The EMG Distribution and Trade Elasticities*

Erick Sager †
Bureau of Labor Statistics
sager.erick@bls.gov

Olga A. Timoshenko ‡
George Washington University
timoshenko@gwu.edu

April 6, 2017

Abstract

The extensive margin of trade elasticities is negligibly small. This paper demonstrates that by omitting small firms, either by selecting a distribution that does not fully match export sales distributions or by using exogenously censored trade data, trade models overstate the magnitude of extensive margin adjustment. This paper shows that the Double Exponentially Modified Gaussian (EMG) distribution parsimoniously captures novel forms of asymmetry within export sales distributions. We provide quantitative evidence that less accurate distributions such, as the log-Normal or Pareto, generate biased trade elasticities. We further find an upward bias in the extensive margin elasticity when data are left-truncated.

Keywords: Trade elasticities, firm size distribution, extensive margin, Exponentially Modified Gaussian distribution

JEL Codes: F12, L11

*The authors thank Mina Kim, Illenin Kondo, Logan Lewis, Toshi Mukoyama, Andrés Rodríguez-Clare, and seminar participants at the Census Bureau, Midwest Trade at Purdue and the GWU Trade Study Group for insightful comments and discussions. We thank Ana Fernandes and the team in charge of implementing the project on "Export Growth and Dynamics" (Trade and International Integration Unit, Development Research Group of the World Bank) for the assistance with the World Bank Exporter Dynamics Database. The views expressed herein are those of the authors and not necessarily those of the Bureau of Labor Statistics or the US Department of Labor. The work was supported in part by the facilities and staff of the George Washington University Colonial One High Performance Computing Initiative. First version: November 2014.

†2 Massachusetts Avenue NE, Postal Square Building, Suite 3105, Washington DC, 20212, United States. E-mail: sager.erick@bls.gov

‡Corresponding author, The George Washington University, Department of Economics and The Elliott School of International Affairs, 2115 G St., NW, Washington DC, 20052, United States. E-mail: timoshenko@gwu.edu

1 Introduction

Recent advances in new trade theory, most notably due to Melitz (2003) and Chaney (2008), establish a central role for firm-level heterogeneity in accounting for the gains from trade. The theory predicts that reductions in trade costs yield welfare gains through two separate channels: the entry of new firms into export markets and the growth of incumbent firms. In fact, these separate channels that generate welfare gains can be characterized completely by the *trade elasticity*, defined as the partial elasticity of trade flows with respect to changes in variable trade costs. This elasticity depends crucially on the distribution that governs firm-level heterogeneity (c.f. Melitz and Redding (2015)). Accordingly, by fitting a parametric distribution to empirical firm size distributions, one can use a model to estimate the trade elasticity. However, when using theory to make such an inference about the trade elasticity from micro data, it is crucial that the parametric distribution accurately characterizes the empirical firm size distribution to which it was fit.

In this paper, we begin by documenting novel forms of asymmetry within empirical export sales distributions. We examine Brazilian export sales data that, unlike many trade datasets, do not censor small sales records. The two novel forms of asymmetry that stand out from the data are (i) substantial heterogeneity in positive and negative skewness across export sales distributions and (ii) the prevalence of power laws in the *left* tails in export sales distributions. These two features contradict the two most frequently utilized distributional assumptions in new trade models: the log-Normal and Pareto distributions.[1] The Normal distribution has zero skew and the Exponential distribution has a constant (parameter independent) skew of 2. Neither the Normal nor the Exponential distribution is sufficiently fat in the left tail to generate this empirical regularity.

In order to parsimoniously capture the prevalent asymmetry and tail fatness in the empirical sales distributions, we introduce the *Double Exponentially Modified Gaussian* (Double EMG) distribution. The Double EMG distribution's microfoundations can be traced to the

[1] Using a simple change of variables, the logarithm of the sales distribution – or *log-sales* distribution – is equivalently described by a *Normal* or an *Exponential* distribution, respectively.

literature on firm size dynamics and power laws (see Gabaix (2009) for an extensive review).[2] The Double EMG distribution is constructed as a convolution of independent Normal and Double Exponential distributions. Hence, it generalizes the most frequently used distributions in the trade literature, including the (Double) Pareto and log-Normal distributions. As such, the Double EMG distribution is a unimodal distribution that also exhibits fat, Pareto-like, tails. We fit the Double Exponentially Modified Gaussian distribution to empirical log-sales distributions across export destinations and demonstrate that the Double EMG distribution matches the micro-data better than either a Normal or an Exponential distribution alone.

The choice of a distribution alters the measurement of the extensive margin for the trade elasticity. In the Melitz (2003) model, the trade elasticity can be decomposed into the sum of an intensive and an extensive margin. The intensive margin arises from changes in incumbent firms' sales, while the extensive margin arises from the entry and exit of new firms into an export destination. We show, for any general distribution, that the intensive margin is constant, while the extensive margin is destination specific and depends on moments of the distribution.

By quantifying the extensive margin in the Melitz (2003) model using the Double EMG distribution, we find that the implied extensive margin contribution to the trade elasticity is quite small (on the order of magnitude of 10^{-5}). Accordingly, our results suggest that in response to changes in trade costs, in an average export destination nearly all trade adjustment occurs on the intensive margin, as incumbent exporters change their sales.

We further uncover two types of bias in measuring the trade elasticity. First, when less accurate distributions are used to approximate the data, the magnitude of the extensive margin contribution will be mismeasured. We refer to this phenomenon as the *distribution specification bias*. In particular, for an average destination, an Exponential distribution over-

[2] Reed and Jorgensen (2004) and Toda (2014) prove that the (Double) EMG is the endogenous steady state distribution of a Brownian motion that is subject to a Poisson process over stopping times (exits) and whose initial points are Normally distributed. Furthermore, the EMG distribution, which is one of the liming cases of the Double EMG distribution, arises naturally in models in which firms draw their productivity from a Pareto distribution, while learning about their Normally distributed product demand (see Timoshenko (2015), Arkolakis et al. (2015), and Bastos et al. (2016)).

estimates the extensive margin contribution by a factor of 10^7, while the Normal distribution underestimates the extensive margin contribution to the trade elasticity by 40%. We further find that the Normal distribution bias is larger when the empirical distribution of log-sales are the more fat tailed.

The size of the bias can be understood as follows. For a Normal distribution our estimates imply that, on average, entry and exit account for only 1 dollar out of one million dollars worth of increased trade. For the Double EMG distribution our estimates increase that magnitude to 10 dollars out of one million dollars. For the Exponential distribution, entry and exit account for about one third of the million dollar increase.

We conclude that the Normal and Exponential implied elasticity estimates provide a lower and upper bounds for the more accurate Double EMG implied partial trade elasticities. For export sales data that exhibits more fatness in the right tail, the Double EMG distribution implied elasticities are closer to the Exponentially implied elasticites and further away from the Normal implied elasticities. This means that if a Normal distribution were used to approximate the data, then the model will wrongly imply extensive margin adjustments that are close to zero when that data have fat right tails. As a consequence, the Normal distribution introduces systematically larger errors when the sales data for large firms are increasingly fat-tailed.

Second, we show that previous estimates of extensive margin elasticities are upwards biased due to using censored export data. We refer to this phenomenon as the *truncation bias*. It is not uncommon in collecting trade and customs data for official records to omit any firm-level exports below a given threshold. This threshold can vary from as low as 1,000 dollars to as high as 200,000 dollars in firm sales, in a given destination. Because our data are not censored we are able to measure the magnitude of the truncation bias generated by censored data. We create a counterfactual dataset that is censored at a modest value, and show that when data are truncated, the average contribution of the extensive margin to the trade elasticity is overestimated by a factor of 10^3.

The quantitative significance of this bias can be understood as follows. For the Double

EMG distribution fitted to a *full sample*, the average contribution of the extensive margin to trade elasticity is on the order of magnitude 10^{-5}, implying that entry and exit account for approximately 10 dollars out of one million dollars worth of increased trade. For the Double EMG distribution fitted to a *truncated sample*, the average contribution becomes 10^{-2}. Therefore, the estimates from the truncated sample imply that \$10,000 (as opposed to \$10) of the increase in trade is due to entry and exit.

More broadly our results demonstrate that by omitting small firms, either by selecting a distribution that does not fully match empirical distributions or by using exogenously censored data, the trade literature may substantially overstate the magnitude of extensive margin adjustment.

Our findings contribute to several literatures. First is the empirical literature on firm size distributions. Axtell (2001) shows that, when measured in number of workers, the right tail of the U.S. firm size distribution closely follows Zipf's law. Studying the French firm size distribution, di Giovanni, Levchenko, and Rancière (2011) provide further evidence on the estimates of the tail parameter of a Power law distribution and show that is lies close to one. Furthermore, while Cabral and Mata (2003) document a positively skewed firm size distribution over number of employees in Portuguese manufacturing firms and Bastos and Dias (2013) extend this result to total Portuguese exports, our work shows that the asymmetric nature of the data is also pronounced in the distribution of export sales by destination. In the Brazilian export data, the majority of destination-level exports are positively skewed and the degree of (positive and negative) skewness varies across destination. We demonstrate that a Double EMG distribution can match this feature of the data, while a Normal distribution is symmetric and an Exponential distribution has a constant skewness of 2.

In contrast, recent work has argued that sales distributions are not well characterized by Zipf's law. For example, Head, Mayer, and Thoenig (2014) show that a Normal distribution provides a better fit to export sales data, primarily due to its superior ability to match the left tail of export sales distributions. Furthermore, similar to this paper, Nigai (2017) shows that a mixture distribution of a log-Normal and Pareto better fits aggregated sales data. The Double EMG as a characterization of the firm size distribution, however, has advantages over both the Normal and the mixture distribution. First, neither the Normal nor the mixture distribution is capable of matching negative skewness that we document in the data. Second, the Double EMG has explicit microfoundations, unlike a mixture distribution, which is appealing on the grounds of theoretical consistency within the model.[3]

This paper also contributes to the literature on the interaction between firm-level heterogeneity and trade elasticities that determine the gains from trade. Melitz and Redding (2015) demonstrate that once firm-level heterogeneity is no longer governed by a Pareto distribution, the elasticity of trade flows with respect to variable trade costs depends on the entire sales distribution. Bas, Mayer, and Thoenig (2015) measure the magnitude of the trade elasticity under different assumptions about heterogeneity and show that the data favor a log-Normal distribution over a Pareto distribution. In this paper, we quantitatively assess the fit of the Double EMG distribution and find that the data favors the Double EMG over either the log-Normal or Pareto. This finding has implications for the trade elasticity, that we develop herein.[4]

Finally, the Double EMG distribution has also been used in various recent macroeconomic applications. Both Badel and Huggett (2014) and Heathcote and Tsujiyama (2015) use the distribution to model idiosyncratic earnings in incomplete markets models with taxation. The distribution helps capture the skewness in log-earnings distributions, as the EMG fits the cross-sectional log-earnings distribution better than a conventionally used Normal distribution. Toda and Walsh (2015) use the Double EMG to model the distribution of consumption growth in the Consumer Expenditure Survey and estimate consumption-based asset pricing models in the presence of fat-tailed consumption growth.

The rest of the paper is organized as follows. Section 2 establishes a set of stylized

[3]A further point of departure from Nigai (2017) is a focus on measuring trade elasticities. Relative to that paper and Head et al. (2014) as well, this paper focuses on the rich heterogeneity of sales distributions across destination-years.

[4] While this paper focuses on exogenous distributions governing firm-level heterogeneity, Mrázová, Neary, and Parenti (2016) focus on how preferences influence sales distributions.

facts about the properties of log-sales distributions across markets. Section 3 constructs the Double Exponentially Modified Gaussian distribution and characterizes its properties. Section 4 fits theoretical distributions to empirical export sales distributions and evaluates goodness of fit. Section 5 demonstrates how the trade elasticity depends on distributional assumptions and defines a theory-based strategy for estimating the trade elasticity using micro-level export data. Section 6 quantifies trade elasticities and documents the presence of distribution specification bias and sample truncation bias in the estimates. Section 7 shows that our results hold for an alternative export sales dataset; shows that our results are robust to sample selection, and, finally, shows that our results are robust to industry heterogeneity. Section 8 concludes. Proofs to all propositions are included in Appendix A, and Appendix B contains a full description of the heterogeneous-firm trade model that we employ.

2 Empirical Facts

In this section we present new stylized facts that describe log-sales distributions across export destinations and discuss how these facts present a puzzle for standard distributional assumptions made in trade models.

The data come from the Brazilian customs declarations collected by SECEX (*Secretaria de Comercio Exterior*).[5] The data cover the period between 1990 and 2001, and include the value of export sales at the firm-product-destination-year level. A product is defined at a six-digit Harmonized Tariff System (HS) level. We focus on exports in manufacturing products.[6] To explore properties of the distribution of export sales across destinations and years, we aggregate the data to the firm-destination-year level and focus on destination-year observations where at least 100 firms export.[7] We define an observation to be an entire

[5] See Molinaz and Muendler (2013) for a detailed description of the dataset. These data have also been recently used by Flach (2016) and Flach and Janeba (Forthcoming).

[6] Manufacturing HS codes lie in the range between 10.00.00 and 97.00.00. In an average year exports in manufacturing products account for 90.82% of total exports.

[7] To be consistent with the literature, we make two decisions on how to use the data. First, we follow the vast majority of research on measuring theoretical trade elasticities by aggregating the data to destination-year,

distribution of log-sales for a given destination in a given year.[8] The final sample consists of 847 destination-year distributions of log-sales.

Table 1 summarizes properties of log-sales distributions across destination-year observations. Each row presents a statistic and, because there is variation in these statistics across destination-year observations, each column reports a statistic's average value, median value, standard deviation, minimum value, and maximum value.

Fact 1 *Across destinations, export sales distributions are highly asymmetric.*

To describe the symmetry of log-sales distributions, we consider three different measures of skewness. The first is the standardized third moment measure of skewness. The second is nonparametric skew, which is defined as the difference between the mean and the median of a distribution divided by its standard deviation. The third is Kelly skewness defined as

$$\text{Kelly skewness} = \frac{(P90 - P50) - (P50 - P10)}{P90 - P10}, \tag{1}$$

where $P10$, $P50$, and $P90$ are the 10th, the 50th and the 90th percentiles of a distribution.[9]

Table 1 shows that the majority of log-sales distributions are asymmetric. The average of each skewness measure across destination-year observations is positive and the averages are statistically different from zero with a maximium p-value of 0.0003 across measures. Among the 847 destination-year observations, 54% have positive skewness, 71% have positive nonparametric skew, and 75% have positive Kelly skewness.

We formally confirm the asymmetry in log-sales distributions through a standard test of Normality, as described in D'Agostino et al. (1990). Based on skewness alone, the test rejects

as opposed to industry-destination-year (see Head et al. (2014), Bas et al. (2015), Nigai (2017)). Second, we follow Fernandes, Klenow, Meleshchuk, Pierola, and Rodríguez-Clare (2015) in requiring at least 100 firms be present within a destination-year observation.

[8] In Appendix E, we consider an alternative definition of an observation that controls for the industrial composition of sales within destinations.

[9] In recent research on the asymmetry of earnings growth over the business cycle using administrative data from the Social Security Administration, Guvenen, Ozkan, and Song (2014) use Kelly skewness to avoid the sensitivity of standardized moments to extreme values. Given that our dataset contains fewer observations, we utilize Kelly skewness for robustness - to better ensure that our results are not generated by a small number of extreme value observations.

normality in 31% of destination-year observations at the 10-percent significance level, 24% of observations at the 5-percent significance level, and 16% of observations at the 1-percent significance level. Based on both skewness and kurtosis, the test rejects normality in 42% of observations at the 10-percent significance level, 32% of observations at the 5-percent significance level, and 20% of observations at the 1-percent significance level.

These empirical facts, however, contradict the properties of the standard theoretical distributions employed by new trade models. First, the Normal distribution is symmetric, and therefore all three skewness measures equal zero for a Normal. Second, the Exponential distribution has skewness of 2, nonparametric skewness of 0.31 and Kelly skewness of 0.47. Across each of the three measures, the Exponential distribution's skewness does not depend on the distribution's parameter values. The majority of log-sales export distributions however are not symmetric, do not share the same level of skewness, and are less skewed than an Exponential distribution. Therefore, according to all three measures of skewness, neither the Normal nor the Exponential distributions characterize the export sales data well.

Fact 2 *Log-sales distributions exhibit a high degree of variation in the fatness of right and left tails.*

We focus on two measures to characterize the tail properties of log-sales distributions across destination-year observations. The first measure is kurtosis, which is the fourth standardized moment of a distribution. Kurtosis measures how much mass is located in the tails of a distribution relative to the mean. The kurtosis of a Normal distribution is constant and equals 3. A leptokurtic distribution has higher kurtosis than a Normal distribution and therefore exhibits fatter tails than a Normal. As can be seen from Panel A in Table 1, the average kurtosis across destination-year observations in the data is 3.17. Therefore, on average, the log-sales distributions are more fat-tailed than a Normal.

Similar findings hold for a percentile based measure of kurtosis defined as

$$\text{Percentile coefficent of kurtosis } = \frac{(P75 - P25)/2}{P90 - P10},$$

where $P25$ and $P75$ are the 25th and 75th percentiles of a distribution. For a Normal distribution the percentile coefficient of kurtosis is equal to 0.26. A smaller value of the coefficient corresponds to a distribution that is more kurtotic than a Normal. As can be seen from Panel B of Table 1, log-sales distributions exhibit substantial variation in kurtosis around the sample mean of 0.26 with a majority of observations being more kurtotic than a Normal.

While kurtosis is informative about the overall fatness across a distribution's tails (relative to a Normal distribution), it does not provide any information about how fat tails are relative to each other. In order to characterize fatness in the left tail relative to fatness in the right tail, we follow Gabaix and Ibragimov (2011) in estimating the right and left tail index parameters for each log-sales distribution across destination-years. Tail index parameters are estimated as the coefficient β from the following regression:

$$\log\left(Rank_i - 0.5\right) = \alpha + \beta \log(Sales_i) + \epsilon_i,$$

where i indexes firms within an export destination, $Sales_i$ is firm i's export sales, and $Rank_i$ is firm i's sales rank out of all firms exporting to a particular destination. We run this regression on a sample of firms in the top or bottom 5%, 10% and 15% of a distribution for each destination-year observation. The smaller is the estimate of the coefficient, the fatter is the corresponding tail of the distribution.

Results are summarized in Panel C of Table 1, which report information about the absolute value of the coefficient β as estimated across destination-year observations. These results indicate that the log-sales distributions exhibit substantial fatness in both the left and right tails. Depending on the sample, the average value of the tail index coefficient varies between 1.01 and 1.42. Notably, the left tail index exhibits more fatness than the right tail index. For example, the sample average of the left tail index for the bottom 15% of firms is 1.01, while for the top 15% the average is 1.08.

Furthermore, we find that both tails are simultaneously fat in a majority of cases. Figure 1 provides a scatter plot of the tail index estimates for the top and bottom 5% of firms in a distribution. Each dot in the Figure corresponds to an estimate of the right tail index (x-axis) and the left tail index (y-axis) for a given destination-year observation. Observe that both tail indexes have values below 2 for a majority of distributions.

Our finding that the right tail of the log-sales distribution tends to be fat is consistent with previous research (see Axtell (2001), di Giovanni and Levchenko (2013)). However, our finding that the *left* tail of log-sales distributions exhibits substantial fatness is, to the best of the authors' knowledge, new to the trade literature. These findings immediately imply that a Normal distribution poorly characterizes the log-sales distributions, as the Normal distribution exhibits too little kurtosis. Furthermore, the log-sales distributions would also be poorly approximated by an Exponential due to the presence of a fat left tail.

In the next section we introduce and characterize the Exponentially Modified Gaussian distribution (EMG). We show that the EMG distribution has a potential to fit the empirical distribution of log-sales better than the most prevalent distributions used in trade models. We then generalize the EMG distribution so that it can fit a larger set of the empirical observations we document above.

3 The Double Exponentially Modified Gaussian Distribution

The Double Exponentially Modified Gaussian (Double EMG) distribution is defined as a convolution of a Normal distribution and a Double Exponential distribution. As a result, one of the key properties of the distribution is its flexible behavior in the right and left tails. Hence, the Double EMG distribution is well suited to generate empirical regularities in the log-sales export data as documented in Section 2. Furthermore, the distribution arises naturally in models that feature both Double Pareto and log-Normal shocks that affect firms' profit. In Section 3.1 we derive several key properties characterizing the distribution including its behavior in the right and left tails.

3.1 Characterization of the Double EMG

Consider a random variable z defined as $z = x + y$, where x and y are two independent random variables. Assume $x \sim \mathcal{N}(\mu, \sigma^2)$ is Normally distributed, and $y \sim \mathcal{DE}(\lambda_L, \lambda_R)$, where $\mathcal{DE}$ denotes the Double Exponential distribution.[10] In this case, random variable z is a convolution of a Normal and a Double Exponential random variables and is said to follow a Double Exponentially Modified Gaussian (Double EMG) distribution with parameters $(\mu, \sigma, \lambda_L, \lambda_R)$. Proposition 1, below, formally characterizes the Double EMG distribution with its cumulative distribution function. Proposition 2 further characterizes the the Double EMG distribution's limiting properties.[11]

Proposition 1 *Let x and y be independent random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim \mathcal{DE}(\lambda_L, \lambda_R)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma > 0$, and $\lambda_L, \lambda_R > 0$. The random variable $z \equiv x + y$ has the cumulative distribution function $G : \mathbb{R} \to [0, 1]$ given by:*

$$G(z) = \Phi\left(\frac{z-\mu}{\sigma}\right) - \frac{\lambda_L}{\lambda_L + \lambda_R} e^{-\lambda_R(z-\mu) + \frac{\sigma^2}{2}\lambda_R^2} \Phi\left(\frac{z-\mu}{\sigma} - \lambda_R \sigma\right) + \frac{\lambda_R}{\lambda_L + \lambda_R} e^{\lambda_L(z-\mu) + \frac{\sigma^2}{2}\lambda_L^2} \Phi\left(-\frac{z-\mu}{\sigma} - \lambda_L \sigma\right).$$

Proposition 2 *(Limiting Results) Let z be a Double Exponentially Modified Gaussian distributed random variable with parameters $(\mu, \sigma, \lambda_L, \lambda_R)$. The random variable z is (i) an Exponentially Modified Gaussian distributed random variable as λ_L goes to infinity, (ii) an Exponentially Modified Gaussian distributed random variable with a Normal right tail and Exponential left tail as λ_R goes to infinity, (iii) a Double Exponentially distributed random variable as σ goes to zero, where if $\mu \neq 0$ then this limiting distribution is a shifted Double Exponential distribution, and (iv) an Exponentially distributed random variable as σ goes to zero and λ_L goes to infinity.*

[10] The Double Exponential distribution is also referred to as an Asymmetric Laplace distribution. The cumulative distribution function is given by $G_{\mathcal{DE}}(y) = \frac{\lambda_R}{\lambda_L+\lambda_R} e^{\lambda_L y}$ if $y < 0$, and $G_{\mathcal{DE}}(y) = \frac{\lambda_R}{\lambda_L+\lambda_R} - \frac{\lambda_L}{\lambda_L+\lambda_R}\left(1 - e^{-\lambda_R y}\right)$ if $y \geq 0$.

[11] The proofs to all propositions are included in Appendix A.

Notice from Proposition 2 that the Double EMG distribution generalizes both the Normal and the Double Exponential distributions. Consider the variance of a Double Exponential distribution with the tail parameters denoted by λ_L and λ_R, which is $\lambda_L^{-2}+\lambda_R^{-2}$. By increasing the value of both of the tail parameters we can make the variance arbitrarily small and the corresponding distribution has a point mass. Next, consider the variance of the Normal distribution. As we decrease the variance parameter σ, the Normal distribution becomes a point mass at μ. Therefore, the Double EMG distribution can be transformed into a Normal distribution when its Double Exponential distribution has zero variance ($(\lambda_L^{-2} + \lambda_R^{-2}) \to 0$) or transformed into a Double Exponential distribution when its Normal distribution has zero variance ($\sigma \to 0$).

The Double EMG distribution can further be transformed into an Exponential distribution. This occurs when the variance parameter of the Normal distribution, σ, declines to zero, and hence the Normal distribution converges to a point mass at μ. Simultaneously, when the value of the left tail parameter of the Double Exponential distribution, λ_L, goes to infinity, Double Exponential distribution converges to Exponential.

Proposition 3, below, shows that, as a consequence of being a convolution of a Normal and a Double Exponential random variable, the Double EMG distribution can generate both positive and negative skewness as well as fatness in both the left and right tails of the distribution.

Proposition 3 *If z is a Double Exponentially Modified Gaussian distributed random variable on $(-\infty, +\infty)$ then the skewness of z is given by*

$$skew(z) \quad = \quad 2\left(\frac{1}{\sigma^3\lambda_R^3} - \frac{1}{\sigma^3\lambda_L^3}\right)\left(1 + \frac{1}{\sigma^2\lambda_R^2} + \frac{1}{\sigma^2\lambda_L^2}\right)^{-\frac{3}{2}}.$$

Furthermore, the sign of $skew(z)$ is determined by the relative size of the tail parameters: (i) $skew(z) > 0$ if $\lambda_L > \lambda_R$, (ii) $skew(z) = 0$ if $\lambda_L = \lambda_R$, and (iii) $skew(z) < 0$ if $\lambda_L < \lambda_R$.

The skewness of the Double EMG distribution exhibits two stark properties. First, the distribution has a potential to generate both positive and negative skewness in the range between -2 and 2. Notably, the sign of the skewness depends on the relative fatness of the right and left tails of the distribution as measured by parameters λ_R and λ_L. Second, the distribution can be symmetric (when $\lambda_L = \lambda_R$), yet exhibit substantial deviations from a Normal in the tails when the values for λ_R and λ_L are small.

Figure 2 provides an example of two probability density functions for two distributions with zero mean and the unit variance. The solid line depicts a probability density function for a symmetric Double EMG distribution (although the tails need not be symmetric), and the dashed line depicts a probability density function for a Normal distribution. Notice that relative to a Normal, the Double EMG, while preserving the unimodal property of the distribution, has more mass in the right and left tails. This is a key distinction between the two distributions which helps the Double EMG to flexibly match novel features of log-export sales distributions.

Hence, the Double EMG is the most flexible distribution among those considered in the trade literature and has the potential to match all of the new stylized facts documented in Section 2. In the next section we describe our strategy for fitting the Double EMG distribution to the data and compare the distribution's fit to that of the Normal and Exponential.

4 Fitting to Empirical Distributions

In this section we describe our strategy for estimating distributional parameters using export sales data from Brazil. Then, equipped with estimated parameters for each destination-year log-sales distribution, we compare the fit of the Double Exponentially Modified Gaussian, Normal and Exponential distributions. We show that the Double Exponentially Modified Gaussian distribution has a superior fit to the data when compared to the Normal and Exponential distributions. Lastly, we document that there is large heterogeneity in estimated parameters and show how the estimates reflect the variation in data moments across destination-year observations.

4.1 Parameter Estimation

We choose distribution parameters so that the percentiles of the theoretical log-sales distribution match the percentiles of the empirical log-sales distribution. Specifically, we recover parameters of a theoretical distribution from non-linear quantile regressions that we implement using a generalized method of moments procedure. Our procedure is a generalization of Head, Mayer, and Thoenig (2014), who use quantile regressions to estimate parameters of the Pareto and a log-Normal distributions, both of which have linear quantile functions and therefore parameters can be estimated using linear regression. In contrast, the Double EMG distribution does not admit a linear quantile function (as can be inferred from Proposition 1) and therefore we estimate the parameters of the Double EMG distribution using a Generalized Method of Moments (GMM) procedure. For the Normal and Exponential distributions, our procedure can recover the parameter estimates implied by linear regression.

Denote by n_q the number of sales quantiles. Let q_i^d denote the i-th quantile of the empirical log-sales distribution and F_i^d denote the corresponding value of the empirical CDF at the i-th quantile.[12] By comparison, let $q_i(\Theta)$ denote the i-th quantile of the theoretical cumulative distribution function with parameters Θ and let $F(q_i|\Theta)$ denote the corresponding value of the theoretical cumulative distribution function at the i-th quantile.

For an arbitrary distribution over log-sales, we can recover the theoretical quantiles by inverting the theoretical cumulative distribution function. Generally, the inverse can be computed numerically for each value of the empirical cumulative distribution function, $\{F_i^d\}_{i=1}^{n_q}$, by using a root-finding procedure to find the value of q such that $F_i^d = F(q|\Theta)$ up to the desired tolerance of error.

For the Double EMG distribution, the parameter vector is $\Theta = (\mu, \sigma, \lambda_L, \lambda_R)$ such that $q \sim f(q|\mu, \sigma, \lambda_L, \lambda_R)$. However, the inverse of the Double EMG distribution does not admit a closed form expression. Therefore, the inverse of the cumulative distribution function must be computed numerically.

[12] Following Head, Mayer, and Thoenig (2014), we define the empirical CDF over log-sales as $F_i^d = (i - 0.3)/(n_q + 0.4)$.

By a change of variables, log-sales are Normally distributed if sales are log-Normally distributed. Similarly, log-sales are Exponentially distributed if sales are distributed according to a Pareto. Both the Normal and Exponential distributions do, in fact, admit closed form expressions for the inverted cumulative distribution functions, of the forms:

$$
\begin{aligned}
q_i^N(\Theta^N) &= \mu^N + \sigma^N \Phi^{-1}(F_i^d) \\
q_i^E(\Theta^E) &= \log(\underline{r}) + (1/\lambda^E)\log(1 - F_i^d),
\end{aligned}
$$

where $\Phi(\cdot)$ is the CDF of a standard normal, and $\Theta^N = (\mu^N, \sigma^N)$ and $\Theta^E = (\underline{r}, \lambda^E)$ denote the parameter vectors for the Normal and Exponential distributions, respectively.

Finally, for a given theoretical distribution $F(\cdot|\Theta)$, we choose parameters Θ that minimize the sum of the squared errors between empirical and theoretical quantiles:

$$
\min_{\Theta} \sum_{i=1}^{n_q} \left(q_i^d - q_i(\Theta)\right)^2. \tag{2}
$$

In estimation, we use the 1st through 99th percentiles of the empirical CDF to estimate parameters. In practice, this choice eases computational burden compared to using each data point, without significantly changing the parameter estimates we recover. Furthermore, note that choosing parameters to minimize the sum of squared residuals is equivalent to Head et al.'s (2014) method of recovering parameters from quantile regressions. Our procedure recovers approximately the same parameter estimates for the Normal and Exponential distributions as those authors' method.

4.2 Double EMG Fit to Empirical Distributions

Having estimated distribution parameters, we now evaluate the fit of each distribution to the log-sales distributions across destination-years.

Result 1 *According to multiple goodness of fit statistics, the Double Exponentially Modified Gaussian distribution fits empirical log export sales distributions better than the Normal and*

Exponential distributions.

We first argue that the Double EMG distribution fits the data better than either the Normal or Exponential distributions by examining fitted distribution functions versus their empirical counterparts. We observe that the Double EMG distribution deviates from the data less than the Normal distribution, especially at the lower and upper percentiles. Panel A of Figure 3 compares the left tail across the empirical, Double EMG and Normal distributions. We observe that the Double EMG distribution provides a superior fit than the Normal in the left tail. Panel B of Figure 3 compares the right tail across distributions. We observe that the Double EMG distribution barely deviates from the empirical distribution up to the 99th percentile. In both tails, the Normal distribution is too thin relative to the data.

To better formalize the suggestive evidence we have put forth thus far, we consider three primary measures of the goodness of fit. Figure 4 presents goodness of fit statistics for each of the distributions under consideration. Specifically, the Figure 4 presents scatter plots of goodness of fit measures from the Exponential distribution (top row) or the Normal distribution (middle row), plotted against goodness of fit measures for the Double EMG distribution.

The first measure is the sum of squared errors (reported in the first column), which is given by the objective criterion from the estimation procedure given in equation (2) when evaluated at the error-minimizing parameters. Panel A and Panel D of Figure 4 show that errors are larger for the Normal and Exponential distributions than the Double EMG distribution. This is unsurprising, since the Double EMG distribution nests both the Normal and Exponential distributions as limiting cases (see Proposition 2). More interesting is the fact that both Panels A and D show that the errors are much larger for the Normal and Exponential distributions. However, the magnitude of the difference in errors is smaller for the Normal than the Exponential distribution.

The second measure is the Mean Absolute Error, which is given by:

$$MAE(\Theta) \equiv \frac{1}{n_q} \sum_{i=1}^{n_q} \left| q_i^d - q_i(\Theta) \right|.$$

The Mean Absolute Error measures the average deviation of the theoretical distribution from the empirical in either direction, but unlike the sum of squared errors does not more harshly penalize infrequent but large deviations. The second column (Panels B and E) of Figure 4 shows that errors are larger for the Normal and Exponential distributions than the EMG distribution. Therefore, the Mean Absolute Error reinforces that the Double EMG distribution has a superior fit, and that the difference in errors across the three distributions are not generated by a small number of large deviations from empirical observations.

The third measure is the Anderson-Darling statistic, which is given by:

$$AD(\Theta) \equiv n_q \sum_{i=1}^{n_q} \frac{(F_i^d - F(q_i|\Theta))^2}{F(q_i|\Theta)(1 - F(q_i|\Theta))} f(q_i|\Theta),$$

where $f(q_i|\Theta)$ is the theoretical probability density function.[13] Compared to our two other goodness of fit measures, the Anderson-Darling statistic places greater weight on observations in the tails of the distributions. To see this, consider the denominator within the integral. As $F(q|\Theta)$ approaches one or zero, $[F(q|\Theta)(1 - F(q|\Theta))]^{-1}$ approaches infinity. Therefore, the denominator is smallest for values of q for which $F(q|\Theta)$ is interior to $[0, 1]$. The third column of Figure 4 shows that the Anderson-Darling statistics are larger for the Normal and Exponential distributions than the Double EMG distribution. Therefore, the deviations of the Normal and Exponential distributions from the data can be, at least partially, attributed to a failure to match tail observations. This is particularly true for the Exponential distribution, which by construction cannot match the left tail of the sales distributions.

Taken together, these three measures show that the Double EMG distribution consistently fits the log-sales distributions better across destination-year observations, and that

[13] We compute this the density function as a numerical approximation to the derivative of the cumulative distribution function: $f(q|\Theta) \equiv (F(q + \Delta|\Theta) - F(q - \Delta|\Theta))/2\Delta$. The constant $\Delta > 0$ is chosen as a tenth of the maximum distance between successive empirical quantiles.

the Normal and Exponential distributions consistently fit the data worse in the tails of the distribution.

Result 2 *The Double Exponentially Modified Gaussian distribution can match the observed dispersion in positive and negative skewness of empirical log export sales distributions while the Normal and Exponential distributions cannot.*

The two most studied distributions in the trade and firm size dynamics literatures have a stark feature: the Normal and Exponential distributions have constant higher order moments that do not vary with parameters. In particular the Normal distribution is symmetric and therefore cannot possibly match the variation in skewness across destination-year observations. Furthermore, the Exponential distribution has constant skewness that does not depend on parameters of the distribution, which again makes it an ill-suited distribution for confronting the data on skewness. Figure 5 plots theoretical moments from the estimated distributions against the empirically observed moments.

Panel I of Figure 5 compares Kelly skewness in the data across destinations to Kelly skewness in the Double EMG distribution, Panel C of Figure 5 compares Kelly skewness in the data to that from the Normal distribution and Panel F from the Exponential distribution. It is immediately clear that the Double EMG distribution is the only distribution that exhibits variation in skewness across destination-year observations.

Panel B, Panel E, and Panel H of Figure 5 compare the interquartile range in the data (x-axes) to that in the three theoretical distributions. We see that all three distributions capture the general relationship in the data, although the Exponential distribution under predicts. Panel A, Panel D, and Panel G of Figure 5 show the comparison with the median of the empirical and theoretical distributions. Only the Double EMG distribution captures some amount of the cross destination-year variation in medians. The median of the Normal distribution is nearly constant across destination-years, which is at odds with the data. Lastly, the Exponential distribution consistently under predicts the median.

5 Theoretical Trade Elasticity

In this section, we employ the workhorse heterogeneous-firm trade model, along the lines of Melitz (2003) and Chaney (2008), to illustrate a relationship between export sales distributions and the partial trade elasticity. We demonstrate that variation in the partial trade elasticity across destinations arises from the variation in the extensive margin of firm entry and exit. We further show that the extensive margin elasticity can be identified from properties of empirical export sales distributions and, finally, develop an estimation approach for quantifying the magnitude of the extensive margin elasticity.

5.1 Sales

We consider an economic environment in which heterogeneous firms are monopolistic competitors and the representative household has constant elasticity of substitution preferences as in Melitz (2003). We further assume that entry is exogenous, as in Chaney (2008).[14] In this environment, a firm in country i has export sales in country j given by

$$r_{ij}(z_{ij}) \quad = \quad \left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} (\tau_{ij} w_i)^{1-\epsilon} \, Y_j P_j^{\epsilon-1} e^{z_{ij}}, \tag{3}$$

where ϵ is the elasticity of substitution, τ_{ij} is the iceberg transportation cost of exporting from country i to country j, w_i is the wage in country i, Y_j and P_j are the income level and the price level in country j, and z_{ij} is a country i firm's idiosyncratic *profitability* in country j.[15] Denote the probability density function and cumulative distribution function over firms' profitability by $g_{ij}(z)$ and $G_{ij}(z)$, respectively.

In a canonical Melitz (2003) environment, the underlying source of heterogeneity in profitability arises from heterogeneity in labor productivity across firms. Chaney (2008) further assumes that firm-level labor productivity, denoted by φ, is drawn from a Pareto distribution

[14] We refer the reader to Appendix B for a richer description of the economic environment that we consider.

[15] Following Foster, Haltiwanger, and Syverson (2008) and Bernard, Redding, and Schott (2010), profitability refers to firm-level shocks that may be the outcome of not only productivity differences but also differences in product demand.

with shape parameter ξ. In this case, $e^{z_{ij}}$ is equal to $\varphi^{\epsilon-1}$ and, by a change of variables, z_{ij} follows an Exponential distribution with shape parameter $\lambda = \xi/(\epsilon - 1)$.

In contrast to Chaney (2008), more recent work by Bas, Mayer, and Thoenig (2015) and Fernandes, Klenow, Meleshchuk, Pierola, and Rodríguez-Clare (2015) assumes that the underlying labor productivity φ is drawn from a log-Normal distribution, $\log \mathcal{N}(m, \upsilon^2)$. In this case, $e^{z_{ij}}$ equals $\varphi^{\epsilon-1}$, and z_{ij} follows a Normal distribution, $\mathcal{N}(\mu, \sigma^2)$ where $\mu = m(\epsilon-1)$ and $\sigma^2 = \upsilon^2(\epsilon - 1)^2$.

Generalizing both sets of distributional assumptions, the literature on firm-level learning and export decisions (see Timoshenko (2015)) assumes that there are two separate sources of heterogeneity in firm-level profitability. Heterogeneity arises from firm-level labor productivity φ drawn from a Pareto distribution with a shape parameter ξ, and firm-level product demand e^{θ} where θ is drawn from a Normal distribution $\mathcal{N}(m, \upsilon^2)$, so that $e^{z_{ij}}$ equals $e^{\theta}\varphi^{\epsilon-1}$.[16] In this case, a firm's profitability draw, $z_{ij} = \theta + \log(\varphi^{\epsilon-1})$, is the sum of a Normal and an Exponential random variable. Hence, z_{ij} is an EMG distributed random variable with parameters (μ, σ, λ), where $\mu = m + (\epsilon - 1)/\xi$, $\sigma^2 = \upsilon^2$, and $\lambda = \xi/\epsilon$.

In the context of the aforementioned firm-level learning literature, equation (3) is a general representation of sales from country i's firms to country j. While variation in profitability across firms may arise from differences in firm-specific labor productivity, destination-specific demand shocks or some combination of both, equation (3) shows that only the cumulative effect, summarized by the profitability draw z_{ij}, determines the level of sales.

5.2 Aggregation

The aggregate trade flow from country i to country j is defined as

$$X_{ij} = M_{ij} \int_{z_{ij}^*}^{+\infty} r_{ij}(z) \frac{g_{ij}(z)}{1 - G_{ij}(z)} dz, \tag{4}$$

[16]In order to be consistent with standard trade models, assume that there is no idiosyncratic or aggregate uncertainty after firms enter the market, firms always observe their product demand, and that the product demand does not vary over time (see Appendix B).

where M_{ij} is the mass of firms exporting from country i to j, and z_{ij}^* is the profitability entry threshold determined by the zero-profit condition. The partial elasticity of trade with respect to variable trade costs is defined as the percent-change in the aggregate trade flows between i and j as a result of a percent-change in variable trade costs τ_{ij} and can be expressed as

$$\frac{\partial \log X_{ij}}{\partial \log \tau_{ij}} = \underbrace{(1-\epsilon)}_{\substack{\text{level of the}\\ \text{partial trade elasticity}}} \Big(\underbrace{1}_{\substack{\text{intensive margin}\\ \text{contribution}}} + \underbrace{\gamma_{ij}}_{\substack{\text{extensive margin}\\ \text{contribution}}} \Big), \tag{5}$$

where γ_{ij} is given by

$$\gamma_{ij} \equiv \frac{g_{ij}(z_{ij}^*)}{(1-G_{ij}(z_{ij}^*))} \cdot \frac{e^{z_{ij}^*}}{E_{ij}(e^z | z > z_{ij}^*)}, \tag{6}$$

and $E_{ij}(\cdot | z > z_{ij}^*)$ is a conditional expectation over profitability.[17]

A conventional way to write equation (5) is $\partial \log X_{ij} / \partial \log \tau_{ij} = (1-\epsilon) + (1-\epsilon)\gamma_{ij}$, where $(1-\epsilon)$ is the intensive margin, and $(1-\epsilon)\gamma_{ij}$ is the extensive margin of the partial trade elasticity.[18] The advantage of the representation in equation (5) is that it highlights the distinct roles for the elasticity of substitution (ϵ) and the parameter governing the extensive margin elasticity (γ_{ij}) in determining the partial trade elasticity.[19] As can be seen from equation (5), ϵ governs the overall *level* of the trade elasticity, the contribution of the *intensive margin* to that level is always unity, and the contribution of the *extensive margin* is governed by γ_{ij}. Hence, every dollar of new trade can be decomposed into an intensive and extensive margin adjustment in the proportion of 1 to γ_{ij}, which is notably independent from ϵ.

Equations (5) and (6) illustrate the important role that micro-level firm heterogeneity plays in the aggregate measures of partial trade elasticity. From equation (5), the main source of variation in the partial trade elasticity across origin-destination country pairs arises from

[17] The partial trade elasticity is derived in Appendix B.

[18] Chaney (2008) first suggested this decomposition in conjunction with a Pareto distribution. With Pareto distributed z_{ij}, equation (6) simplifies to $\gamma_{ij} = (\lambda - 1)$, where $\lambda = \xi/(\epsilon - 1)$ and ξ is the Pareto tail parameter. Substituting the definition for λ yields Chaney's (2008) familiar formula for the partial trade elasticity: $\partial \log X_{ij} / \partial \log \tau_{ij} = -\xi$.

[19] Melitz and Redding (2015) provide a representation of the partial trade elasticity consistent with equation (5).

variation in the extensive margin elasticity, γ_{ij}.[20] In turn, equation (6) demonstrates that the extensive margin elasticity is solely determined by the shape of the log-sales distributions summarized by the probability density and the cumulative distribution functions, $g_{ij}(\cdot)$ and $G_{ij}(\cdot)$, respectively, and the entry profitability threshold, z^*_{ij}. We describe the estimation method for γ_{ij} in the next subsection.[21]

5.3 Estimation Method

To compute extensive margin elasticities in equation (6) we proceed by, first, estimating the origin-destination specific distributions of firm profitability, $G_{ij}(.)$ and, second, using estimated parameters of the distributions to recover the profitability entry thresholds, z^*_{ij}.

We recover parameters of the distribution $G_{ij}(\cdot)$ from micro-data on log-sales distributions by applying the estimation procedure in Section 4.1. From equation (3), we can write log-sales, $\log(r_{ij}(z_{ij}))$, as

$$\log(r_{ij}(z_{ij})) = \log\left(C_{ij}\right) + z_{ij}, \tag{7}$$

where $C_{ij} = \left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}\left(\tau_{ij}w_i\right)^{1-\epsilon} Y_j P_j^{\epsilon-1}$. Equation (7) highlights a one-to-one mapping between the distribution of log-sales and the distribution of the underlying profitability shocks. The two distributions are equal up to a scale parameter (C_{ij}). Hence, we uncover parameters of the firms' profitability distribution, $g_{ij}(z)$, by fitting that distribution to the empirical distribution of log-sales.

Given the estimated origin-destination specific distribution parameters, we follow the

[20] Variation in the level of the partial trade elasticity can, in principle, arise from the elasticity of substitution being destination specific. Following the vast majority of the literature we abstract away from this generalization and leave it for future research.

[21] In our generalization, notice that the partial trade elasticity is origin-destination specific due to the profitability entry thresholds *and* the profitability distributions being origin-destination specific. This is in contrast to a more restrictive assumption of a constant partial trade elasticity in Arkolakis et al.'s (2012) framework. A constant partial trade elasticity requires the elasticity to be independent of the endogenous profitability entry threshold *and* the profitability distribution not be destination specific. As a consequence, the most common gravity estimation approach will uncover only a sample average of origin-destination specific partial trade elasticities, which in fact may vary due to origin-destination specific distributions. This is also a key point in Melitz and Redding (2015).

approach developed in Bas, Mayer, and Thoenig (2015) for recovering the threshold z^*_{ij} from the average-to-minimum ratio of a sales distribution.[22] Using equation (3), we can express the theoretical average-to-minimum ratio as a function of z^*_{ij} alone:

$$\frac{E_{ij}(r_{ij}(z_{ij})|z_{ij} > z^*_{ij})}{r_{ij}(z^*_{ij})} = \frac{E_{ij}(e^{z_{ij}}|z_{ij} > z^*_{ij})}{e^{z^*_{ij}}}. \tag{8}$$

Subsequently, we compute the value of z^*_{ij} for which expression (8) equals the empirical average-to-minimum ratio.

Our approach slightly differs from Bas, Mayer, and Thoenig (2015) in that it does not require any knowledge of the elasticity of substitution when computing the profitability entry threshold z^*_{ij} or the corresponding extensive margin contribution to the partial trade elasticity, γ_{ij}. The main distinction is that our firm-level profitability construct consolidates various types of (productivity or demand) shocks and the elasticity of substitution into a single parameter, z_{ij}. As a result, we can use equation (7) to directly estimate parameters of the profitability distribution, $g_{ij}(.)$, from log export sales data, and then use equations (6) and (8) to compute z^*_{ij} and γ_{ij} using only the estimated distribution $g_{ij}(.)$. The advantage of our approach lies in its ability to estimate the extensive margin of the partial trade elasticity without estimating the elasticity of substitution, ϵ, which has it own challenges as extensively discussed in Bas, Mayer, and Thoenig (2015).[23]

6 Quantifying Trade Elasticities

In this section, we report the quantitative magnitudes of extensive margin elasticities and the resulting extent of variation in the partial trade elasticity across origin-destination country pairs. We show that the Double EMG implied extensive margin elasticity estimates are small

[22] As the name suggests, the "average-to-minimum ratio" in a destination-year distribution is constructed as the ratio of average sales to smallest sales record observed.

[23] Without loss of generality for the decomposition of trade elasticity into intensive and extensive margins, we assume $\epsilon = 6$. This value lies within the range used in the literature, see Broda and Weinstein (2006) and Bas et al. (2015). Subsequent results on the extensive margin contribution to the partial trade elasticity do not depend on the particular value we choose.

and only slightly larger than those implied by a Normal distribution. As a result, there is little variation in the partial trade elasticity. In contrast, the Exponential distribution overestimates the extensive margin elasticity by a large order of magnitude, which falsely attributes a large role to firm entry and exit in accounting for trade adjustments. Finally, we show that the size of the extensive margin adjustment is exaggerated when left truncated data are used in estimation.

6.1 Trade Elasticity Estimates

In Section 4 we showed that the Double EMG distribution closely fits the log-sales data and both fit the data better than other distributions commonly used in the trade literature. In this section, we proceed by comparing estimates of the extensive margin elasticity implied by the Double EMG distribution to the estimates implied by either Normal or Exponential distributions.

Table 2 reports summary statistics of the estimated values of the extensive margin elasticities, γ_{ij}, for the Double EMG, Normal, and Exponential distributions.[24] Result 3 summarizes the comparisons, as follows.

Result 3 *(Quantitative magnitude of the extensive margin elasticities)*

(i) The extensive margin contribution to the partial trade elasticity implied by the Double EMG distributions is small, with the average order of magnitude being 10^{-5}.

(ii) There is little variation in the extensive margin elasticity across origin-destination country pairs, as the standard deviation across estimates implied by the Double EMG distributions is on the order of 10^{-4}.

[24]The estimated values of the Exponential tail parameter from Section 4 all lie below unity. Such small values of the tail parameter occur due to fitting the Exponential distribution to data percentiles 1 through 99, to be consistent with the moments used to estimate parameters of the other considered distributions (Double EMG, Normal). When the value of the Exponential distribution tail parameter falls below unity, the moments of the distribution necessary to compute γ_{ij} are not defined. We nevertheless would like to provide some meaningful comparison between the Double EMG and Exponentially implied trade elasticities. We proceed by following the approach taken by Bas et al. (2015) and use estimates of Exponential tail parameter inferred from fitting the distribution to the top 5% of exporters for each each destination-year. The summary statistics for these estimated parameters are contained in Panel C of Table 1.

To put the magnitudes reported in Result 3 in perspective, consider the average sample value of the extensive margin elasticity implied by the Double EMG distribution, which equals $4.2 \cdot 10^{-5}$. This value should be understood in the context of equation (5), where the partial trade elasticity for an average observation equals $(1 - \epsilon) \cdot (1 + 4.2 \cdot 10^{-5})$. Given the elasticity of substitution is $\epsilon = 6$, a 1% decline in variable trade costs will increase trade by 5.002%, to which the entry and exit of firms contribute only 0.002%. To further emphasize the small magnitude of the extensive margin adjustment, suppose that a 1% decline in variable trade costs leads to 500 million dollars in increased export sales. According to equation (5), every one dollar of new trade can be decomposed into an intensive and extensive margin adjustment in the proportion of 1 to γ_{ij}, or 1 to $4.2 \cdot 10^{-5}$. Therefore, those 500 million dollars of new trade amount to \$499,979,000 of intensive versus \$21,000 of extensive margin adjustment. Hence, the extensive margin is quantitatively and economically small.

Furthermore, while the theory advanced in equation (5) attributes all origin-destination specific variation in the partial trade elasticity to variation in the extensive margin component, estimates of the Double EMG implied extensive margin component are so small that there is essentially no variation. As a result, the majority of trade adjustment in response to a decline in variable trade costs is accounted for by changes on the intensive margin. This is confirmed in the last two columns of Table 2, which reports that the partial trade elasticity exhibits negligible variation across destinations $(5.5 \cdot 10^{-4})$ and has an average value that approximately equals the level of the partial trade elasticity $(\epsilon - 1 = 5)$.[25]

6.2 Distribution Specification Bias

Relative to the Double EMG estimates, the often used Normal and Exponential distributions generate biased estimates of the extensive margin elasticity and should rather be viewed as providing upper (Exponential distribution) and lower (Normal distribution) bounds on the

[25] Our analysis does not exclude the possibility that there could be variation in the partial trade elasticity due to variation in the elasticity of substitution, ϵ, across destinations. Our results merely indicate that if there is variation in the partial trade elasticity across destinations, then it is not due to the extensive margin.

extensive margin elasticity estimates.This is the subject of Result 4, below.

Result 4 *(Distribution Specification Bias) The Normal and Exponential distribution extensive margin elasticity estimates provide lower and upper bounds, respectively, on the Double EMG distribution implied estimates.*

(i) The Normal distribution generates extensive margin elasticities that under-predict magnitudes by an average of 40% relative to the Double EMG distribution.

(ii) The Exponential distribution generates extensive margin elasticities that over-predict magnitudes by an average factor of 10^7 *relative to the Double EMG distribution.*

(iii) The larger is the mass in the right tail of a distribution, the larger is the distribution specification bias implied by a Normal relative to the Double EMG distribution fit.

Notice from Table 2, that the value of an average extensive margin elasticity implied by a Normal distribution is about 5 times smaller than that implied by the Double EMG distribution, and the value implied by an Exponential distribution is about 10,000 times larger. Quantitatively, these average magnitudes imply that for a hypothetical 500 million dollars increase in total newly created trade the Normal, Double EMG, and Exponential distributions, would attribute \$4,050, \$21,000, and \$175,000,000, respectively, to trade generated by new entrants.

For further comparison, Panel A of Figure 6 depicts estimates of the extensive margin elasticity γ_{ij} computed using a fitted Normal (dots) versus Exponential (stars) distribution (on the y-axis) with that computed using a fitted Double EMG distribution (on the x-axis). Each point corresponds to an elasticity that was computed for a given destination-year. The solid line is a 45-degrees line. For the Exponentially implied elasticities, each observation lies above the 45-degree line, and hence, the value of the elasticity is above that implied by the Double EMG distribution. In sharp contrast, the Normally implied elasticities for each destination-year observation lie on or below the 45-degree line indicating that the Double EMG implied elasticities are larger or equal to those implied by a Normal.

Table 3 further demonstrates the extent of the *distribution specification bias* in the extensive margin elasticities generated by a Normal and Exponential distributions. The table summarizes the size of and variation in the ratio of the Normal and Exponential elasticity estimates to Double EMG estimates across destination-year observations. As can be seen from the Table, a Normal distribution under-predicts the extensive margin elasticities by about 40 percent, on average, relative to the Double EMG estimates. In contrast, an Exponential distribution over-predicts the extensive margin elasticities by an average factor of 10^7 relative to the Double EMG estimates.

Lastly, the Normal distribution will exhibit its largest deviation from an empirical log-sales distribution that is positively skewed, and hence the extent of *distribution specification bias* generated by the Normal distribution heavily depends on the behavior of the right tail. To show this, Panel B of Figure 6 plots the ratio of extensive margin elasticity estimates from the Double EMG to those of the Normal distribution ($\gamma_{ij}^{DEMG}/\gamma_{ij}^{N}$), against the estimated Double EMG tail parameter, λ_R, for a subsample of distributions with a fat right tail, in particular $\lambda_R < 2$. Panel B of Figure 6 shows that the smaller is the estimated value of λ_R, the more fat-tailed is the log-sales distribution and the larger is the distribution specification bias implied by the Normal distribution. When the estimated tail parameter λ_R for a destination-year observation is less than 1.05, the Double EMG implied extensive margin elasticity is 10 to 1,000 times larger than the elasticity implied by a Normal distribution. Furthermore, this bias is pervasive, as 55% of observations have an estimated λ_R less than 2.

6.3 Sample Truncation Bias

Data restrictions pose additional challenges to correctly identifying bilateral trade elasticities. In that regard, many customs-level data sets are truncated. For example, in the French trade data used by Bas, Mayer, and Thoenig (2015), firms are not required to report their exports to an EU member country, unless the value of the shipment exceeds 250,000 euros. For non-EU member countries, firms need not report trade values below 1,000 euros. These reporting rules are exogenous to the researcher, but they are not without consequence for estimating policy relevant trade statistics.

At a conceptual level, left data censoring disproportionately reduces the size of the smallest firm that is observable to the econometrician and exaggerates its role in the sales distribution. Specifically, as is apparent from equation (6), omitting firms below a certain threshold, increases the absolute size of the smallest firm, z_{ij}^*, and reduces its size relative to an average firm, $E_{ij}(e^z|z > z_{ij}^*)/e^{z_{ij}^*}$. As a result, truncated samples are likely to overstate the contribution of the extensive margin to the partial trade elasticity and hence generate misleadingly high variation in bilateral elasticities.

In order to quantify the *truncation bias* in elasticity estimates due to exogenous data censoring, we conduct the following counterfactual experiment. We take the original log-sales data and drop all firm-destination-year observations with a value of exports below \$5,000.[26] We then re-fit a Double EMG distributions to these truncated data, recompute the average-to-minimum ratio based on the truncated sample, and finally recompute the extensive margin elasticities using the counterfactual Double EMG distributions' parameters and average-to-minimum ratios.

Result 5 *(Sample Truncation Bias) Consider a truncated sample due to dropping all firm-destination export sales lower than \$5,000.*
(i) Data truncation generates an upward bias in the extensive margin elasticity estimates with an average order of magnitude 10^3 *for the Double EMG distribution.*
(ii) Data truncation increases the standard deviation of the extensive margin and partial trade elasticities by a factor of 110 *for the Double EMG implied estimates.*

Table 5 reports estimates of the trade elasticities for a truncated sample and compares them to the estimates from a full sample. As can be seen from Panel B, a small data truncation of \$5,000 in firm sales per destination yields an upward bias in an extensive margin elasticity estimate by an average factor of $6.7 \cdot 10^3$ for a Double EMG distribution. Notice the increase in the average the extensive margin trade elasticity estimate implied by the Double EMG distribution, from $2.4 \cdot 10^{-5}$ to 0.009. To motivate the size of this bias,

[26] We have also computed results for thresholds of \$1,000 and \$10,000. The choice of threshold does not change our qualitative results.

suppose again that there were a reduction in variable trade costs that generates a 500 million dollar increase in export sales. The truncated sample estimates attribute \$4 million of the increase to trade generated by entering firms, while the non-truncated sample estimates would attribute only \$21,000.

Finally, the truncated sample generates larger variation in the partial trade elasticity estimates across destination-year observations. The standard deviation of Double EMG distribution generated partial trade elasticities is $5.5 \cdot 10^{-4}$ on a non-truncated sample and increases to 0.09 on the truncated sample. The latter standard deviation moves our estimated elasticities closer to those reported in Bas, Mayer, and Thoenig (2015).

Hence, using truncated samples of our data generates false conclusions regarding the magnitude, variation and, therefore, economic significance of extensive margin adjustments.

7 Robustness

7.1 Alternative Origin Country

In this robustness check, we ask whether our results are specific to Brazilian export sales during 1990-2001. We replicate each of our results using export sales data from a second country and verify that our results are robust to changes in economic environment.

We use Peruvian export data for the period between 1993 and 2009 from the World Bank Exporter Dynamics Database. For the detailed description of the data see Cebeci et al. (2012), Fernandes et al. (2016), and Freund and Pierola (2012). The dataset is comparable to the Brazilian export data and reports the value of export sales at the firm-product-destination-year level. We find that all results are qualitatively reproduced in the Peruvian data and in many cases we find that relationships and parameter estimates are quantitatively similar. The complete results are presented in Appendix C.

7.2 Product Definition

In this section, we ask whether our results are driven by the particular way in which manufacturing trade is defined in our paper. Specifically, prior to aggregating the data at the firm-destination-year level, we drop any firm-product-destination-year observations for agricultural products. If a firm simultaneously exports manufacturing and agricultural products, our approach can potentially create an abundance of small firms that might not primarily export manufacturing-industry products. Our dataset does not contain an indicator of a firm's primary industry of operation. Hence, we check the robustness of our results by dropping all firms that export at least one non-manufacturing product within a destination-year. The firms that remain only export manufacturing products.

Across firm-destination-year bins, 10% of firms export any non-manufacturing products. For an average firm, measured as the unconditional mean across firm-destination-years, non-manufacturing products account for 9% of export revenue. However, for those firms that export any non-manufacturing products, revenues are highly concentrated in non-manufacturing products with non-manufacturing products accounting for 92% of export revenue on average. Therefore, the main text included the remaining 8% of export sales from these 10% of firms. This section altogether excludes all sales, manufacturing and non-manufacturing, from these 10% of firms. We find that all quantitative results are nearly unchanged. The complete results are presented in Appendix D.

7.3 Industrial Composition

In Section 4.2 we applied the estimation procedure outlined in Section 4.1 to estimate the distribution parameters for each of the observations in our sample. Recall that we define an observation to be a distribution of log-export sales for a given export destination in a given year. We conduct our analysis at the country, rather than country-industry, level to make our results comparable to those in the literature.[27] We acknowledge that properties of distribu-

[27] Head et al. (2014) estimate the distributions of log-export sales of French firms in Belgium and Chinese firms in Japan, Bas et al. (2015) estimate distributions of log-export sales for each of the French and

tions might vary with the industrial composition of exports across destinations. To check the robustness of our results, Appendix E repeats our estimation for the log-export sales at the destination-year-industry level. We find that there is no statistically significant relationship between industry shares and skewness within destination-year observations. Therefore, no single industry drives tail fatness or skewness across destination-years, and the distribution estimation results remain quantitatively similar.

8 Conclusion

New trade theory predicts that welfare gains can be characterized by the partial trade elasticity (the elasticity of trade flows with respect to changes in variable trade costs), which can be decomposed into an intensive margin elasticity of changes in sales by incumbent firms and an extensive margin elasticity of changes in sales by entering and exiting firms.

In this paper we focus on the extensive margin elasticity, which depends crucially on the distribution governing firm-level heterogeneity, and ask: what is the role of small firms in determining the gains from trade? We find that small firms substantially attenuate the gains from trade.

We arrive at this answer by using a dataset on Brazilian export sales that, unlike standard trade datasets, has not been exogenously left-censored as a result of custom office rules. We observe the full export sales distribution. Exploiting the special features of this data, we contribute two stylized facts to the trade literature. First, export sales distributions are not symmetric and, in fact, exhibit high variation in skewness, mostly positive but also negative, across destinations. Second, export sales distributions have both fat right tails *and* fat left tails. While it is well known that the right tail of sales distributions tend to be fat, that the left tail is fat is new to the trade literature.

We show that these stylized facts are puzzling from the perspective of standard distributional assumptions in new trade models: neither the Pareto nor log-Normal distributions

Chinese export destinations, Nigai (2017) for total French exports.

exhibit a fat left tail or an ability to generate differences in skewness across export destinations. We confront this puzzle by introducing a distribution that generalizes both the Pareto and log-Normal distributions, the Double Exponentially Modified Gaussian distribution. We demonstrate that, due to its ability to generate different behavior in its two tails and generate variable skewness, the Double Exponentially Modified Gaussian distribution fits the export sales data better than either the Pareto or log-Normal distribution.

Given parameter estimates for each distribution across each export destination, we highlight two sets of issues that arise elsewhere in the trade literature. First, a *distribution selection bias* arises from making the wrong distributional assumptions. While the log-Normal distribution slightly underestimates the extensive margin trade elasticity and the Pareto distribution severely overestimates it, we find that the Double Exponentially Modified Gaussian distribution generates a very small (near zero) extensive margin trade elasticity. Second, we show that if our dataset were left-censored in a way that was consistent with other datasets, then there would be a *truncation bias*. That is, the extensive margin trade elasticity would be too large. We find the severity of the upward bias is largest for export destinations with fat right tails, which includes the range of standard Pareto tail estimates.

Therefore, small firms in the left tail of export distributions tend to drive the extensive margin elasticity down. Large extensive margin elasticities that have been computed using left-censored data overstate the contribution of the extensive margin to the gains from trade.

References

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Arkolakis, C., T. Papageorgiou, and O. Timoshenko (2015): "Firm Learning and Growth," Unpublished.

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

Badel, A. and M. Huggett (2014): "Taxing top earners: a human capital perspective," Unpublished.

Bas, M., T. Mayer, and M. Thoenig (2015): "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity," Unpublished.

Bastos, P., D. Dias, and O. Timoshenko (2016): "Learning, Prices, and Firm Dynamics," Unpublished.

Bastos, P. and D. A. Dias (2013): "The Life Cycle of Exporting Firms," Unpublished.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multi-Product Firms and Product Switching," *American Economic Review*, 100, 70–97.

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

Cabral, L. and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93, 1075–1090.

Cebeci, T., A. M. Fernandes, and C. Freund (2012): "Exporter Dynamics Database," World Bank Policy Research Working Paper No. 6229.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–21.

D'Agostino, R. B., A. J. Belanger, and R. B. D'Agostino Jr. (1990): "A suggestion for using powerful and informative tests of normality." *American Statistician*, 44, 316–321.

di Giovanni, J. and A. A. Levchenko (2013): "Firm entry, trade, and welfare in Zipf's world," *Journal of International Economics*, 89, 283–296.

di Giovanni, J., A. A. Levchenko, and R. Rancière (2011): "Power Laws in Firm Size and Openness to Trade: Measurement and Implications," *Journal of International Economics*, 58, 42–52.

Fernandes, A., C. Freund, and C. Pierola (2016): "Exporter behavior, country size and stage of development: Evidence from the Exporter Dynamics Database," *Journal of Development Economics*, 119, 121–137.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, M. D. Pierola, and A. Rodríguez-Clare (2015): "The intensive margin in trade: moving beyond pareto," Unpublished.

Flach, L. (2016): "Quality upgrading and price heterogeneity: Evidence from Brazilian exporters," *Journal of International Economics*, 102, 282–290.

Flach, L. and E. Janeba (Forthcoming): "Income inequality and export prices across countries," *Canadian Journal of Economics*.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98, 394–425.

Freund, C. and M. D. Pierola (2012): "Export Superstars," World Bank Policy Research Working Paper No. 6222.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294.

Gabaix, X. and R. Ibragimov (2011): "Rank-1/2: A simple way to improve the OLS Estimation of tail exponents," *Journal of Business and Economic Statistics*, 29, 24–29.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122, 621 – 660.

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

Heathcote, J. and H. Tsujiyama (2015): "Optimal Income Taxation: Mirrlees Meets Ramsey," Federal Reserve Bank of Minneapolis Research Department Staff Report 507.

Melitz, M. and S. Redding (2015): "New Trade Models, New Welfare Implications," *American Economic Review*, 105, 1105–46.

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

Molinaz, D. and M.-A. Muendler (2013): "Preparing to Export," NBER Working Paper No. 18962.

Mrázová, M., J. P. Neary, and M. Parenti (2016): "Technology, Demand, and the Size Distribution of Firms," Unpublished.

Nigai, S. (2017): "A tale of two tails: Productivity distribution and the gains from trade," *Journal of International Economics*, 104, 44–62.

Reed, W. J. and M. Jorgensen (2004): "The double Pareto-lognormal distributiona new parametric model for size distributions," *Communications in Statistics-Theory and Methods*, 33, 1733–1753.

Timoshenko, O. A. (2015): "Product Switching in a Model of Learning," *Journal of International Economics*, 95, 233–249.

Toda, A. A. (2014): "Incomplete market dynamics and cross-sectional distributions," *Journal of Economic Theory*, 154, 310–348.

Toda, A. A. and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123, 1177 – 1200.

Figures and Tables

Table 1: Properties of the log-sales distribution across destination-year observations over 1990-2001.

Statistic	Mean	Median	Standard Deviation	Min	Max
Panel A: Moment based statistics					
Standard Deviation	2.11	2.12	0.28	1.28	2.77
Skewness	0.03	0.02	0.24	-1.08	1.29
Nonparametric Skew	0.03	0.03	0.06	-0.19	0.21
Kurtosis	3.17	3.03	0.60	2.08	8.14
Panel B: Percentile based statistics					
Interquartile Range	2.82	2.83	0.49	1.50	4.44
Kelly Skewness	0.04	0.04	0.08	-0.30	0.35
Percentile Coefficient of Kurtosis	0.26	0.26	0.02	0.19	0.34
Panel C: Tail parameter estimates					
Top 5%	1.42	1.27	0.62	0.39	6.67
Top 10%	1.18	1.13	0.31	0.49	2.78
Top 15%	1.08	1.04	0.25	0.52	2.58
Bottom 5%	1.21	1.13	0.50	0.44	4.77
Bottom 10%	1.07	1.04	0.29	0.45	3.67
Bottom 15%	1.01	0.98	0.23	0.48	2.77

Note: the statistics are reported across 847 destination-year observations where at least 100 firms export.

Table 2: Trade elasticity estimates.

	Extensive Margin Elasticity, γ_{ij}		Partial Trade Elasticity, $\lvert(1-\epsilon)(1+\gamma_{ij})\rvert$	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Normal	$8.1 \cdot 10^{-6}$	$4.9 \cdot 10^{-5}$	5.00	$2.5 \cdot 10^{-4}$
Double EMG	$4.2 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	5.00	$5.5 \cdot 10^{-4}$
Exponential	0.54	0.61	7.68	3.04

Note: the table reports sample means and standard deviations of the corresponding elasticity estimates for various distributional assumptions. For the Double EMG, and Exponential distributions the means are reported across 281, and 705 observations respectively for which the estimates of λ_R, or the tail index, respectively, are greater than 1. The elasticities are not defined if values of the corresponding parameters are less than 1. To compute the partial trade elasticity, the value of $\epsilon = 6$ is assumed.

Table 3: Distribution specification bias.

	Normal		Exponential	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Double EMG	0.61	0.37	$7.6 \cdot 10^{7}$	$5.9 \cdot 10^{8}$

Note: the table reports the sample mean and standard deviation of ratios between extensive margin elasticity estimates. The columns indicate the numerator of Normal or Exponential distribution implied extensive margin elasticity estimates, while the row indicates the denomenator of Double EMG implied extensive margin elasticity estimates.

Table 4: Properties of the log-sales distribution across destination-year observations over 1990-2001, Sample truncated at $5,000.

Statistic	Mean	Median	Standard Deviation	Min	Max
Panel A: Moment based statistics					
Standard Deviation	1.74	1.74	0.26	1.05	2.44
Skewness	0.59	0.60	0.27	-0.40	2.43
Nonparametric Skew	0.12	0.13	0.07	-0.17	0.38
Kurtosis	2.99	2.86	0.83	1.85	13.02
Panel B: Percentile based statistics					
Interquartile Range	2.52	2.50	0.47	1.36	4.05
Kelly Skewness	0.15	0.16	0.09	-0.22	0.48
Percentile Coefficient of Kurtosis	0.28	0.28	0.02	0.21	0.36
Panel C: Tail parameter estimates					
Top 5%	1.47	1.32	0.73	0.38	11.91
Top 10%	1.22	1.17	0.33	0.47	2.87
Top 15%	1.11	1.07	0.26	0.51	2.70
Bottom 5%	8.61	7.57	4.12	1.04	37.40
Bottom 10%	4.61	4.35	1.45	1.12	11.27
Bottom 15%	3.35	3.22	0.92	1.11	7.07

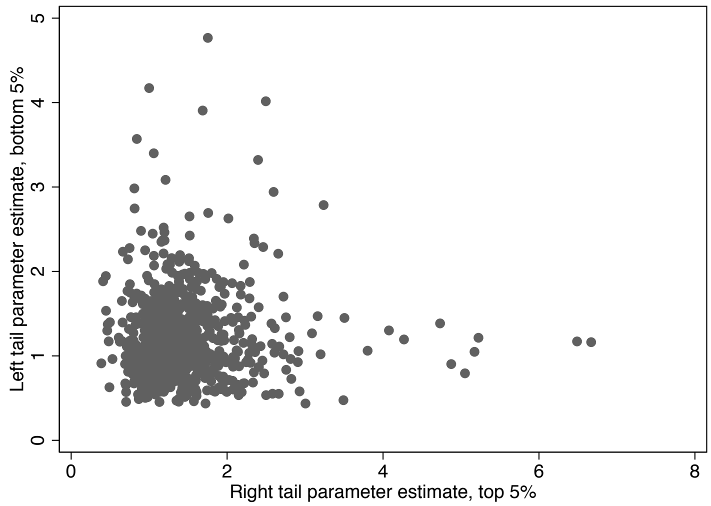
Note: the statistics are reported across 813 destination-year observations where at least 100 firms export and exprt value per firm-destiantion-year is $5,000 or more.

Table 5: Sample truncation bias.

Distribution	Extensive Margin Elasticity, γ_{ij}		Partial Trade Elasticity, $\lvert(1-\epsilon)(1+\gamma_{ij})\rvert$	
	Mean	Std. Dev.	Mean	Std. Dev.
Panel A: Elasticity Estimates				
Double EMG	0.009	0.018	5.05	0.09
Panel B: Magnitude of Bias				
Double EMG	$6.7 \cdot 10^3$	$2.3 \cdot 10^4$	1.009	0.018

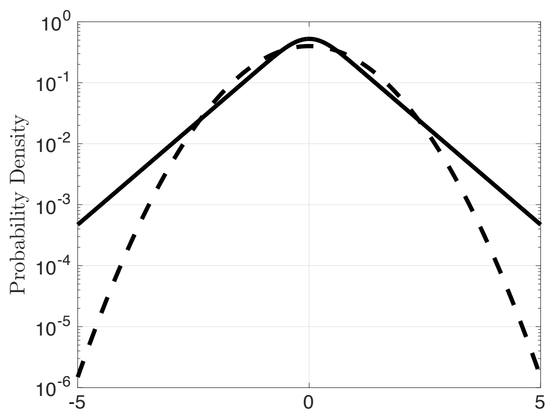
Note: Panel A of the table reports sample means and standard deviations of the corresponding elasticity estimates for various distributional assumptions. The values are reported for a truncated sample, where the truncation point is $5,000 sales per firm-destination. To compute the partial trade elasticity, the value of $\epsilon = 6$ is assumed. For the Double EMG distributions, the means are reported across 87 observations for which the estimates of λ_R is greater than 1. The elasticities are not defined if values of the corresponding parameters are less than 1. Panel B reports statistics for the ratio of the corresponding elasticity estimates from a truncated sample relative to the full sample for the two distributions.

Figure 1: Heterogeneity in the tail index estimates of log-sales distributions across export destinations.



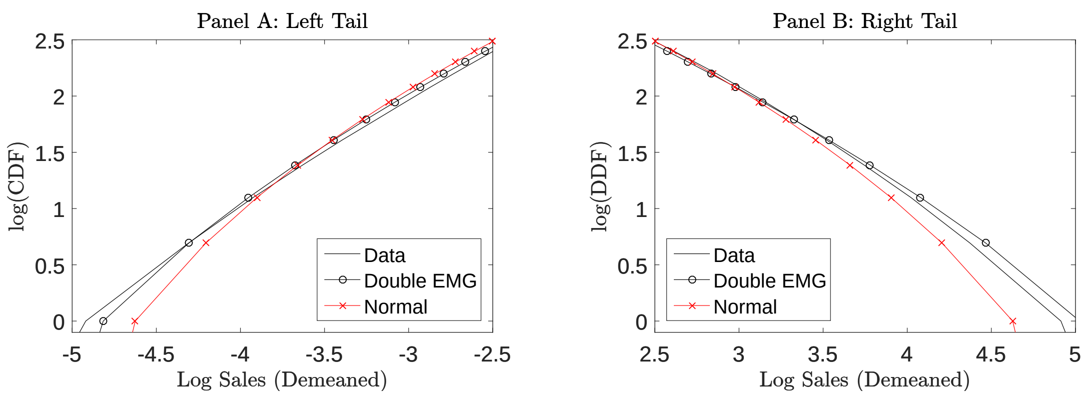
Notes: The figure depicts a scatter plot of the right and left tail index estimates for for the top and bottom 5% of firms. Each dot in the figure corresponds to an estimate of the right and left tail indexes for a given destination-year observation. A sample of 847 destination-year observations where at least 100 firms export.

Figure 2: An example of a Normal and a Double EMG distribution.



Notes: The figure depicts two probability density function (pdf) for two distribution with zero mean and unit variance. The solid line depicts a pdf for an symmetric EMG disquisition with tail parameters equal to 1.5. The dashed line depicts a pdf for a Normal distribution. The y-axis is plotted on the log scale.

Figure 3: Comparison of model errors and tail properties.



Notes: Each panel of the figure presents observations that have been averaged over each destination-year pair.

Figure 4: Goodness of fit statistics across each destination-year observation.

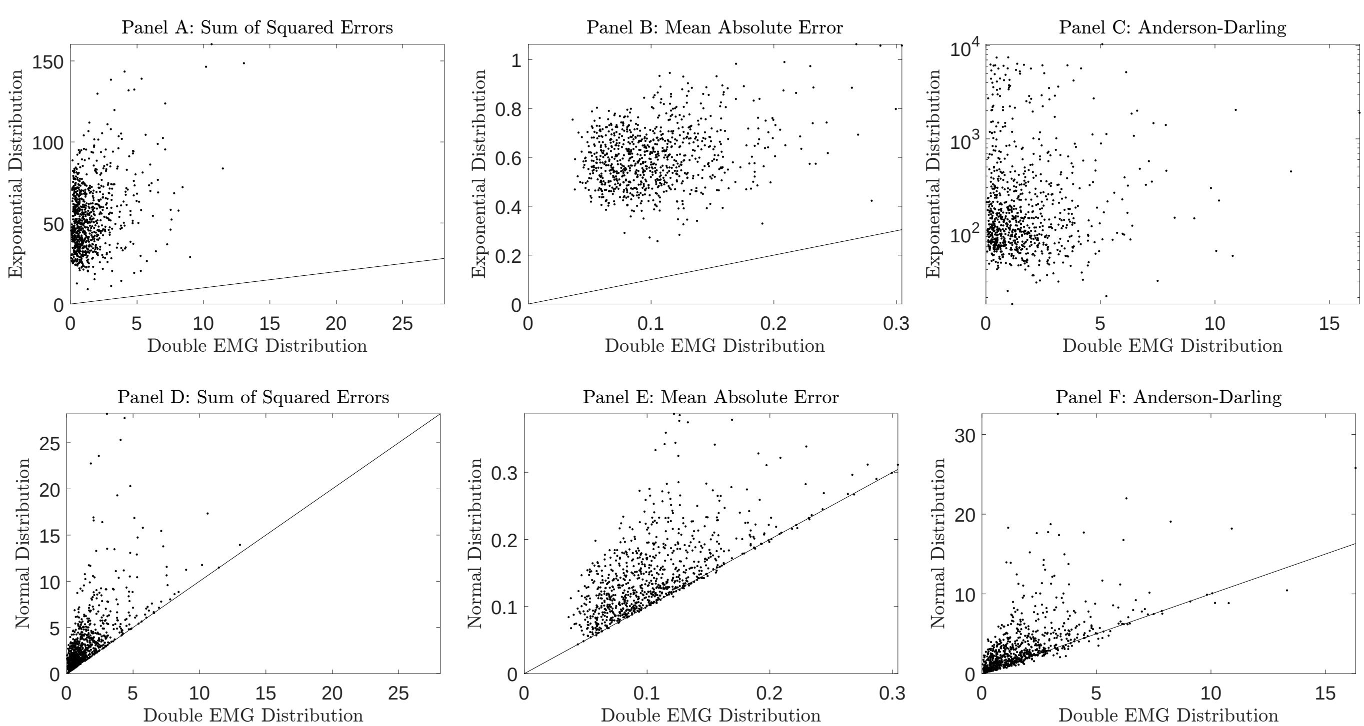
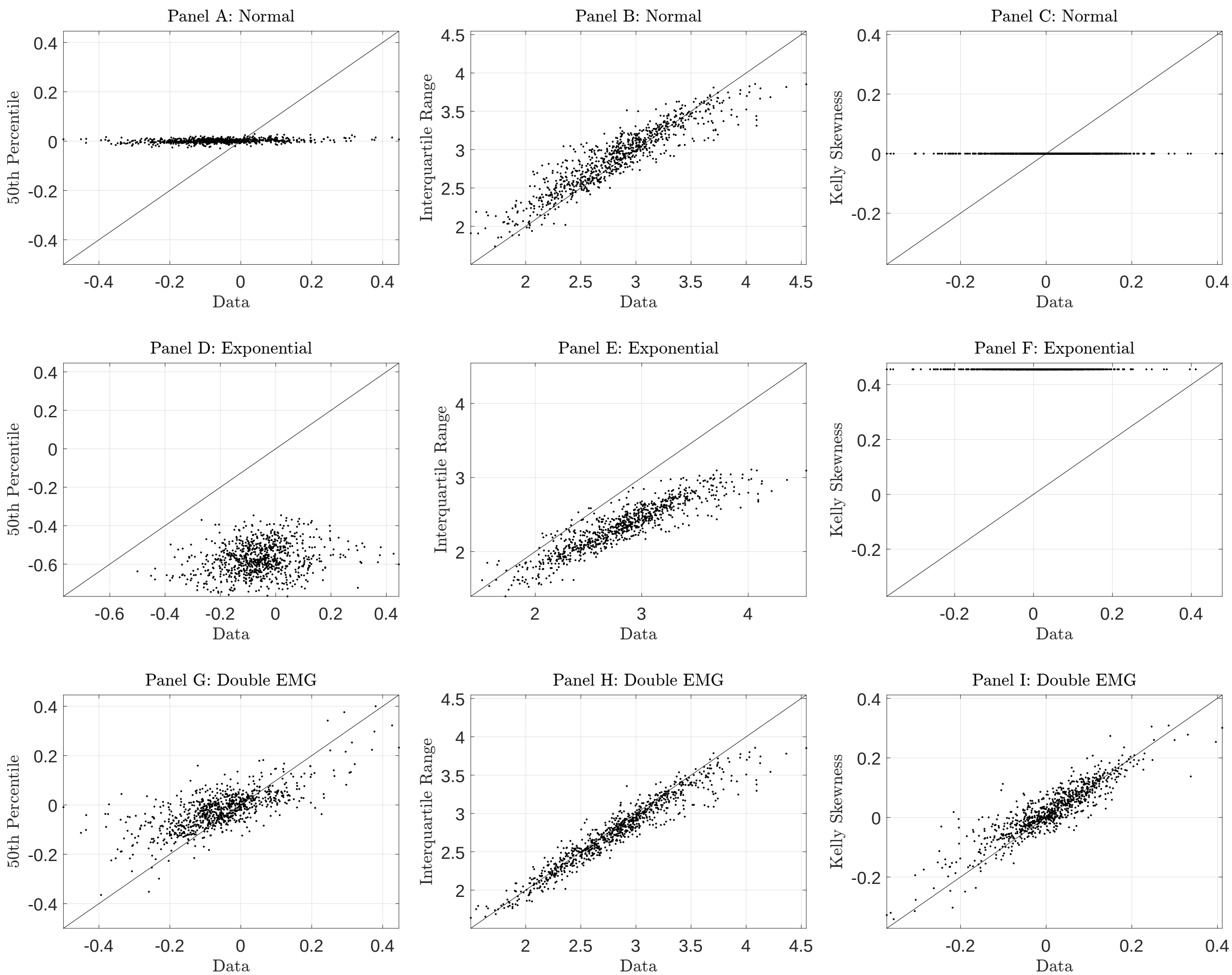
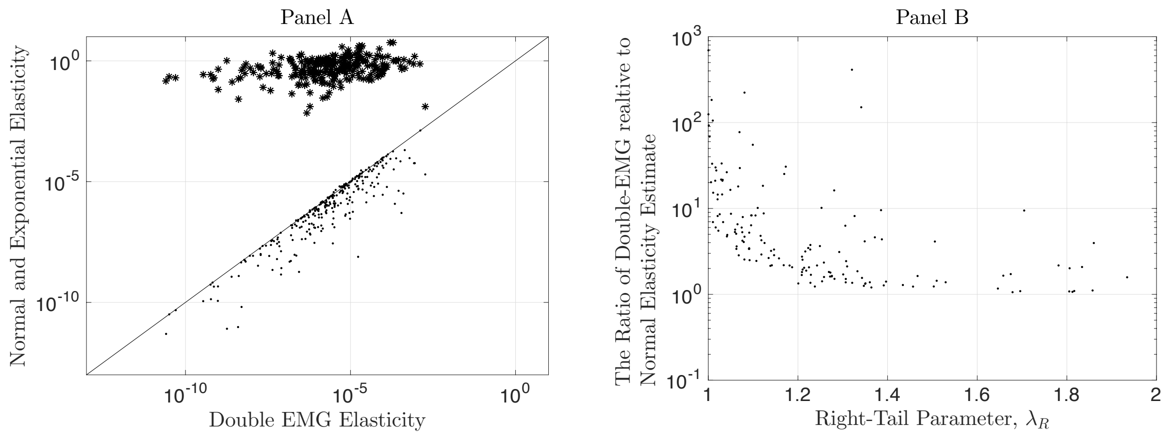


Figure 5: Comparison of empirical to model-generated moments.



Notes: Scatter plots show 50th percentile, interquartile range and Kelly skewness generated by the Normal distribution (Panels A, B and C), Exponential distribution (Panels D, E and F), and Double EMG distribution (Panels G, H and I) against the empirically observed statistics (x-axis) across destination-year pairs.

Figure 6: Extensive margin elasticity estimates.



Notes: Panel A of the figure depicts the estimates of the extensive margin elasticity for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. Dots (stars) plot Double EMG against Normal (Exponential) elasticity estimates. Panel B of the figure depicts the ratio of the the Double EMG relative to Normal extensive margin elasticity estimates for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. The elasticity is not defined for $\lambda_R \leq 1$.

A Proofs of Propositions (For Online Publication Only)

Proposition 1 *Let x and y be independent random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim \mathcal{DE}(\lambda_L, \lambda_R)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma > 0$, and $\lambda_L, \lambda_R > 0$. The random variable $z \equiv x + y$ has the cumulative distribution function $G : \mathbb{R} \to [0, 1]$ given by:*

$$G(z) = \Phi\left(\frac{z-\mu}{\sigma}\right) - \frac{\lambda_L}{\lambda_L + \lambda_R} e^{-\lambda_R(z-\mu)+\frac{\sigma^2}{2}\lambda_R^2} \Phi\left(\frac{z-\mu}{\sigma} - \lambda_R\sigma\right) + \frac{\lambda_R}{\lambda_L + \lambda_R} e^{\lambda_L(z-\mu)+\frac{\sigma^2}{2}\lambda_L^2} \Phi\left(-\frac{z-\mu}{\sigma} - \lambda_L\sigma\right),$$

the density function:

$$g(z) = \frac{\lambda_L\lambda_R}{\lambda_L + \lambda_R}\left[e^{-\lambda_R(z-\mu)+\frac{\sigma^2}{2}\lambda_R^2} \Phi\left(\frac{z-\mu}{\sigma} - \lambda_R\sigma\right) + e^{\lambda_L(z-\mu)+\frac{\sigma^2}{2}\lambda_L^2} \Phi\left(-\frac{z-\mu}{\sigma} - \lambda_L\sigma\right)\right],$$

and the moment generating function:

$$M_z(t) = \frac{\lambda_L\lambda_R}{(\lambda_L + t)(\lambda_R - t)} e^{\mu t + \frac{\sigma^2}{2}t^2}.$$

Proof of Proposition 1

Consider Lemma 1 below.

Lemma 1 *Let x and y be independent random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim \mathcal{E}(\lambda)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma > 0$, and $\lambda > 0$. The random variable $z \equiv x + y$ has the cumulative distribution function $G : \mathbb{R} \to [0, 1]$ given by:*

$$G(z) = \Phi\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)} \Phi\left(\frac{z-\mu}{\sigma} - \lambda\sigma\right),$$

the density function:

$$g(z) = \lambda e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)} \Phi\left(\frac{z-\mu}{\sigma} - \lambda\sigma\right),$$

and the moment generating function:

$$M_z(t) = \frac{\lambda}{\lambda - t} e^{\mu t + \frac{\sigma^2}{2} t^2}.$$

Proof of Lemma 1

Let x and y be random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim \mathcal{E}(\lambda)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda > 0$. For notational convenience, denote the density function that corresponds to the Normal distribution $\mathcal{N}(\mu, \sigma^2)$ by $f(x) = (1/\sigma)\phi((x - \mu)/\sigma)$. In the following derivations, we will make use of the conditional expectation for log-Normal random variables:

$$\int_{x^*}^{+\infty} (e^x)^\kappa f(x) dx = e^{\kappa\mu + \frac{1}{2}\kappa^2\sigma^2} \left(1 - \Phi\left(\frac{x^* - \mu}{\sigma} - \kappa\sigma\right)\right)$$

Let the random variable $z \equiv x + y$ have the distribution function $G : \mathbb{R} \to [0, 1]$, which we now derive:

$$\int_{-\infty}^{z^*} zg(z)dz = Prob\,(x + y < z^*) = \int_{-\infty}^{z*} \left(1 - e^{-\lambda(z^* - x)}\right) f(x)dx$$

Using the conditional expectation for log-Normal random variables, we obtain:

$$G(z^*) = \Phi\left(\frac{z^* - \mu}{\sigma}\right) - e^{-\lambda z^* + \left(\lambda\mu + \frac{1}{2}\lambda^2\sigma^2\right)} \Phi\left(\frac{z^* - \mu - \lambda\sigma^2}{\sigma}\right)$$

Next we derive the density function:

$$\begin{aligned}
\frac{\partial}{\partial z} \int_{-\infty}^{z} zdG(z) &= \int_{-\infty}^{z} \lambda e^{-\lambda y} f(z - y) dy \\
&= \frac{\lambda}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} e^{-\lambda y - \frac{1}{2}\left(\frac{z - y - \mu}{\sigma}\right)^2} dy \\
&= \frac{\lambda}{\sqrt{2\pi}\sigma} e^{-\lambda z + \lambda\mu + \frac{1}{2}\lambda^2\sigma^2} \int_{-\infty}^{z} e^{-\frac{1}{2}\left(\frac{z - y - \mu - \lambda\sigma^2}{\sigma}\right)^2} dy \\
g(z) &= \lambda e^{-\lambda z + \left(\lambda\mu + \frac{1}{2}\lambda^2\sigma^2\right)} \Phi\left(\frac{z - \mu - \lambda\sigma^2}{\sigma}\right)
\end{aligned}$$

Lastly, we derive the moment generating function. To do so, we will appeal to an intermediate result, that if $g(z)$ is a density function then it must integrate to one:

$$\int_{-\infty}^{+\infty} g(z)dz = \int_{-\infty}^{+\infty} \lambda e^{-\lambda z+\left(\lambda\mu+\frac{1}{2}\lambda^2\sigma^2\right)}\Phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right)dz$$
$$= e^{-\frac{1}{2}\lambda^2\sigma^2}\int_{-\infty}^{+\infty}\lambda\sigma e^{-\lambda\sigma y}\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dxdy$$

where we have used the change of variables $y = (z-\mu-\lambda\sigma^2)/\sigma$. Then we know that:

$$\int_{-\infty}^{+\infty}\lambda\sigma e^{-\lambda\sigma y}\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dxdy = e^{\frac{1}{2}\lambda^2\sigma^2}$$

Given this result, we can use the change of variables $y = (z-\mu-\lambda\sigma^2)/\sigma$ to derive the moment generating function:

$$M_z(t) = \int_{-\infty}^{+\infty} e^{-tz}\lambda e^{-\lambda z+\left(\lambda\mu+\frac{1}{2}\lambda^2\sigma^2\right)}\Phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right)dz$$
$$= \frac{\lambda}{\lambda-t}e^{-\frac{1}{2}\lambda^2\sigma^2+t(\mu+\lambda\sigma^2)}\cdot\int_{-\infty}^{+\infty}(\lambda-t)\sigma e^{-(\lambda-t)\sigma y}\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dxdy$$
$$= \frac{\lambda}{\lambda-t}e^{-\frac{1}{2}\lambda^2\sigma^2+t(\mu+\lambda\sigma^2)}\cdot e^{\frac{1}{2}(\lambda-t)^2\sigma^2}$$
$$= \frac{\lambda}{\lambda-t}\cdot e^{\mu t+\frac{\sigma^2}{2}t^2}$$

Note that the MGF for the EMG is the product of the MGF for the Exponential distribution and the MGF for the $\mathcal{N}(\mu,\sigma^2)$ distribution. QED.

Deriving the cumulative distribution function, density function and moment generating function of the Double Exponentially Modified Gaussian distribution follows steps from the proof for Lemma 1. The main difference is that the Double Exponential distribution changes functional form at its kink, $y = 0$. ■

Proposition 2 *(Limiting Results) Let z be a Double Exponentially Modified Gaussian distributed random variable with parameters $(\mu,\sigma,\lambda_L,\lambda_R)$. The random variable z is (i) an*

Exponentially Modified Gaussian distributed random variable as λ_L goes to infinity, (ii) an Exponentially Modified Gaussian distributed random variable with a Normal right tail and Exponential left tail as λ_R goes to infinity, (iii) a Double Exponentially distributed random variable as σ goes to zero, where if $\mu \neq 0$ then this limiting distribution is a shifted Double Exponential distribution, and (iv) an Exponentially distributed random variable as σ goes to zero and λ_L goes to infinity.

Proof of Proposition 2

Consider Lemma 2 below.

Lemma 2 *Let z be an Exponentially Modified Gaussian distributed random variable with parameters (μ, σ, λ). The random variable z is Normally distributed in the limit as λ goes to infinity, that is,*

$$\lim_{\lambda \to +\infty} \left[\Phi\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)} \Phi\left(\frac{z-\mu}{\sigma} - \lambda\sigma\right) \right] = \Phi\left(\frac{z-\mu}{\sigma}\right).$$

Furthermore, the random variable z is exponentially distributed in the limit as σ goes to zero. That is

$$\lim_{\sigma \to 0} \left[\Phi\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)} \Phi\left(\frac{z-\mu}{\sigma} - \lambda\sigma\right) \right] = 1 - e^{-\lambda(z-\mu)},$$

where, if $\mu > 0$ then this limiting distribution is a shifted Exponential distribution on $(\mu, +\infty)$. Lastly, consider the limit with respect to the value of the random variable z. There exists a value of z denoted $\bar{z}$ such that $\forall\ z \geq \bar{z}$ the distribution $G(z)$ approaches a shifted Exponential distribution:

$$\left[\Phi\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)} \Phi\left(\frac{z-\mu}{\sigma} - \lambda\sigma\right) \right] \approx 1 - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)}.$$

Proof of Lemma 2

We will consider each of the three limits of $G(z)$ in turn:

$$\text{(a) } \lambda \to +\infty, \quad \text{(b) } \sigma \to 0, \quad \text{(c) } z \to +\infty$$

(a) We first take the limit of $G(z)$ as $\lambda \to +\infty$. We know that

$$\lim_{\lambda \to +\infty} \Phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right) = \lim_{\lambda \to +\infty} e^{-\lambda z} = 0 \qquad \forall\, z \in \bar{\mathbb{R}}, z \neq 0$$

We must now show that $\exp(\lambda\mu + \lambda^2\sigma^2/2)$ reaches $+\infty$ at a slower rate than $\exp(-\lambda z) \times \Phi((z-\mu-\lambda\sigma^2)/\sigma)$ reaches 0. To do so, we appeal to l'Hôpital's rule:

$$\lim_{\lambda \to +\infty} \frac{\frac{\partial}{\partial \lambda} e^{-\lambda z} \Phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right)}{\frac{\partial}{\partial \lambda} e^{\lambda\mu + \frac{1}{2}\lambda^2\sigma^2}} = \lim_{\lambda \to +\infty} \frac{-z\Phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right) + \frac{1}{\sigma}\phi\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right)}{\mu + \lambda\sigma^2} e^{-\lambda z - \lambda\mu - \frac{1}{2}\lambda^2\sigma^2} = 0$$

The limit equals zero since $e^{\lambda^2\sigma^2}$ converges to zero faster than linearly, e.g. faster than $\lambda\sigma^2$.

(b) Next take the limit as $\sigma \to 0$. Let $\mu > 0$. As σ approaches 0, the Normal density becomes a point mass at μ and therefore: $\Phi\left(\frac{z-\mu}{\sigma}\right) = 1[z \geq \mu]$. Then clearly the limit of $G(z)$ as σ approaches 0 equals $1 - \exp(-\lambda(z-\mu))$ on $(\mu, +\infty)$ and zero elsewhere.

(c) Lastly, we show that there exists some $\bar{z}$ such that for all $z \geq \bar{z}$, $G(z) \approx 1 - \exp(-\lambda(z - \mu - \frac{1}{2}\lambda\sigma^2))$. We must show that as $z \to +\infty$, $\exp(-\lambda z)$ approaches 0 at a slower rate than $\Phi(\frac{z-\mu-\lambda\sigma^2}{\sigma})$ approaches 1. To do so, apply l'Hôpital's rule:

$$\lim_{z \to +\infty} \frac{e^{-\lambda z}}{\Phi(\frac{z-\mu-\lambda\sigma^2}{\sigma})} = \lim_{z \to +\infty} \frac{-\lambda e^{-\lambda z}}{\frac{1}{\sigma}\phi(\frac{z-\mu-\lambda\sigma^2}{\sigma})} \propto \lim_{z \to +\infty} e^{-\left(\lambda + \frac{\mu+\lambda\sigma^2}{\sigma^2}\right)z + \frac{1}{2}\left(\frac{z}{\sigma}\right)^2} = +\infty$$

Therefore, since both functions are decreasing in z, $\exp(-\lambda z)$ approaches 0 slower than $\Phi(\frac{z-\mu-\lambda\sigma^2}{\sigma})$ approaches 1. Therefore, there exists $\bar{z}$ sufficiently large such that:

$$\forall\, z \geq \bar{z} \qquad \Phi\left(\frac{z-\mu}{\sigma}\right) \approx 1 \qquad \text{and} \qquad \Phi(\frac{z-\mu-\lambda\sigma^2}{\sigma}) \approx 1$$

and

$$G(z) \approx 1 - e^{-\lambda z + \left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right)}$$

Therefore for sufficiently large values of z, the EMG is approximated by a shifted Exponential distribution. QED.

Deriving limiting results for the Double EMG distributions follows similar steps to the proof of Lemma 2. ■

Proposition 3 *If z is a Double Exponentially Modified Gaussian distributed random variable on $(-\infty, +\infty)$ then the skewness of z is given by*

$$skew(z) = 2\left(\frac{1}{\sigma^3\lambda_R^3} - \frac{1}{\sigma^3\lambda_L^3}\right)\left(1 + \frac{1}{\sigma^2\lambda_R^2} + \frac{1}{\sigma^2\lambda_L^2}\right)^{-\frac{3}{2}}.$$

Furthermore, the sign of $skew(z)$ is determined by the relative size of the tail parameters:

$$\left\{\begin{array}{llll} skew(z) & > 0 & if & \lambda_L > \lambda_R, \\ skew(z) & = 0 & if & \lambda_L = \lambda_R, \\ skew(z) & < 0 & if & \lambda_L < \lambda_R. \end{array}\right\}$$

Proof of Proposition 3

Given the moment generating function for the Double Exponentially Modified Gaussian distribution, we use the cumulant generating function defined as:

$$C_z(t) \equiv \log(M_z(t)) = \log(\lambda_R\lambda_L) - \log(\lambda_R - t) - \log(\lambda_L + t) + \left(\mu t + \frac{\sigma^2}{2}t^2\right).$$

The n-th centered moment is given by the n-th derivative of $C_z(t)$ evaluated at zero, or

$C_z^{(n)}(0)$. Therefore, the mean and variance are:

$$\begin{aligned}
C_z'(t) &= \mu - \frac{1}{\lambda_R - t}(-1) - \frac{1}{\lambda_L + t} \\
C_z''(t) &= \sigma^2 + \frac{-1}{(\lambda_R - t)^2}(-1) - \frac{-1}{(\lambda_L + t)^2} \\
C_z'''(t) &= \frac{-2}{(\lambda_R - t)^3}(-1) + \frac{-2}{(\lambda_L + t)^2}
\end{aligned}$$

which yield the first three centered moments of the Double Exponentially Modified Gaussian Distribution:

$$\begin{aligned}
E[x] = C_z'(0) &= \mu + \frac{1}{\lambda_R} - \frac{1}{\lambda_L} \\
E[(x - E[x])^2] = C_z''(0) &= \sigma^2 + \frac{1}{\lambda_R^2} + \frac{1}{\lambda_L^2} \\
E[(x - E[x])^3] = C_z'''(0) &= 2\left(\frac{1}{\lambda_R^3} - \frac{1}{\lambda_L^3}\right)
\end{aligned}$$

Therefore the skewness of the Double EMG distribution is:

$$skew(z) = \frac{2\left(\frac{1}{\lambda_R^3} - \frac{1}{\lambda_L^3}\right)}{\left(\sigma^2 + \frac{1}{\lambda_R^2} + \frac{1}{\lambda_L^2}\right)^{3/2}} = 2\left(\frac{1}{\sigma^3\lambda_R^3} - \frac{1}{\sigma^3\lambda_L^3}\right)\left(1 + \frac{1}{\sigma^2\lambda_R^2} + \frac{1}{\sigma^2\lambda_L^2}\right)^{-3/2}.$$

Notice that the skewness can also be expressed as

$$skew(z) = 2\frac{\lambda_L^3 - \lambda_R^3}{(\sigma^2\lambda_R^2\lambda_L^2 + \lambda_L^2 + \lambda_R^2)^{\frac{3}{2}}}.$$

The sign properties of the skewness follow immediately. ■

B Baseline Trade Model (For Online Publication Only)

B.1 Economic Environment

There are N countries. We will denote by i the origin country and by j a destination country. Each country j is populated by L_j identical consumers with preferences given by a constant elasticity of substitution utility function given by

$$U_j = \left(\sum_{i=1}^{N} \int_{\omega \in \Omega_{ij}} \left(e^{\theta_{ij}(\omega)} \right)^{\frac{1}{\epsilon}} c_{ij}(\omega)^{\frac{\epsilon-1}{\epsilon}} d\omega \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{9}$$

where Ω_{ij} is the set of varieties consumed in country j originating from country i, $c_{ij}(\omega)$ is the consumption of variety $\omega \in \Omega_{ij}$, ϵ is the elasticity of substitution, and $\theta_{ij}(\omega)$ is the demand parameter for variety $\omega \in \Omega_{ij}$.[28]

Each consumer owns a share of domestic firms and is endowed with one unit of labor that is inelastically supplied to the market. Cost minimization yields optimal demand for variety $\omega \in \Omega_{ij}$ given by

$$c_{ij}(\omega) = e^{\theta_{ij}(\omega)} p_{ij}(\omega)^{-\epsilon} Y_j P_j^{\epsilon-1}, \tag{10}$$

where $p_{ij}(\omega)$ is the price of variety $\omega \in \Omega_{ij}$, Y_j is income in country j and P_j is the aggregate price index in country j. The aggregate price index is given by

$$P_j^{1-\epsilon} = \sum_{i=1}^{N} \int_{\omega \in \Omega_{ij}} e^{\theta_{ij}(\omega)} p_{ij}(\omega)^{1-\epsilon} d\omega. \tag{11}$$

B.2 Supply

As in Chaney (2008), each country is endowed with the exogenous mass J_i of prospective entrants. Upon entry, a firm is endowed with an idiosyncratic labor productivity level φ and a destination-specific demand parameter θ_j. Productivity and destination-specific demand parameters are drawn from separate independent distributions. Firms face fixed f_{ij} and

[28] Bernard, Redding, and Schott (2010) interpret $\theta_{ij}(\omega)$ as variations in consumer tastes or relative demand across different varieties. In Timoshenko (2015) $\theta_{ij}(\omega)$ represents product demand that firms need to learn over time through market participation.

variable τ_{ij} costs of selling from country i to country j denominated in terms of units of labor.

Once productivity and demand are realized, firms compete in a monopolistically competitive environment. Firms maximize profits subject to the consumer demand (10) yielding the optimal price given by

$$p_{ij}(\varphi) = \frac{\epsilon}{\epsilon - 1} \frac{\tau_{ij} w_i}{\varphi},$$

where w_i is the wage in country i. The corresponding firm's optimal revenues and profits are given by

$$r_{ij}(\theta_{ij}, \varphi) = \left(\frac{\epsilon - 1}{\epsilon}\right)^{\epsilon - 1} (\tau_{ij} w_i)^{1-\epsilon} Y_j P_j^{\epsilon - 1} e^{\theta_{ij}} \varphi^{\epsilon - 1}, \tag{12}$$

$$\pi_{ij}(\theta_{ij}, \varphi) = \frac{r_{ij}(\theta_{ij}, \varphi)}{\epsilon} - w_i f_{ij}. \tag{13}$$

Notice from equations (12) and (13) that a firm's profitability in market j depends on both a firm's productivity φ and a demand parameter θ_j in a multiplicative way. Hence a low productivity firm can generate positive profits if the demand for its product is high, and vise versa. Thus, selection into a market occurs based on a firm's profitability, and not productivity or demand alone. Denote by z_{ij} the firm's payoff relevant state variable given by

$$z_{ij} = \theta_{ij} + \log\left(\varphi^{\epsilon - 1}\right). \tag{14}$$

We will refer to z_{ij} as a firm's *profitability* in market j. Given z_{ij}, we can rewrite the firm's optimal revenue and profit as a function of profitability z_{ij} as

$$r_{ij}(z_{ij}) = \left(\frac{\epsilon - 1}{\epsilon}\right)^{\epsilon - 1} (\tau_{ij} w_i)^{1-\epsilon} Y_j P_j^{\epsilon - 1} e^{z_{ij}}. \tag{15}$$

$$\pi_{ij}(z_{ij}) = \frac{r_{ij}(z_{ij})}{\epsilon} - w_i f_{ij}. \tag{16}$$

Since there are no sunk entry costs, the profitability entry threshold is determined by the zero-profit condition $\pi_{ij}(z_{ij}^*) = 0$ and is given by

$$e^{z_{ij}^*} = \frac{\epsilon w_i f_{ij} (w_i \tau_{ij})^{\epsilon-1}}{\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} Y_j P_j^{\epsilon-1}}. \tag{17}$$

The firm's optimal revenue can then be written as a function of a firm's profitability, z_{ij}, and the profitability entry threshold z_{ij}^* as

$$r_{ij}(z_{ij}) = \epsilon w_i f_{ij} \frac{e^{z_{ij}}}{e^{z_{ij}^*}}. \tag{18}$$

B.3 Trade Elasticity

The value of exports from country i to country j is defined as

$$X_{ij} = M_{ij} \int_{z_{ij}^*}^{+\infty} r_{ij}(z) \frac{g_{ij}(z)}{1 - G_{ij}(z)} dz, \tag{19}$$

where M_{ij} is the equilibrium mass of firms selling from country i to country j and is given by

$$M_{ij} = J_i(1 - G_{ij}(z_{ij}^*)). \tag{20}$$

The cumulative and the probability distribution functions of firms profitabilities are denoted by $G_{ij}(z)$ and $g_{ij}(z)$ correspondingly.

Proposition 4 below establishes the partial trade elasticity result.

Proposition 4 *The partial elasticity of trade flows with respect to variable trade costs is given by*

$$\frac{\partial \log X_{ij}}{\partial \log \tau_{ij}} = (1 - \epsilon)(1 + \gamma_{ij}),$$

where γ_{ij} *given by*

$$\gamma_{ij} \;\;=\;\; \frac{g_{ij}(z^*_{ij})}{(1-G_{ij}(z^*_{ij}))}\frac{e^{z^*_{ij}}}{E_{ij}(e^z|z>z^*_{ij})}.$$

Proof: Substitute equations (18) and (20) into equation (19) to obtain

$$X_{ij} = \epsilon J_i w_i f_{ij}\int_{z^*_{ij}}^{+\infty}\left(e^{z_{ij}-z^*_{ij}}\right)g_{ij}(z)dz.$$

Using the Leibniz's Integration Rule:

$$\begin{aligned}\frac{\partial X_{ij}}{\partial \tau_{ij}} &= \epsilon J_i w_i f_{ij}\left[-\frac{\partial z^*_{ij}}{\partial \tau_{ij}}\int_{z^*_{ij}}^{+\infty}\left(e^{z_{ij}-z^*_{ij}}\right)g_{ij}(z)dz - g(z^*_{ij})\frac{\partial z^*_{ij}}{\partial \tau_{ij}}\right]\\ &= \epsilon J_i w_i f_{ij}\left[-\frac{\partial z^*_{ij}}{\partial \tau_{ij}}e^{-z^*_{ij}}(1-G_{ij}(z^*_{ij}))E_{ij}(e^z|z>z^*_{ij}) - g(z^*_{ij})\frac{\partial z^*_{ij}}{\partial \tau_{ij}}\right].\end{aligned}$$

Now we must derive the partial derivative of the profitability threshold with respect to a change in variable costs. To do so, we use the expression characterizing the threshold in equation (17):

$$\frac{\partial z^*_{ij}}{\partial \tau_{ij}} = \frac{\partial}{\partial \tau_{ij}}\log\left(\frac{\epsilon w_i f_{ij}(w_i\tau_{ij})^{\epsilon-1}}{\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}}Y_j P_J^{\epsilon-1}\right) = \frac{\epsilon-1}{\tau_{ij}}.$$

Notice that the value of trade flows can be expressed as

$$X_{ij} = \epsilon J_i w_i f_{ij} e^{-z^*_{ij}}(1-G_{ij}(z^*_{ij}))E_{ij}(e^z|z>z^*_{ij}).$$

Therefore, the partial elasticity of trade is:

$$
\begin{aligned}
\frac{\partial \log X_{ij}}{\partial \log \tau_{ij}} &= \frac{\tau_{ij}}{X_{ij}} \cdot \epsilon J_i w_i f_{ij} \left[\frac{1-\epsilon}{\tau_{ij}} e^{-z_{ij}^*} (1 - G_{ij}(z_{ij}^*)) E_{ij}(e^z | z > z_{ij}^*) + g(z_{ij}^*) \frac{1-\epsilon}{\tau_{ij}} \right] \\
&= (1-\epsilon) + \frac{g(z_{ij}^*)}{1 - G_{ij}(z_{ij}^*)} \cdot \frac{(1-\epsilon) e^{z_{ij}^*}}{E_{ij}(e^z | z > z_{ij}^*)} \\
&= (1-\epsilon) + (1-\epsilon)\gamma_{ij} \\
&= (1-\epsilon)(1+\gamma_{ij}).
\end{aligned}
$$

as desired. ■

B.4 Conditional Expectations

Finally, we derive the conditional expectation for the Double EMG distribution. Given conditional expectations, it is possible to compute the extensive margin elasticity from Proposition 4. The conditional expectation of the Double EMG distribution in Proposition 5 as follows.

Proposition 5 *If z is a Double Exponentially Modified Gaussian distributed random variable on $(-\infty, +\infty)$ then the conditional first moment on $(z^*, +\infty)$ is*

$$
\begin{aligned}
\int_{z^*}^{+\infty} e^z g(z) dz = M_z(1) \left(1 - \Phi\left(\frac{z^* - \mu}{\sigma} - \sigma \right) \right) \\
+ \frac{1}{\lambda_R - 1} \frac{\lambda_L \lambda_R}{\lambda_L + \lambda_R} \; e^{z^* - \lambda_R (z^* - \mu) + \frac{\sigma^2}{2} \lambda_R^2} \Phi\left(\frac{z^* - \mu}{\sigma} - \lambda_R \sigma \right) \\
- \frac{1}{\lambda_L + 1} \frac{\lambda_L \lambda_R}{\lambda_L + \lambda_R} \; e^{z^* + \lambda_L (z^* - \mu) + \frac{\sigma^2}{2} \lambda_L^2} \Phi\left(-\frac{z^* - \mu}{\sigma} - \lambda_L \sigma \right).
\end{aligned}
$$

Proof: Let $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim \mathcal{DE}(\lambda_L, \lambda_R)$ and z be a Double EMG distributed random variable on $(-\infty, +\infty)$. Then the conditional first moment on $(z^*, +\infty)$ is:

$$
\int_{z^*}^{+\infty} e^z G(dz) = \left[\int_{x > z^*} \int_{y > 0} + \int_{x > z^*} \int_{z^* - x}^{0} + \int_{x < z^*} \int_{y > z^* - x} \right] e^{x+y} f(x) g(y) dx dy
$$

First take each integral in turn and define cases. The first (case is $y > 0$):

$$\int_{x>z^*}\int_{y>0} e^{x+y} f(x)g(y)dxdy = \int_{x>z^*} e^x f(x)dx \cdot \int_{y>0} e^y g(y)dy$$

the second (case is $y < 0$):

$$\int_{x>z^*}\int_{z^*-x}^{0} e^{x+y} f(x)g(y)dxdy = \int_{x>z^*} e^x \left[\int_{z^*-x}^{0} e^y g(y)dy\right] f(x)dx$$

and, lastly, the third (case is $y > 0$):

$$\int_{x<z^*}\int_{y>z^*-x} e^{x+y} f(x)g(y)dxdy = \int_{x<z^*} e^x \left[\int_{y>z^*-x} e^y g(y)dy\right] f(x)dx$$

We can simplify the integrals in each case. Simplifying the first case:

$$\begin{aligned}\int_{z^*}^{+\infty} e^x f(x)dx \cdot \int_0^{+\infty} e^y g(y)dy &= e^{\mu+\frac{\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu-\sigma^2}{\sigma}\right)\cdot\int_0^{\infty} e^y \frac{\lambda_L\lambda_R}{\lambda_L+\lambda_R}e^{-\lambda_L y}dy \\ &= e^{\mu+\frac{\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu-\sigma^2}{\sigma}\right)\cdot\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{-1}{\lambda_R-1}e^{-(\lambda_R-1)y}\Big|_0^{\infty} \\ &= e^{\mu+\frac{\sigma^2}{2}}\left(1-\Phi\left(\frac{z^*-\mu-\sigma^2}{\sigma}\right)\right)\cdot\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1}\end{aligned}$$

the second case:

$$\begin{aligned}&\int_{x>z^*} e^x\left(\int_{z^*-x}^{0} e^y \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}e^{\lambda_L y}dy\right)f(x)dx \\ &\quad= \int_{x>z^*} e^x\left(\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_L+1}\left(1-e^{(\lambda_L+1)(z^*-x)}\right)\right)f(x)dx \\ &\quad= \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_L+1}\left(e^{\mu+\frac{\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu-\sigma^2}{\sigma}\right)-e^{(\lambda_L+1)z^*}\int_{x>z^*}e^{-\lambda_L x}f(x)dx\right) \\ &\quad= \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_L+1}\left(e^{\mu+\frac{\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu-\sigma^2}{\sigma}\right)-e^{z^*+\lambda_L(z^*-\mu)+\frac{\lambda_L^2\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu+\lambda_L\sigma^2}{\sigma}\right)\right)\end{aligned}$$

and, lastly, the third case:

$$\begin{aligned}
&\int_{x<z^*} e^x \left[\int_{y>z^*-x} e^y g(y)dy\right] f(x)dx \\
&\quad = \int_{x<z^*} e^x \left[\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{-1}{\lambda_R-1}\left(0-e^{-(\lambda_R-1)(z^*-x)}\right)\right] f(x)dx \\
&\quad = \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1}e^{-(\lambda_R-1)z^*}\int_{x<z^*} e^{\lambda_R x} f(x)dx \\
&\quad = \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1}e^{z^*-\lambda_R(z^*-\mu)+\frac{\lambda_R^2\sigma^2}{2}}\Phi\left(\frac{z^*-\mu-\lambda_R\sigma^2}{\sigma}\right)
\end{aligned}$$

Summing these integrals together we obtain:

$$\begin{aligned}
\int_{z^*}^{+\infty} e^z H(dz) &= e^{\mu+\frac{\sigma^2}{2}}\left(1-\Phi\left(\frac{z^*-\mu-\sigma^2}{\sigma}\right)\right)\cdot\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1} \\
&+\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_L+1}\left(e^{\mu+\frac{\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu-\sigma^2}{\sigma}\right)-e^{z^*+\lambda_L(z^*-\mu)+\frac{\lambda_L^2\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu+\lambda_L\sigma^2}{\sigma}\right)\right) \\
&+\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1}e^{z^*-\lambda_R(z^*-\mu)+\frac{\lambda_R^2\sigma^2}{2}}\Phi\left(\frac{z^*-\mu-\lambda_R\sigma^2}{\sigma}\right)
\end{aligned}$$

Therefore, the final conditional expectation is:

$$\begin{aligned}
\int_{z^*}^{+\infty} e^z H(dz) &= \frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\left(\frac{1}{\lambda_R-1}+\frac{1}{\lambda_L+1}\right)e^{\mu+\frac{\sigma^2}{2}}\left(1-\Phi\left(\frac{z^*-\mu-\sigma^2}{\sigma}\right)\right) \\
&+\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_R-1}e^{z^*-\lambda_R(z^*-\mu)+\frac{\lambda_R^2\sigma^2}{2}}\Phi\left(\frac{z^*-\mu-\lambda_R\sigma^2}{\sigma}\right) \\
&-\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\frac{1}{\lambda_L+1}e^{z^*+\lambda_L(z^*-\mu)+\frac{\lambda_L^2\sigma^2}{2}}\Phi\left(-\frac{z^*-\mu+\lambda_L\sigma^2}{\sigma}\right)
\end{aligned}$$

where

$$\frac{\lambda_R\lambda_L}{\lambda_R+\lambda_L}\left(\frac{1}{\lambda_R-1}+\frac{1}{\lambda_L+1}\right)=\frac{\lambda_R\lambda_L}{(\lambda_R-1)(\lambda_L+1)}$$

as desired. ■

C Robustness on Peruvian Data (For Online Publication Only)

In this appendix, we ask whether our results are specific to Brazilian export sales during 1990-2001. We replicate each of our results using export sales data from a second country and verify that our results are robust to changes in economic environment.

We use Peruvian export data for the period between 1993 and 2009 from the World Bank Exporter Dynamics Database. For the detailed description of the data see Cebeci et al. (2012), Fernandes et al. (2016), and Freund and Pierola (2012). The dataset is comparable to the Brazilian export data and reports the value of export sales at the firm-product-destination-year level.

Below, we reproduce Table 1 to Table 3 and Figure 1 to Figure 6. We find that all results are qualitatively reproduced in the Peruvian data and in many cases we find that relationships and parameter estimates are quantitatively similar.

Table C1: Properties of the log-sales distribution across destination-year observations over 1993-2009, Peru.

Statistic	Mean	Meian	Standard Deviation	Min	Max
Panel A: Moment based statistics					
Standard Deviation	2.64	2.62	0.37	1.86	3.82
Skewness	-0.12	-0.13	0.34	-1.23	1.16
Nonparametric Skew	0.02	0.02	0.07	-0.29	0.31
Kurtosis	3.59	3.49	0.75	2.11	7.12
Panel B: Percentile based statistics					
Interquartile Range	3.42	3.37	0.47	2.35	4.96
Kelly Skewness	0.05	0.05	0.11	-0.37	0.53
Percentile Coefficient of Kurtosis	0.26	0.26	0.02	0.15	0.33
Panel C: Tail parameter estimates					
Top 5%	1.17	1.08	0.46	0.42	3.77
Top 10%	0.997	0.96	0.33	0.37	3.16
Top 15%	0.89	0.86	0.25	0.36	1.81
Bottom 5%	0.74	0.63	0.46	0.25	5.22
Bottom 10%	0.69	0.60	0.31	0.31	3.38
Bottom 15%	0.69	0.62	0.25	0.31	2.21

Note: the statistics are reported across 415 destination-year observations where at least 100 firms export.

Table C2: Trade elasticity estimates, Peru.

	Extensive Margin Elasticity, γ_{ij}		Partial Trade Elasticity, $\lvert(1-\epsilon)(1+\gamma_{ij})\rvert$	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Normal	$1.6 \cdot 10^{-6}$	$1.4 \cdot 10^{-5}$	5.00	$6.8 \cdot 10^{-5}$
Double EMG	$7.7 \cdot 10^{-6}$	$3.6 \cdot 10^{-5}$	5.00	$1.8 \cdot 10^{-4}$
Exponential	0.43	0.43	7.16	2.14

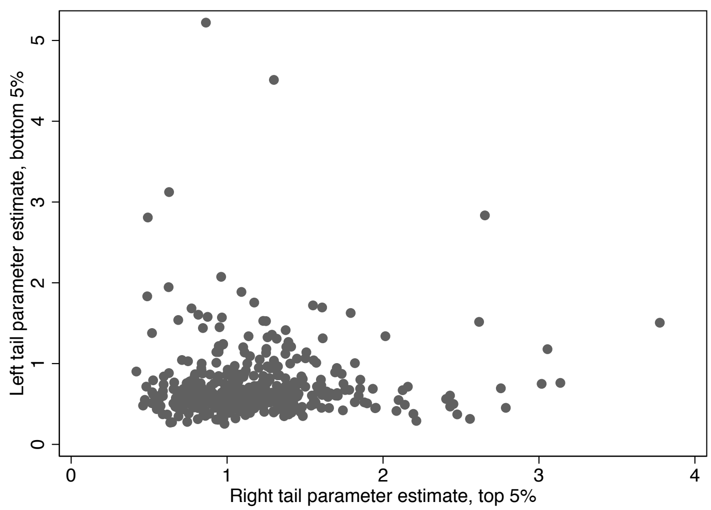
Note: the table reports sample means and standard deviations of the corresponding elasticity estimates for various distributional assumptions. For the Double EMG, and Exponential distributions the means are reported across 81, and 244 observations respectively for which the estimates of λ_R, or the tail index, respectively, are greater than 1. The elasticities are not defined if values of the corresponding parameters are less than 1. To compute the partial trade elasticity, the value of $\epsilon = 6$ is assumed.

Table C3: Distribution specification bias, Peru.

	Normal		Exponential	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Double EMG	0.64	0.42	$2.0 \cdot 10^{8}$	$7.2 \cdot 10^{8}$

Note: the table reports the sample mean and standard deviation of ratios between extensive margin elasticity estimates. The columns indicate the numerator of Normal or Exponential distribution implied extensive margin elasticity estimates, while the rows indicate the denomenator of Double EMG implied extensive margin elasticity estimates.

Figure C1: Heterogeneity in the tail index estimates of log-sales distributions across export destinations, a scatter plot, Peru.



Notes: The figure depicts a scatter plot of the right and left tail index estimates for for the top and bottom 5% of firms. Each dot in the figure corresponds to an estimate of the right and left tail indexes for a given destination-year observation. A sample of 415 destination-year observations where at least 100 first export.

Figure C2: Goodness of fit statistics across each destination-year observation, Peru.

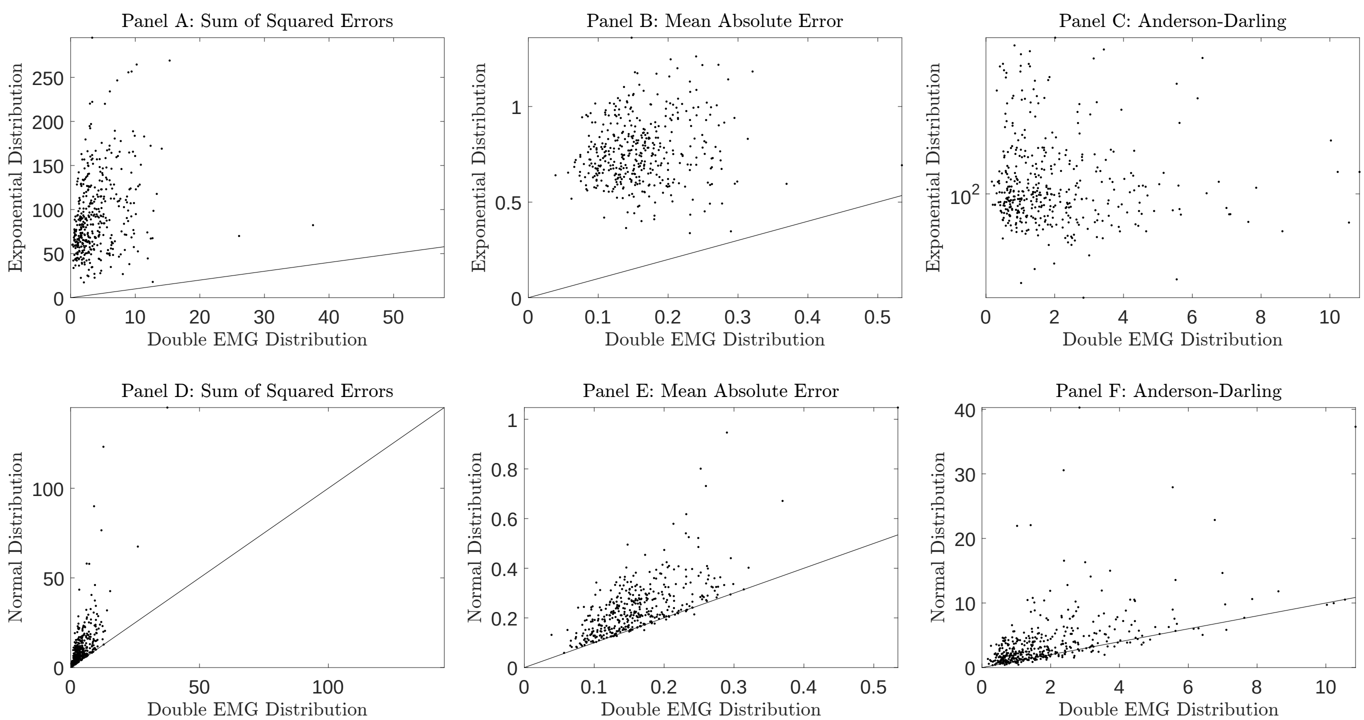
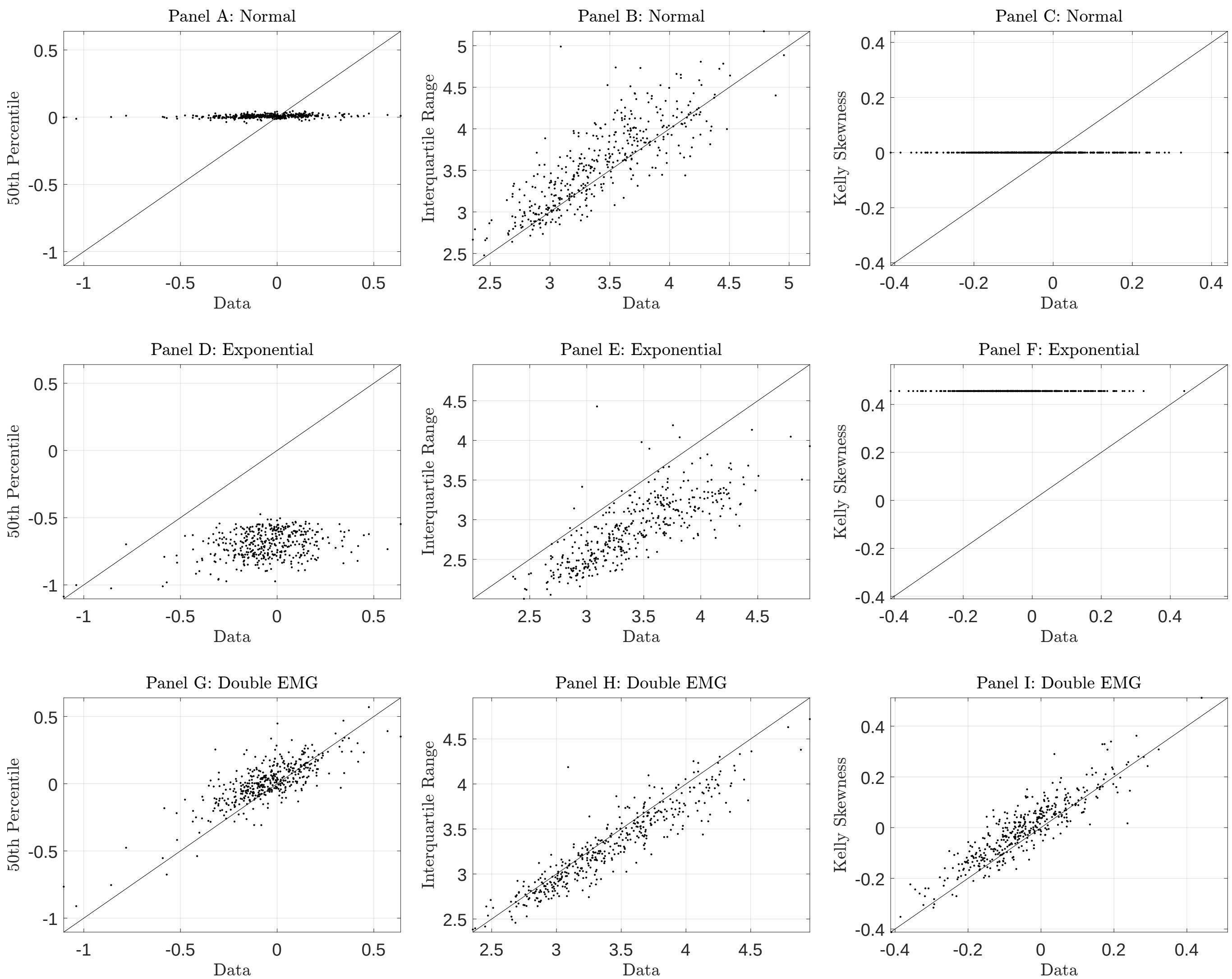
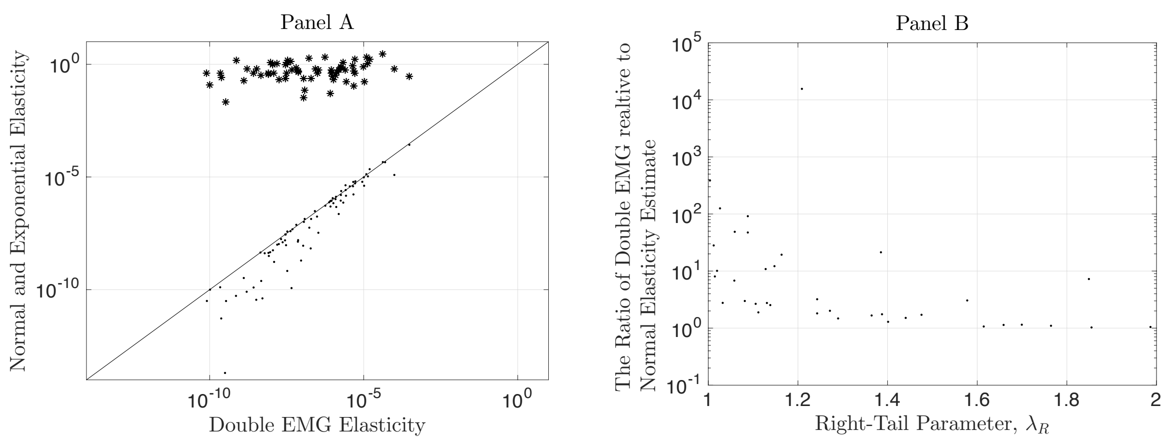


Figure C3: Comparison of empirical to model-generated moments, Peru.



Notes: Scatter plots show 50th percentile, interquartile range and Kelly skewness generated by the Normal distribution (Panels A, B and C), Exponential distribution (Panels D, E and F), and Double EMG distribution (Panels G, H and I) against the empirically observed statistics (x-axis) across destination-year pairs.

Figure C4: Extensive margin elasticity estimates, Peru.



Notes: Panel A of the figure depicts the estimates of the extensive margin elasticity for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. Dots (stars) plot Double EMG against Normal (Exponential) elasticity estimates. Panel B of the figure depicts the ratio of the Double EMG relative to Normal extensive margin elasticity estimates for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. The elasticity is not defined for $\lambda_R \leq 1$.

D Robustness on Sample Selection (For Online Publication Only)

In this appendix, we ask whether our results are driven by the particular way in which manufacturing trade is defined in our paper. Specifically, prior to aggregating the data at the firm-destination-year level, we drop any firm-product-destination-year observations for agricultural products. If a firm simultaneously exports manufacturing and agricultural products, our approach can potentially create an abundance of small firms that might not primarily export manufacturing-industry products. Our dataset does not contain an indicator of a firm's primary industry of operation. Hence, we check the robustness of our results by dropping all firms that export at least one non-manufacturing product within a destination-year. The firms that remain only export manufacturing products.

Across firm-destination-year bins, 10% of firms export any non-manufacturing products. For an average firm, measured as the unconditional mean across firm-destination-years, non-manufacturing products account for 9% of export revenue. However, for those firms that export any non-manufacturing products, revenues are highly concentrated in non-manufacturing products with non-manufacturing products accounting for 92% of export revenue on average. Therefore, the main text included the remaining 8% of export sales from these 10% of firms. This section altogether excludes all sales, manufacturing and non-manufacturing, from these 10% of firms.

Below, we reproduce Table 1 to Table 3 and Figure 4 to Figure 6. We find that all quantitative results are nearly unchanged.

Table D1: Properties of the log-sales distribution across destination-year observations over 1990-2001, sample selection robustness.

Statistic	Mean	Median	Standard Deviation	Min	Max
Panel A: Moment based statistics					
Standard Deviation	2.10	2.11	0.28	1.28	2.76
Skewness	0.03	0.02	0.24	-1.07	1.25
Nonparametric Skew	0.03	0.03	0.06	-0.20	0.20
Kurtosis	3.16	3.03	0.58	2.08	8.17
Panel B: Percentile based statistics					
Interquartile Range	2.81	2.82	0.49	1.49	4.31
Kelly Skewness	0.05	0.05	0.08	-0.30	0.35
Percentile Coefficient of Kurtosis	0.26	0.26	0.02	0.19	0.34
Panel C: Tail parameter estimates					
Top 5%	1.43	1.29	0.62	0.40	6.67
Top 10%	1.19	1.14	0.31	0.48	3.05
Top 15%	1.08	1.05	0.25	0.53	2.75
Bottom 5%	1.22	1.14	0.53	0.44	7.51
Bottom 10%	1.08	1.04	0.29	0.45	3.67
Bottom 15%	1.02	0.99	0.24	0.48	2.77

Note: the statistics are reported across 845 destination-year observations where at least 100 firms export.

Table D2: Trade elasticity estimates, sample selection robustness.

	Extensive Margin Elasticity, γ_{ij}		Partial Trade Elasticity, $\lvert(1-\epsilon)(1+\gamma_{ij})\rvert$	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Normal	$8.5 \cdot 10^{-6}$	$5.0 \cdot 10^{-5}$	5.00	$2.5 \cdot 10^{-4}$
Double EMG	$3.8 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$	5.00	$8.3 \cdot 10^{-4}$
Exponential	0.54	0.61	7.70	3.07

Note: the table reports sample means and standard deviations of the corresponding elasticity estimates for various distributional assumptions. For the Double EMG, and Exponential distributions the means are reported across 279, and 713 observations respectively for which the estimates of λ_R, or the tail index, respectively, are greater than 1. The elasticities are not defined if values of the corresponding parameters are less than 1. To compute the partial trade elasticity, the value of $\epsilon = 6$ is assumed.

Table D3: Distribution specification bias, sample selection robustness.

	Normal		Exponential	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Double EMG	0.62	0.36	$6.8 \cdot 10^{7}$	$5.2 \cdot 10^{8}$

Note: the table reports the sample mean and standard deviation of ratios between extensive margin elasticity estimates. The columns indicate the numerator of Normal or Exponential distribution implied extensive margin elasticity estimates, while the rows indicate the denomenator of Double EMG implied extensive margin elasticity estimates.

Figure D1: Goodness of fit statistics across each destination-year observation, sample selection robustness.

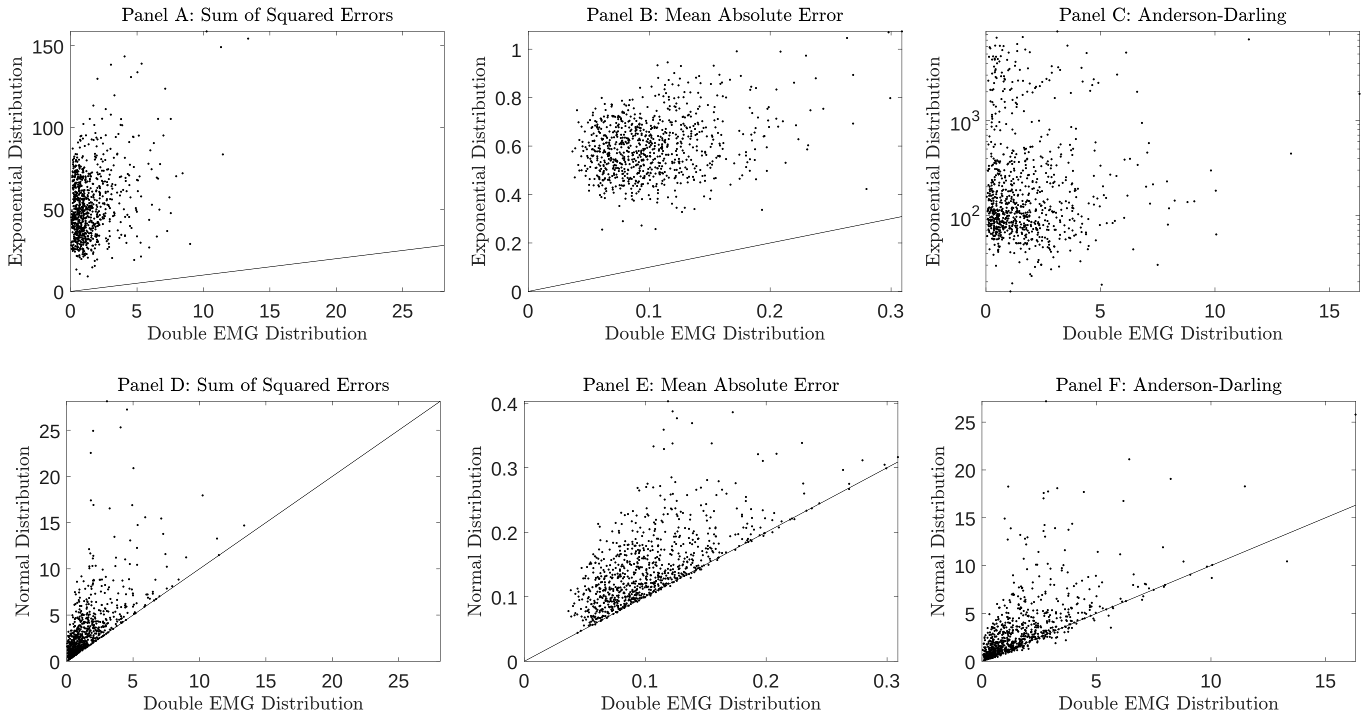
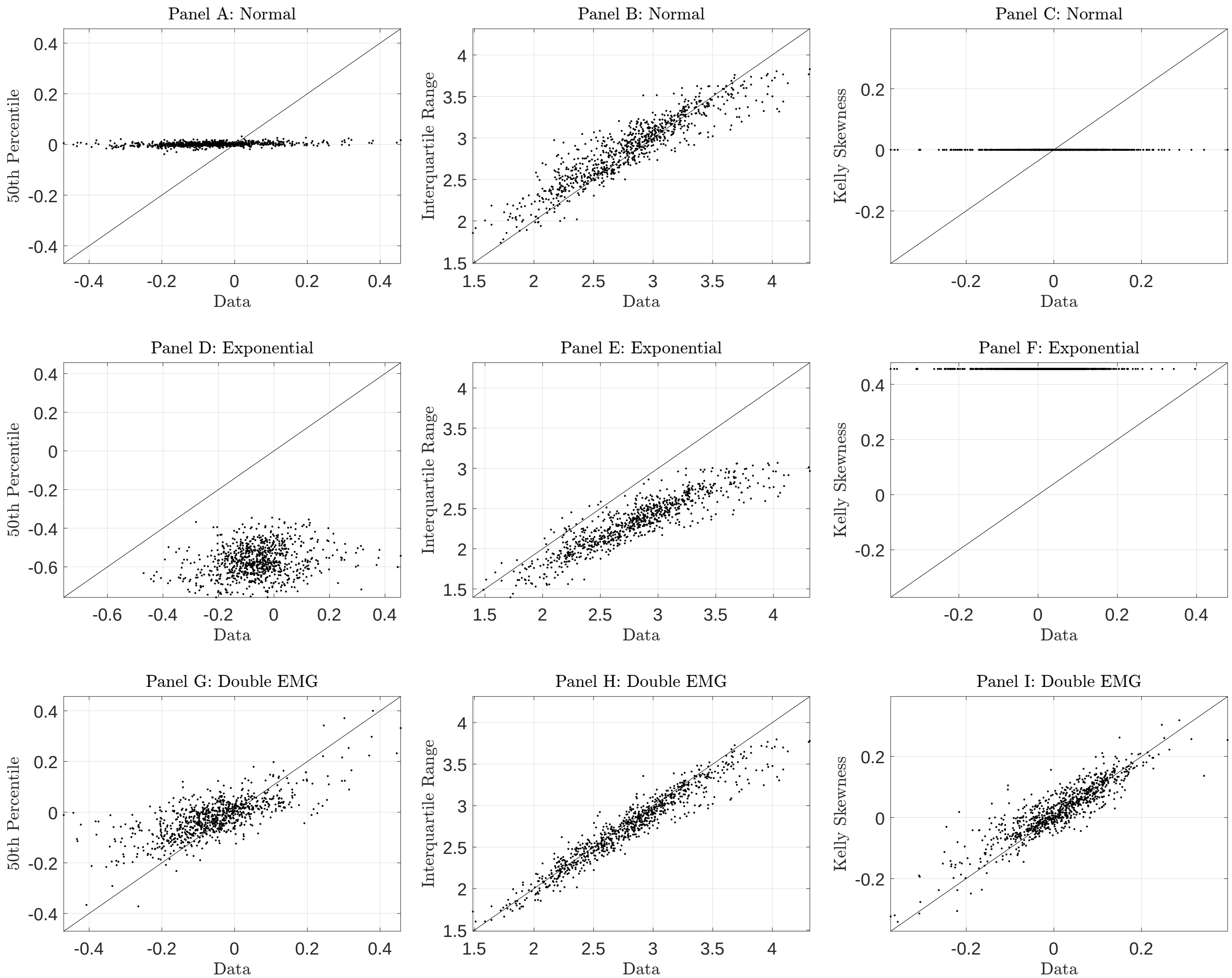
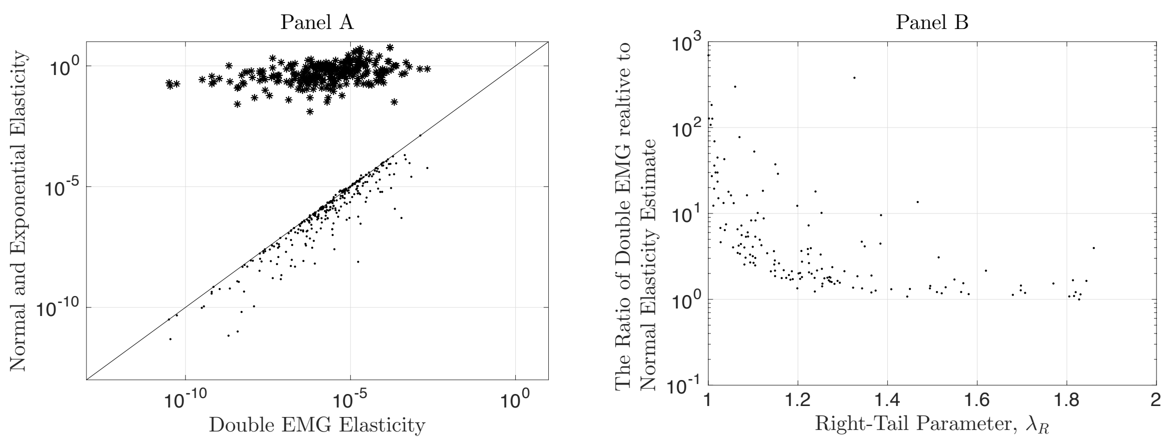


Figure D2: Comparison of empirical to model-generated moments, sample selection robustness.



Notes: Scatter plots show 50th percentile, interquartile range and Kelly skewness generated by the Normal distribution (Panels A, B and C), Exponential distribution (Panels D, E and F), and Double EMG distribution (Panels G, H and I) against the empirically observed statistics (x-axis) across destination-year pairs.

Figure D3: Extensive margin elasticity estimates, sample selection robustness.



Notes: Panel A of the figure depicts the estimates of the extensive margin elasticity for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. Dots (stars) plot Double EMG against Normal (Exponential) elasticity estimates. Panel B of the figure depicts the ratio of the Double EMG relative to Normal extensive margin elasticity estimates for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. The elasticity is not defined for $\lambda_R \leq 1$.

E Robustness on Industrial Composition (For Online Publication Only)

In this appendix, we ask whether the documented properties of export sales distributions across destination-years are systematically related to industry heterogeneity. We find that there is no statistically significant relationship between industry shares and skewness within destination-year observations. Therefore, no single industry drives tail fatness or skewness across destination-years.

Table E1 reproduces results from Table 1 and reports statistics over destination-year observations in which each firm's sales of products within a particular industry are demeaned by that industry's destination-year average. The table shows that controlling for industry composition induces negative skewness, fatter left tails and thinner left tails of destination-year(-industry) observations. Therefore, controlling for industry composition artificially shrinks the size of firms across industries and has no economic significance within the class of trade models studied in this paper.

We reproduce Figure 4 to Figure 6 and show the Double EMG provides a superior fit to export sales data across destination-year observations. This is because left tails become fatter after controlling for industry composition.

Finally, we reproduce Table 2 to Table 3 and Figure 4 to Figure 6. We find that the magnitude of the *distribution specification bias* is quantitatively similar.

Table E1: Properties of the log-sales distribution across destination-year observations over 1990-2001, demeaned by industry.

Statistic	Mean	Median	Standard Deviation	Min	Max
Panel A: Moment based statistics					
Standard Deviation	1.99	2.01	0.27	1.25	2.70
Skewness	-0.31	-0.28	0.28	-1.51	0.40
Nonparametric Skew	-0.04	-0.04	0.06	-0.26	0.14
Kurtosis	3.16	3.01	0.65	2.15	8.12
Panel B: Percentile based statistics					
Interquartile Range	2.70	2.72	0.41	1.62	3.86
Kelly Skewness	-0.04	-0.03	0.09	-0.41	0.21
Percentile Coefficient of Kurtosis	0.27	0.27	0.02	0.18	0.34
Panel C: Tail parameter estimates					
Top 5%	2.25	1.93	3.18	0.63	90.45
Top 10%	1.73	1.63	0.54	0.83	8.69
Top 15%	1.52	1.46	0.40	0.92	5.35
Bottom 5%	1.14	1.05	0.49	0.36	4.44
Bottom 10%	1.02	0.97	0.30	0.38	2.65
Bottom 15%	0.96	0.93	0.24	0.42	2.13

Note: the statistics are reported across 847 destination-year observations where at least 100 firms export.

Table E2: Trade elasticity estimates, industry composition robustness.

	Extensive Margin Elasticity, γ_{ij}		Partial Trade Elasticity, $\lvert(1-\epsilon)(1+\gamma_{ij})\rvert$	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Normal	$1.4 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$	5.00	$2.9 \cdot 10^{-4}$
Double EMG	$2.7 \cdot 10^{-5}$	$2.8 \cdot 10^{-4}$	5.00	0.0014
Exponential	1.26	3.19	11.31	15.95

Note: the table reports sample means and standard deviations of the corresponding elasticity estimates for various distributional assumptions. For the Double EMG, and Exponential distributions the means are reported across 734, and 839 observations respectively for which the estimates of λ_R, or the tail index, respectively, are greater than 1. The elasticities are not defined if values of the corresponding parameters are less than 1. To compute the partial trade elasticity, the value of $\epsilon = 6$ is assumed.

Table E3: Distribution specification bias, industry composition robustness.

	Normal		Exponential	
Distribution	Mean	Std. Dev.	Mean	Std. Dev.
Double EMG	0.54	0.36	$9.1 \cdot 10^{8}$	$9.9 \cdot 10^{9}$

Note: the table reports the sample mean and standard deviation of ratios between extensive margin elasticity estimates. The columns indicate the numerator of Normal or Exponential distribution implied extensive margin elasticity estimates, while the rows indicate the denomenator of Double EMG implied extensive margin elasticity estimates.

Figure E1: Goodness of fit statistics across each destination-year observation, industry composition robustness.

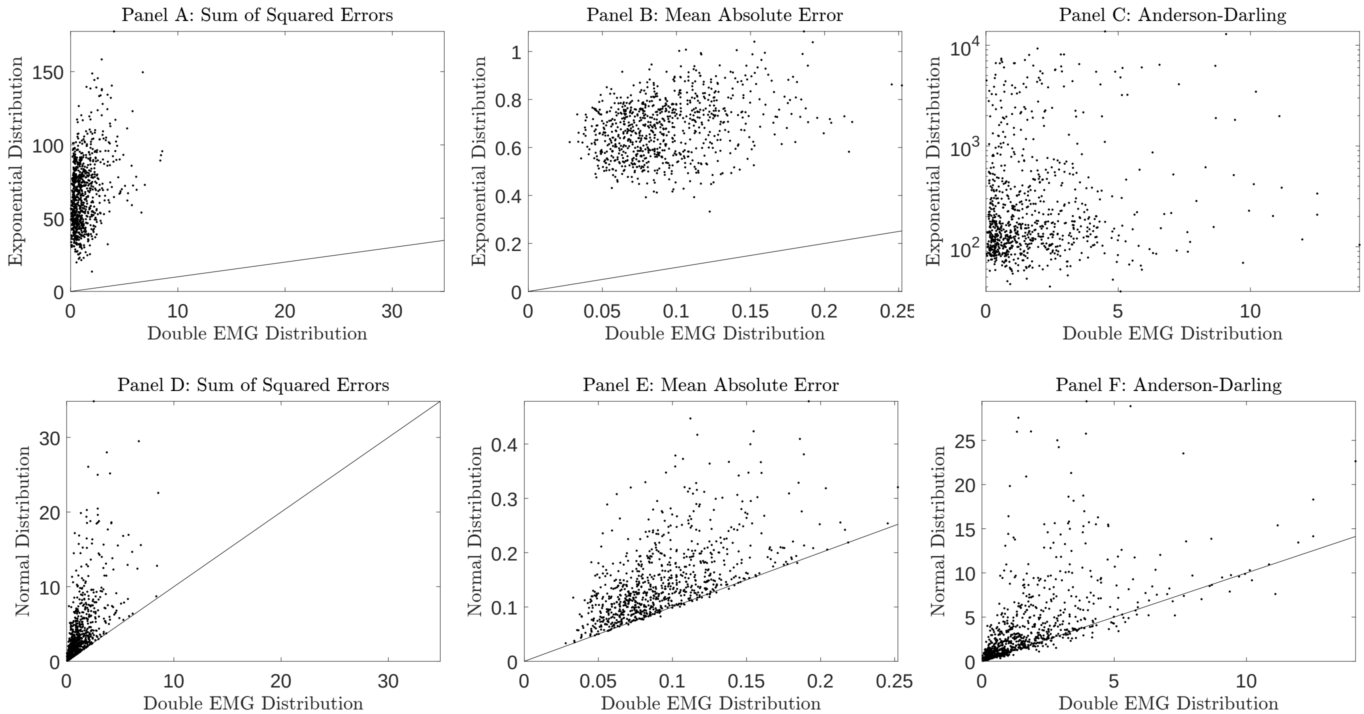
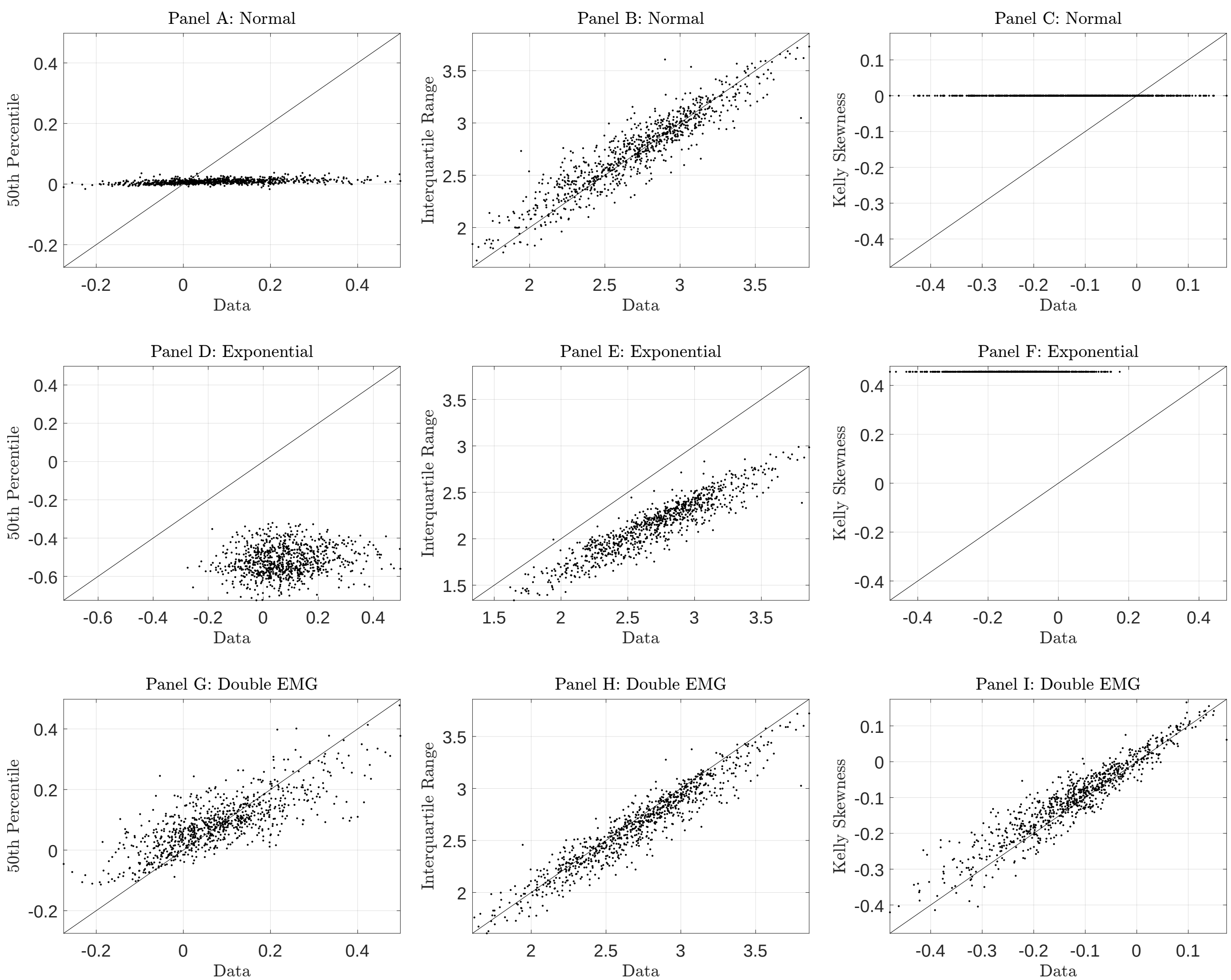
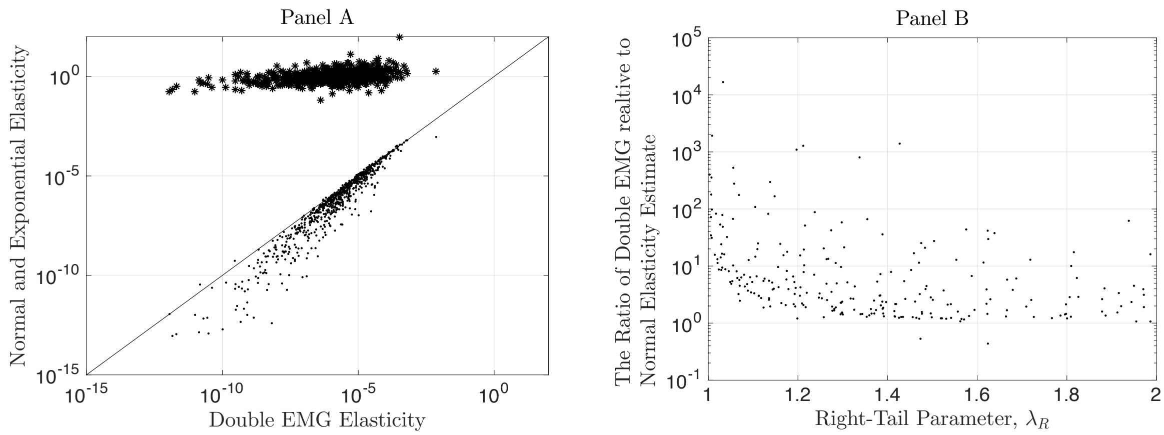


Figure E2: Comparison of empirical to model-generated moments, industry composition robustness.



Notes: Scatter plots show 50th percentile, interquartile range and Kelly skewness generated by the Normal distribution (Panels A, B and C), Exponential distribution (Panels D, E and F), and Double EMG distribution (Panels G, H and I) against the empirically observed statistics (x-axis) across destination-year pairs.

Figure E3: Extensive margin elasticity estimates, industry composition robustness.



Notes: Panel A of the figure depicts the estimates of the extensive margin elasticity for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. Dots (stars) plot Double EMG against Normal (Exponential) elasticity estimates. Panel B of the figure depicts the ratio of the Double EMG relative to Normal extensive margin elasticity estimates for destination-year observations with a Double EMG estimate of the tail parameter $\lambda_R > 1$. The elasticity is not defined for $\lambda_R \leq 1$.