sales distributed log-normally and Pareto. In both figures average sales per product tend to increase with scope in the data and are virtually independent of scope in the simulations. This is consistent with Prediction 3 and suggests that average sales per product is a useful moment to characterize multi-product exporters. Furthermore, average sales in Figures 7(a) and 7(b) suggest that large scale exporters on average sell more than smaller scope exporters, although the relationship between scope and average sales per product is not monotone.

When sales are distributed log-normally, the statistical model matches the magnitude for average sales per product, even though it does not replicate the relationship between scope and average sales per product in the data. With Pareto distribution, average sales per product predicted by the statistical model depend on how the parameters of the distribution were estimated. Average sales per product based on the SMM estimates provide an unreasonably low estimate of average sales per product, even though the mean sales at rank simulated from the SMM estimates produced a better fit to the data than the moments simulated from the ML estimates. This further highlights that Pareto poorly fits the entire distribution of sales.



(a) Sales distributed Log-Normally.

(b) Sales distributed Pareto.

Figure 7 – Average sales per product by exporter scope.

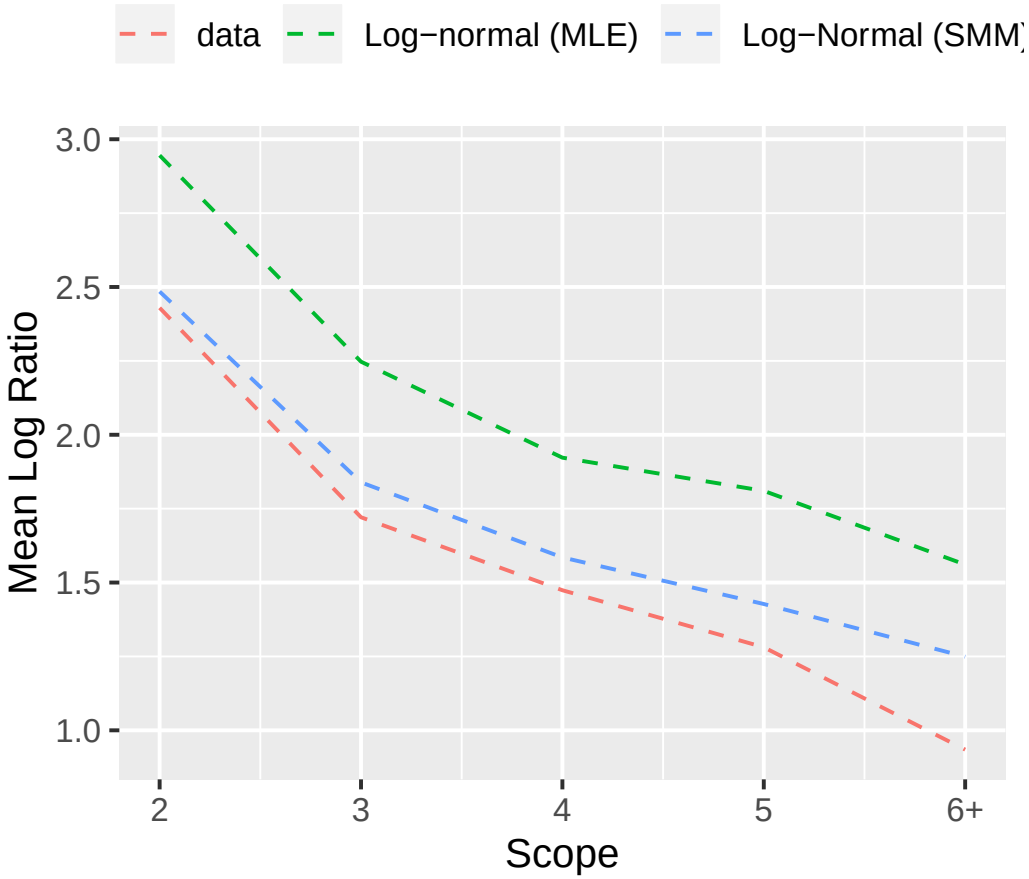


Figure 8 – Mean ratio of the best selling to the second best selling product.

3.2.4 Prediction 4

According to Prediction 4, the log ratio of sales between the best selling and the second best selling product is constant with scope when sales are distributed Pareto and decreases monotonically with scope when sales follow log-normal, Weibull or exponential distributions. In Figure 8, we show that when sales are log-normally distributed the mean of the log-ratio decreases with exporter scope both in the data and the simulation. The model misses the levels but replicates the rate of decline in the ratio with scope. Note that we do not directly target the ratio moments in the data.

The relationship between the ratio of sales and the number of products is rarely of interest in its own right. Instead, researchers use the ratio as a measure of skewness of the exporter's product mix and focus on how competition impacts it. However, to the extent that changes in the competitive environment influence the number of products firms export, some of the variation in the ratio will be driven by the changes on the extensive margin rather than exporter response on the intensive margin. In the next section, we explore when the log-ratio is a useful measure of sales concentration, and when it mechanically reflects the variation in the number of products.

4 Multi-destination Extension

In the previous section we have shown that when sales are log-normally distributed the log ratio of sales between the best selling and the second

best selling product is systematically related to the number of products a firm sells in the market: smaller scope exporters have a higher ratio. The ratio, in turn, has been used to measure how multi-product firms respond to changes in the intensity of competition. The best known example of this is Mayer et al. (2014). In their seminal paper on multi-product exporters, they show that tougher competition in an export destination shifts down the entire distribution of markups across products and induces multi-product exporters to concentrate on their core products[9]. Sales in the model are distributed Pareto so an increase in sales concentration is equivalent to an increase in the log-ratio. To test the theoretical predictions of their model Mayer et al. (2014) regress the log-ratio of sales between the best and the second best selling product on destination market size, it's foreign supply potential[10] and other variables that proxy toughness of competition in a market[11].

While Mayer et al. (2014) consider alternative measures of concentration, their preferred formulation of the model's testable prediction relies on the log-ratio. As long as firm-product sales are distributed Pareto, using log-

[9] A core product of a firm is its lowest marginal cost product.

[10] The supply potential is used to proxy for the geography of a destination that does not rely on country-level data for that destination. It is typically constructed as the aggregate predicted exports to a destination based on a bilateral trade gravity equation (in logs) with both exporter and importer fixed effects as well as the standard bilateral measures of trade barriers/enhancers. Following Mayer et al. (2014) the foreign supply potential is a related measure of a destination's foreign supply potential that does not use the importer's fixed effect when predicting aggregate exports to that destination. By construction, foreign supply potential is thus uncorrelated with the importer's fixed-effect.
The supply potential data by Head and Mayer (2011) is available online at http://www.cepii.fr/anglaisgraph/bdd/marketpotentials.htm.

[11] The regressions include country-specific random effects on firm-demeaned data.

ratio to capture the reallocation of resources across the fixed product range is a perfectly valid approach. However, this is unlikely to be the case in the data, and the log-ratio is likely to reflect variation in the number of products a firm exports to a market above and beyond the reallocation of resources across products within the firm. This is a problem because exporter scope at a destination itself varies with market toughness. Bernard et al. (2010) and Feenstra and Ma (2007), among others, document that firms export a wider product range to larger markets and trim their scope in high supply potential destinations.

Prediction 4 of the statistical model highlights that the log-ratio is related to the number of products for which it is calculated. Smaller scope exporters have a higher ratio. If firms export more products to large market size destinations then the statistical model predicts a negative correlation between market size and the log-ratio. Pro-competitive effects of a larger market that work through firms reallocating resources to lower cost (higher mark-up) products in contrast imply a positive correlation. It's even trickier with the supply potential. Firms export fewer products to high supply potential destinations, so the statistical model predicts a positive correlation between the log-ratio and the market supply potential. Pro-competitive effects captured by the high-supply potential also imply a positive correlation with the log-ratio. So, when one regresses the log-ratio on the measures of market competitiveness, the coefficients reflect a combination of product mix adjustments and product scope adjustments. To the extent that one

wants to measure the effects of competition on the reallocation of resources across a given product range, one will underestimate the effect of market size on changes in product mix and overestimate the effect of market supply potential.

Whether variation in the number of products exported to a destination is sufficient to generate statistically significant relationships between the log-ratio and the variables that proxy intensity of competition in the market is an empirical question. To answer it we extend the statistical model to the multi-country set-up. We then simulate the statistical model to obtain a data set where firm outcomes are driven by randomness alone. We then replicate the regressions in Mayer et al. (2014) to evaluate the impact of the market competitiveness measures on the log-ratio in the actual and simulated data sets. Any statistically significant results using the simulated data set will indicate that the results are driven by variation in the number of products rather than changes on the intensive margin.

We extend the statistical model to the multi-product set-up as follows. As with a single destination case, a multi-product exporter indexed by f is a collection of products K_d that a firm exports in a given year to a given destination d. Products that a firm exports to a destination are indexed by $k \in \{1, ..., K_d\}$. The number of products a firm exports to a destination d (K_d) is treated as exogenous. Product sales S_{fkd} are iid draws from the same log-normal distribution.

To estimate the parameters of the sales distribution $F()$ we use the Maxi-

Table 3 – Country Level Summary Statistics

	Mean	Std. Dev.	25%	50%	75%	95%
Mean Country Log Ratio	1.78	0.63	1.54	1.72	1.91	2.46
Mean Country Scope	2.80	0.45	2.60	2.73	2.95	3.72
Number of Exporters	584	1473	17	82	512	2627
Log GDP	23.4	2.27	21.9	23.1	25.1	27.3
Log Supply Potential	14.9	0.99	14.3	14.6	15.4	16.9
Log GDP per Capita	7.85	1.64	6.45	7.78	9.10	10.5
Log Distance	9.02	0.53	8.83	9.06	9.39	9.63
Observations	176					

The sample contains all countries to which at least one Chinese multi-product exporter sold at least two products in 2003 and includes only manufacturing producers. The first three rows describe the variation in the outcome variables across different destinations. The *Mean Country Log Ratio* is the mean log-ratio of sales of the best selling to the second best selling product at a destination. The *Mean Country Scope* is the average number of products that firms export to a destination and *Number of Exporters* is the number of firms exporting to a destination. The table also contains information on the country level variables that we will use to proxy market toughness of a destination: *Log GDP*, *Log Supply Potential*, *Log GDP per Capita* and *Log Distance*.

mum Likelihood (ML) approach where we treat sales of the Chinese exporters to 176 destinations as if they come from a single distribution.

Table 3 provides some summary statistics for the country level variables the we use. The sample includes all destinations to which at least one Chinese exporter sold at least two products in 2003 and only includes manufacturing firms involved in production of its exports. The number of multi-product exporters selling to a destination is highly skewed: the median number of exporters is only 82 while the 95th percentile is 2,627. Average exporter scope, on the other hand, is much more balanced across destinations with the 25th percentile given by 2.6 and the 95th by 3.72. The table also contains information on the determinants of the market toughness (*Log GDP*, *Log Supply Potential*, *Log GDP per Capita*, and *Log Distance*) that we will use in replicating the regression analysis in Mayer et al. (2014).

In its simplest form the multi-country extension of the model has two parameters: the location and the shape parameter that govern the distribution of sales. In the data, we have to contend with the fact that sales systematically vary with destinations and product categories (6-digit HS categories). To take this into account, we purge the sales of the destination-product category fixed effects. We do this by first estimating Equation 1

$$lnS_{fkd} = \gamma_{kd} + \epsilon_{fkd} \tag{1}$$

where the dependent variable is the log of sales and γ_{kd} captures the destination-product category fixed effects.

We then use the residual ϵ_{fkd} to estimate the parameters of sales distribution $F()$. Using the ML approach and the assumption that ϵ_{fkd} is normally distributed, we estimate the variance of 2.14 with a standard error of 0.002 and the mean equal to 0 by construction. Figure 9 compares the estimated theoretical distribution and the actual distribution of the log-sales. While the log-normal distribution slightly underestimates the thickness of the left tail, it is otherwise able to approximate the empirical distribution of log sales fairly accurately. [12] We use the estimated distribution to simulate sales data for each exporter at a destination which we use to study the relationship between the simulated log-ratio and country-level variables that measure competition at the destination.

[12] Using raw data on log sales does not change the estimate of the variance. The results can be found in the online appendix.

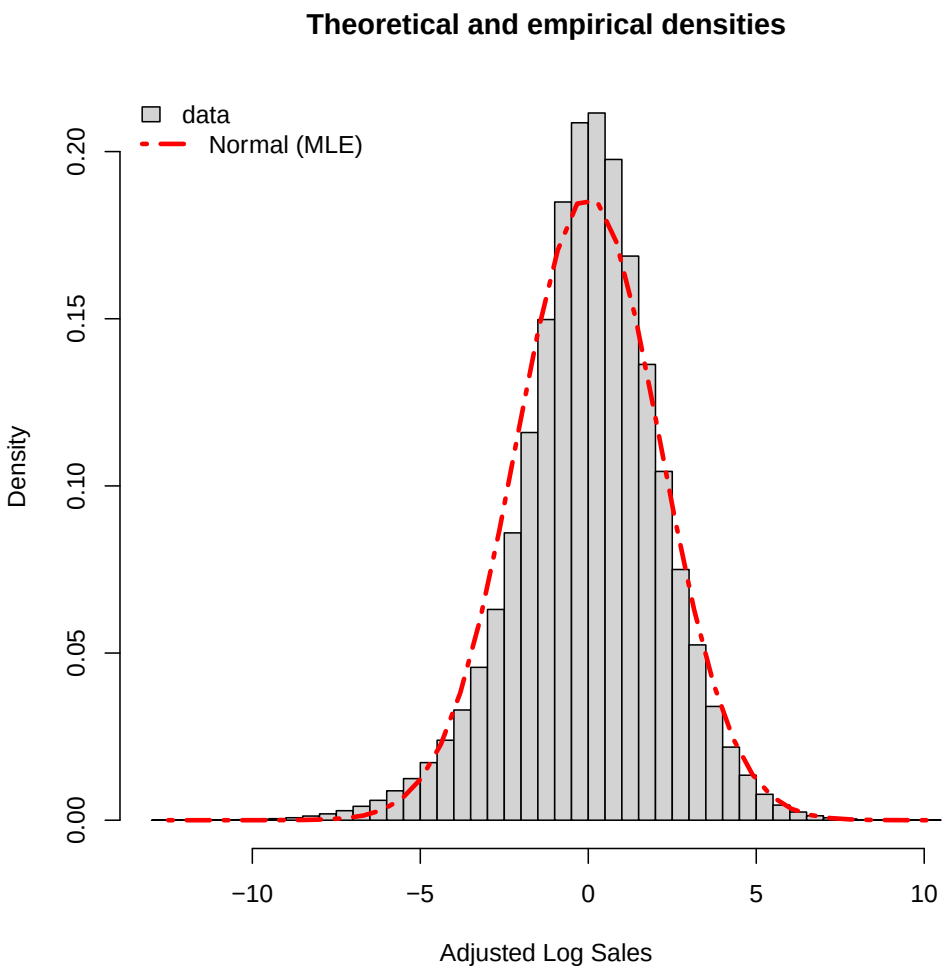


Figure 9 – The density of firm-product-destination log sales and the estimated density $N(0, 2.14)$.

We then use the residual ϵ_{fkd} to estimate the parameters of sales distribution $F()$. Using the ML approach and the assumption that ϵ_{fkd} is normally distributed, we estimate the variance of 2.14 with a standard error of 0.002 and the mean equal to 0 by construction. Figure 9 compares the estimated theoretical distribution and the actual distribution of the log-sales. While the log-normal distribution slightly underestimates the thickness of the left tail, it is otherwise able to approximate the empirical distribution of log sales fairly accurately. [13] We use the estimated distribution to simulate sales data for each exporter at a destination which we use to study the relationship between the simulated log-ratio and country-level variables that measure competition at the destination.

Before we make use of the simulated data, we establish the relationship between the ratio and the various measures of market toughness in the Chinese data by replicating the regressions from Table 2 in Mayer et al. (2014) in columns (1) and (5) in Table 4. In column (1), we look only at the effect of market size and foreign supply potential. In column (5), we also include geographic barriers. Only the *log of GDP* variable is highly significant across regressions. In magnitude, the coefficient on the *Log GDP* variable is similar to what Mayer et al. (2014) find for the French exporters. To the extent that GDP captures the intensity of competition, we observe that the effect on the log-ratio is positive - firms concentrate their resources in their best selling

[13] Using raw data on log sales does not change the estimate of the variance. The results can be found in the online appendix.

products in tougher markets. The foreign supply potential that is significant in French data is not significant in any of the formulations in the Chinese data.

Next, we reproduce the regressions with the simulated data in columns (3) and (7). In the simulated data, the product sales are just iid draws from the log-normal distribution. Hence any effect of market competitiveness measures that we find is driven by aggregating data on sales across the different number of products to the firm level. If a lot of small scope exporters sell at a destination, the ratio will be mechanically higher than if the destination was dominated by large scope exporters. Both in columns (3) and (7) the coefficient on GDP is negative and significant, reflecting that large market size is associated with firms exporting more products, which mechanically lowers the log-ratio. The foreign supply potential variable is positive in both regressions. Interestingly it becomes significant in the simulated data when we control for bilateral trade barriers between China and its trading partners in column (7). This is because including both geographic barriers and supply potential helps differentiate between good geography and distance as many high supply potential markets are located far away from China[14]. Distance and contiguity in column (7) are also statistically significant. The log of distance enters with a negative sign, while contiguity has a positive effect.

[14]In the French data that Mayer et al. (2014) use distance from France is highly correlated with good geography and hence a high supply potential for that destination. The correlation between the log-distance and log-supply potential is 78 percent. This is not the case in the Chinese data, the correlation between the foreign supply potential and distance in logs is negative and relatively small (-0.2191).

These signs are consistent with firms exporting a narrow range of products to faraway destinations and a wider range to countries nearby. To check that the variation in the simulated data is indeed driven by the variation in scope, we use the number of products as the dependent variable in columns (4) and (8) respectively. The signs on the independent variables are consistent with variation in scope being responsible for the results in the regressions with the simulated data.

Finally, we introduce the scope variable as an additional control in the regressions with the data on Chinese exporters. In columns (2) and (6) we regress the log-ratio on exporter scope along with the measures of market competitiveness. In both cases the coefficient on the scope variable is negative and statistically significant. What is more interesting is that the inclusion of the scope variable changes the magnitude of the market size coefficient. Comparing the regressions in column (1) and (2,) the coefficient on the GDP variable increased from 0.035 to 0.05, which is a 30 percent increase. Similarly, the coefficient on the GDP variable increases from 0.034 in column (5) to 0.05 in column (6), an increase of 32 percent. This shows that the coefficients on market size in columns (1) and (2) underestimate the effect of market size on the reallocation of sales across products because it also captures variation in the number of products exported to the destination. This further indicates that the effect of market size in Mayer et al. (2014) could have been underestimated by as much as 30 percent. The coefficient on the supply potential variable is still statistically insignificant, but changes the

sign from positive to negative again implying that in the original formulation the coefficient underestimates the negative effect of the supply potential on product mix. Note, however, that the scope and product mix (which we approximate with the log-ratio) are likely to be simultaneously determined so using scope as an independent variable is not econometrically correct.

Table 4 – Country Level Summary Statistics

	Log Ratio			Scope	Log Ratio			Scope
	Data		Simulation		Data		Simulation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP	0.0347***	0.0505***	-0.0443***	0.189***	0.0338***	0.0509***	-0.0479***	0.209***
	(10.40)	(14.54)	(-7.42)	(6.05)	(8.24)	(15.15)	(-7.80)	(9.34)
Supply Potential	0.00124	-0.00195	0.00912	-0.0490	0.00612	-0.00488	0.0225*	-0.118***
	(0.11)	(-0.21)	(0.86)	(-1.53)	(0.47)	(-0.45)	(2.12)	(-4.19)
Scope		-0.0749***				-0.0752***		
		(-6.97)				(-6.96)		
Distance					0.0331**	-0.00206	0.0617***	-0.430***
					(2.96)	(-0.17)	(4.10)	(-4.43)
Contiguity					0.000720	0.0329*	-0.0901*	0.338
					(0.04)	(2.04)	(-2.45)	(1.37)
GATT					0.00971	0.0377	-0.0554	0.272*
					(0.26)	(1.01)	(-1.46)	(2.06)
Observations	102758	102758	102758	102758	102758	102758	102758	102758
Within R^2	0.000770	0.0148	0.00105	0.0116	0.000883	0.0148	0.00156	0.0252
R^2	0.00193	0.0160	0.00214	0.0198	0.00213	0.0161	0.00281	0.0411

Notes: All columns use Wooldridge's (2006) procedure with country-specific random effects on firm and HS2 demeaned data, with a robust covariance matrix estimation. t-statistics in parentheses.

In Table 4 the regressions estimated from the simulated data highlight that variation in the number of products can drive variation in the log-ratio even when firms don't optimize on the intensive margin. As such, the log-ratio may be less than an ideal measure of changes in sales concentration in response to changes in competition. We have shown that the coefficient on the market size is biased downward, while the coefficient on the foreign supply potential is biased upward. This has implications for the measures of

economic significance of the product mix adjustment mechanism. The GDP coefficient in the baseline regression can be interpreted as the average elasticity of the log-ratio with respect to the destination's GDP. Mayer et al. (2014) use this elasticity to estimate that the within-firm resource reallocation on the intensive margin contributes 19 percent to the aggregate productivity growth. Our results suggest that the true effect of GDP on skewness measured by the log ratio of sales is larger than the baseline regression estimates, and would imply that Mayer et al. (2014) underestimate the contribution of the product mix channel to aggregate productivity growth.

More generally, the results of this section illustrate that the log-ratio, as well as other measures that are based on order statistics, should be treated with care in the multi-product firm analysis as they systematically vary with the number of products per firm.

5 Conclusion

In this paper we propose a stark statistical model of multi-product exporters to help separate which well-documented facts about multi-product exporters are genuinely informative about the economic forces behind the observed outcomes, and which ones arise as an artifact of aggregating data across a different number of products. Our results show that patterns that rely on order statistics should be treated with caution. While large scope exporters sell more in their top selling products than small scope exporters, these

differences can be reproduced with a model where pure chance drives exporter sales per product. In fact, if large scope exporters drew sales from the same distribution as the single product exporters, the differences between the large and small scope exporters would be even bigger. These differences will be bigger the bigger the variance of firm-product sales. Hence, one should be cautious when interpreting the observation that large scope exporters sell more in their top selling products than small scope exporters as evidence of large scope exporters being more productive. Similarly, differences between how much large and small scope exporters sell in their least selling products can be rationalized by randomness and aggregation and so should not be used as evidence for large scope exporters benefiting from lower product introduction fees.

In the context of multi-product exporters we show that using order statistics as dependent variables should be treated with caution as well. For example, we show that the ratio of the best selling to the second best selling product that is often used as a measure of exporter product mix systematically varies with scope. So, when used as a dependent variable to measure changes in the product mix in response to variation in measures of market toughness, it may produce biased results. This occurs when independent variables proxying market toughness are correlated with exporter scope.

Ultimately, in this paper we have shown that comparing outcomes between the large and small scope exporters when it involves aggregating product level data to the firm level may be misleading when order statistics are

involved. Our insight is that this process of aggregation when combined with pure random chance itself gives rise to differences between large and small scope exporters.

References

Albornoz, F., H. F. C. Pardo, G. Corcos, and E. Ornelas (2012): "Sequential exporting," *Journal of International Economics*, 88, 17 – 31.

Allanson, P. and C. Montagna (2005): "Multiproduct firms and market structure: An explorative application to the product life cycle," *International Journal of Industrial Organization*, 23, 587–97.

Arkolakis, C., S. Ganapati, and M.-A. Muendler (2020): "The Extensive Margin of Exporting Products: A Firm-level Analysis," *American Economic Journal: Macro*, fothcoming.

Armenter, R. and M. Koren (2014): "A Balls-and-Bins Model of Trade," *American Economic Review*, 104, 2127–51.

Bee, M. and S. Schiavo (2018): "Powerless: gains from trade when firm productivity is not Pareto distributed," *Review of World Economics*, 154, 15–45.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-

product firms and product switching," *American Economic Review*, 100, 70–97.

——— (2011): "Multiproduct firms and trade liberalization," *The Quarterly Journal of Economics*, 126, 1271–1318.

Boehm, J., S. Dhingra, and J. Morrow (2019): "The Comparative Advantage of Firms," Tech. rep., CEPR Discussion Paper No. DP13699 Available at SSRN: https://ssrn.com/abstract=3379615.

De Loecker, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79, 1407–51.

Dhingra, S. (2013): "Trading Away Wide Brands for Cheap Brands," *The American Economic Review*, 103, 2554–84.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2017): "Multi product firms, import competition, and the evolution of firm-product technical efficiencies," Tech. rep., National Bureau of Economic Research.

Eckel, C., L. Iacovone, B. Javorcik, and J. P. Neary (2015): "Multi-product firms at home and away: Cost- versus quality-based competence," *Journal of International Economics*, 95, 216–32.

Eckel, C. and J. P. Neary (2010): "Multi-product firms and flexible manufacturing in the global economy," *The Review of Economic Studies*, 77, 188–217.

Ellison, G. and E. L. Glaeser (1997): "Geographic concentration in U.S. manufacturing industries: a dartboard approach," *Journal of Political Economy*, 105, 889–927.

Feenstra, R. and H. Ma (2007): "Optimal choice of product scope for multiproduct firms under monopolistic competition," Tech. rep., National Bureau of Economic Research.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, M. D. Pierola, and A. Rodriguez-Clare (2015): "The intensive margin in trade: moving beyond Pareto," *Manuscript, Stanford University.*

Head, K. and T. Mayer (2011): "Gravity, market potential and economic development," *Journal of Economic Geography*, 11, 281–94.

Head, K., T. Mayer, and M. Thoenig (2014): "Welfare and trade without Pareto," *American Economic Review*, 104, 310–16.

Helpman, E. (1985): "Multinational Corporations and Trade Structure," *The Review of Economic Studies*, 52, 443–57.

Iacovone, L. and B. S. Javorcik (2010): "Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries," *The Economic Journal*, 120, 481–99.

Ju, J. (2003): "Oligopolistic Competition, Technology Innovation, and Multiproduct Firms," *Review of International Economics*, 11, 346–59.

Klette, T. J. and S. Kortum (2004): "Innovating firms and aggregate innovation," *Journal of political economy*, 112, 986–1018.

Macedoni, L. and M. J. Xu (2020): "Flexibility and Productivity: Towards the Understanding of Firm Heterogeneity," .

Malik, H. J. and R. Trudel (1982): "Probability density function of quotient of order statistics from the areto, power and weibull distributions," *Communications in Statistics-Theory and Methods*, 11, 801–14.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014): "Market size, competition, and the product mix of exporters," *American Economic Review*, 104, 495–536.

Nocke, V. and S. Yeaple (2015): "GLOBALIZATION AND MULTI-PRODUCT FIRMS," *International Economic Review*, 55.

Qiu, L. D. and W. Zhou (2013): "Multiproduct firms and scope adjustment in globalization," *Journal of International Economics*, 91, 142–153.

Roberts, M. J., D. Yi Xu, X. Fan, and S. Zhang (2018): "The role of firm factors in demand, cost, and export market selection for Chinese footwear producers," *The Review of Economic Studies*, 85, 2429–61.

Timoshenko, O. A. (2015): "Product switching in a model of learning," *Journal of International Economics*, 95, 233–49.

Topalova, P. and A. Khandelwal (2011): "Trade liberalization and firm productivity: The case of India," *Review of economics and statistics*, 93, 995–1009.

6 Appendix

A Proof of Prediction 4

A.1 Pareto distribution

Let S_{fk} be iid draws from the Pareto distribution with cdf:

$$F(S_{fk}) = 1 - \left(\frac{S_{fk}}{a}\right)^{-v} \tag{2}$$

where a is the location parameter such that S_{fk} and $a > 0$, $v > 0$ is the shape parameter. Let $X_{j:n}$ and $X_{i:n}$ be the ith and the jth order statistic from the random sample of size n from $F(S_{fk})$. With $i < j$ the ratio $Z_{i:j} = \frac{X_{i:n}}{X_{j:n}}$. Using the inverse Mellin's transform Malik and Trudel (1982) show that the distribution of the ratio is

$$h(z_{i:j}) = \frac{v z_{i:j}^{v+vn-vj-1}}{B(j-i, n-j+1)} * (1 - z_{i:j}^{v})^{j-i-1} \tag{3}$$

where $0 \leq z \leq 1$ and $0 < i < j \leq n$.

Letting $j = n$ and the $i = n - 1$, the expression in 3 reduces to

$$h(z_{n-1:n}) = v z_{n-1:n}^{v-1} \tag{4}$$

The expression $h_{n-1,n}(z)$ is independent of the sample size, and so will be the expected value of $E[ln(1/z_{i:j})]$

A.2 Weibull distribution

Let S_{fk} be iid draws from the Weibul distribution with cdf:

$$F(S_{fk}) = 1 - e^{-\frac{S_{fk}^{\alpha}}{\theta}} \tag{5}$$

where $S_{fk} > 0$, $\alpha > 0$, $\theta > 0$. Let $X_{j:n}$ and $X_{i:n}$ be the ith and the jth order statistic from the random sample of size n from $F(S_{fk})$. With $i < j \leq n$ the ratio $Z_{i:j} = \frac{X_{i:n}}{X_{j:n}}$. Using the inverse Mellin's transform Malik and Trudel (1982) show that the distribution of the ratio is

$$h(z_{i:j}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \sum_{r=0}^{j-i-1} \sum_{s=0}^{i-1} (-1)^{r+s} \frac{\alpha z^{\alpha-1}}{[(n-j+r+1)+(j-i-r+s)z^{\alpha}]^2} \tag{6}$$

where $0 \leq z \leq 1$ and $0 < i < j \leq n$.

Letting $j = n$ and the $i = n - 1$, the expression in 6 reduces to:

$$h(z_{n-1:n}) = n(n-1)\sum_{s=0}^{n-2}(-1)^s\binom{n-2}{s}\frac{\alpha z^{\alpha-1}}{[1+(s+1)z^{\alpha}]^2} \tag{7}$$

The expression $h_{n-1,n}(z)$ depends on the size of the sample from which the ordered statistics are calculated, and so will be the expected value of $E[ln(1/z_{i:j})]$. The expected value of $E[ln(1/z_{i:j})] = E[ln(1/z_{i:j})]$

$$\begin{aligned}
E[ln(1/z_{n-1:n})] &= \int_0^1 ln\left(\frac{1}{z}\right) * h(z_{n-1:n})dz_{n-1:n} \\
&= \int_0^1 ln(1) * h(z_{n-1:n})dz_{n-1:n} - \int_0^1 ln(z) * h(z_{n-1:n})dz_{n-1:n} \\
&= -\int_0^1 ln(z) * h(z_{n-1:n})dz_{n-1:n} \\
&= -\int_0^1 ln(z)\left(n(n-1)\sum_{s=0}^{n-2}(-1)^s\binom{n-2}{s}\frac{\alpha z^{\alpha-1}}{[1+(s+1)z^{\alpha}]^2}\right)dz_{n-1:n}
\end{aligned}$$

Interchanging summation and integration order obtain:

$$-n(n-1)\sum_{s=0}^{n-2}(-1)^s\binom{n-2}{s}\left(\int_0^1 \frac{ln(z)\alpha z^{\alpha-1}}{[1+(s+1)z^{\alpha}]^2}dz_{n-1:n}\right)$$

The integral

$$\int_0^1 \frac{ln(z)\alpha z^{\alpha-1}}{[1+(i+1)z^{\alpha}]^2}dz_{n-1:n} = -\frac{ln(2+s)}{(s+1)\alpha}$$

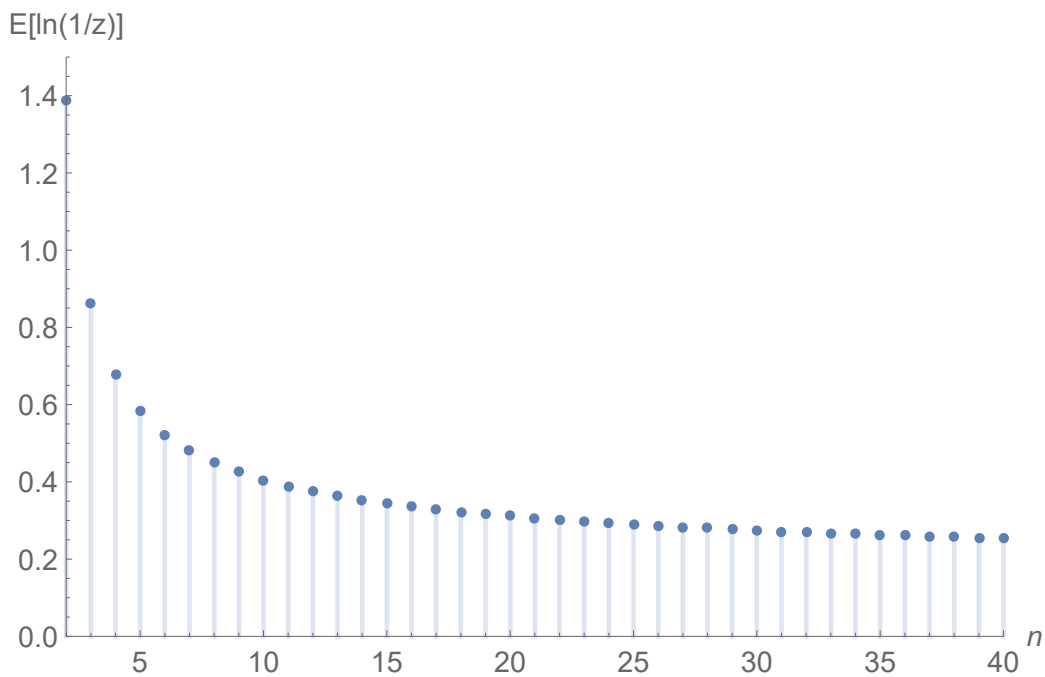


Figure 10 – $E[ln(X_{n:n} - ln(X_{n-1:n})]$ as a function of the size of the ordered sample drawn from Weibull distribution

$$
\begin{aligned}
E[ln(1/z_{n-1:n})] &= -n(n-1)\sum_{s=0}^{n-2}(-1)^i\binom{n-2}{s}\left(-\frac{ln(2+s)}{(s+1)\alpha}\right) \\
&= \frac{1}{\alpha}\sum_{s=0}^{n-2}(-1)^i\binom{n}{s}\frac{ln(2+s)}{(s+1)}
\end{aligned}
$$

The expected value of the ratio drawn from the Weibull distribution is given by the oscillating sum and in general cannot be shown to be monotonically decreasing or increasing in the size of the sample n. However for small value of n it can be shown to be monotonically decreasing in n.

A.3 Log-normal distribution

Let S_{fk} be iid draws from the log-normal distribution with cdf:

$$F(S_{fk}) = \frac{1}{2} + \frac{1}{2} Erf\left(\frac{-(lnS_{fk} - \mu)^2}{2}\sigma^2\right) \tag{8}$$

where $S_{fk} > 0$, $-\infty \leq \mu \leq \infty$, $\sigma > 0$. $X_{j:n}$ and $X_{i:n}$ are the ith and the jth order statistic from the random sample of size n from $F(S_{fk})$. With $i < j \leq n$ the ratio $Z_{i:j} = \frac{X_{i:n}}{X_{j:n}}$. In the case of log-normal distribution it is not possible to derive the distribution of the ratio for an arbitrary n. Instead we adopt numerical integration approach and calculate the expected log-ratio for various values of μ and n.

The joint distribution of two order statistics $X_{j:n}$ and $X_{i:n}$ such that $0 < X_{i:n} \leq X_{j:n} < \infty$ is given by

$$g(X_{j:n}, X_{i:n}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(X_{i:n}) \left[F(X_{j:n}) - F(X_{i:n})\right]^{j-i-1}$$
$$\left[1 - F(X_{j:n}\right]^{n-j} f(X_{j:n}) f(X_{i:n})$$

Letting $j = n$ and $i = n - 1$ obtain

$$g(X_{n-1:n}, X_{n:n}) = n(n-1)F^{n-2}(X_{n-1:n}) f(X_{n:n}) f(X_{n-1:n})$$

The expected value of $E\left[ln\left(\frac{X_{j:n}}{X_{i:n}}\right)\right]$ or $E\left[ln(X_{j:n}) - ln(X_{i:n})\right]$

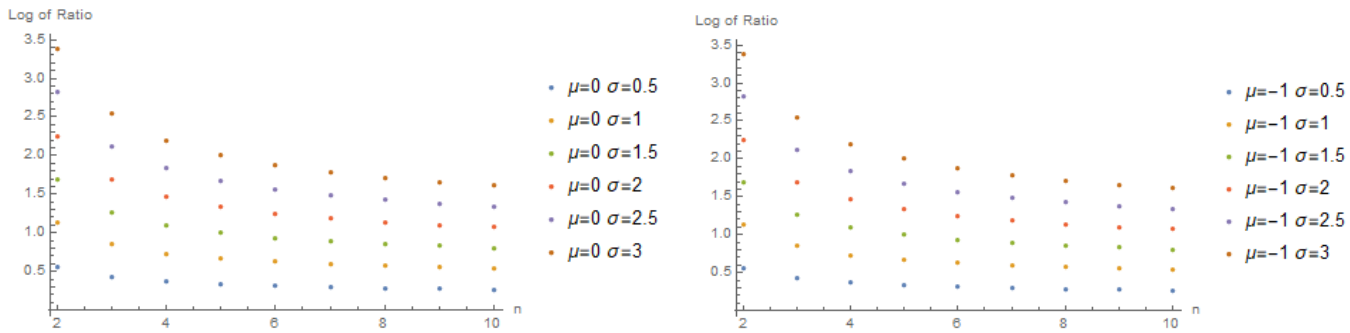


Figure 11 – $E[ln(X_{n:n} - ln(X_{n-1:n})]$ as a function of the size of the ordered sample drawn from log-normal distribution.

$$E\left[ln\left(\frac{X_{j:n}}{X_{i:n}}\right)\right] = \int_0^\infty \int_{X_{n:n}}^\infty \left(ln(X_{n:n}) - ln(X_{n-1:n})\right) g(X_{n:n}, X_{n-1:n}) dX_{n:n} dX_{n-1:n}$$

While the integral is hard to evaluate analytically we show results of numerical integration in the table below. The expected value decreases with the size of the ordered sample. The decrease is particularly important for small samples.

B Quantifying the Fit

To quantify the fit of the statistical model we use a modified version of R^2 which measures its ability to explain the sales at rank pattern. Let r^D_{kK} be the expected value of sales of a product with rank k for firms with K products calculated from the data; let r^S_{kK} be the simulated analogue, and let $\bar{r}^D$ be the simple mean of sales in the data. G is given in Equation 9

$$G = 1 - \frac{\sum_{k,K}\left(r^D_{kK} - r^S_{kK}\right)^2}{\sum_{k,K}\left(r^D_{kK} - \bar{r}^D\right)^2} \tag{9}$$

and it measures the share of variation in the order statistics explained by the statistical model relative to the share of variation explained by the simple mean. The values of G for the log-normal distribution and Pareto respectively are shown in Table 5. The statistical model when sales are log-normally distributed accounts for 99.3 percent of variation in order statistics

with ML estimates and 99.9 with SMM. Pareto distribution, in contrast, performs worse than a simple average of sales with the SMM estimate and accounts for 83.2 percent with ML estimates. The results show that if sales are distributed log-normally then the statistical model replicates the variation in sales at rank much better than if sales are Pareto distributed.

Table 5 – The value of G for simulated method of moments and maximum likelihood estimators for the log-normal and Pareto distributions.

	MLE	SMM
Pareto	0.832	-1.919
Log-normal	0.993	0.999