Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University

The EMG Distribution and Aggregate Trade Elasticities

IIEP-WP-2016-15

Olga A. Timoshenko
George Washington University

Erick Sager
U.S. Bureau of Labor Statistics

May 2016

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **www.gwu.edu/~iiep**

The EMG Distribution and Aggregate Trade Elasticities*

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

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

May 9, 2016

Abstract

When firm-level productivity is not assumed to be Pareto distributed, new trade models predict that micro-data such as sales distributions determine the aggregate trade elasticity. In this paper, we document novel features of export sales distributions across destination markets. Notably there is large variation in dispersion and degrees of asymmetry. To capture these features of the export sales data we introduce a novel distribution: the Exponentially Modified Gaussian (EMG) distribution. We show that the EMG distribution fits sales data far better than either of the often assumed log-Normal or Pareto distributions and we provide quantitative evidence that these less accurate distributions can generate highly biased trade elasticities.

1 Introduction

The aggregate elasticity of trade with respect to variable trade costs is an important component in measuring the welfare gains from trade. In the context of new trade theories with firm-level heterogeneity (most notably Melitz (2003)), this trade elasticity is intimately linked to the distributional assumption made with respect to the fundamental sources of firm heterogeneity, as shown by Melitz and Redding (2015). Accordingly, one way to estimate a trade elasticity is by fitting a model-implied parametric distribution of export sales to the data. When making such an inference about the aggregate trade elasticity from micro data, it is crucial that the theoretical distribution is able to accurately characterize the sales data.

*The authors thank Mina Kim, Illenin Kondo, Logan Lewis, Toshi Mukoyama and seminar participants at the Census Bureau and the GWU Trade Study Group for insightful comments and discussions. 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.

In this paper, first we document new features of log-sales distributions of exports that are inconsistent with the salient features of the two most frequently utilized distributional assumptions in workhorse trade models: the Normal and Exponential distributions. In the context of Brazilian export data for the period between 1990 and 2001 we document a number of novel striking features of log-sales distributions across export destinations. First, in contrast to a Normal distribution, the majority of the log-sales distributions are not symmetric. Not only do statistical tests reject the assumption of Normality in 42 percent of log-sales distributions across destinations, but nearly 75 percent of distributions are positively skewed. Second, in contrast to an Exponential distribution which has constant skewness of 2, the skewness of log-sales in the Brazilian data varies between -1.08 and 1.29 with an average value of 0.03 being far below 2. Finally, the dispersion of log-sales varies substantially across destinations. For example, if across all destinations the 25th percentile firm always made 100 dollars in sales, then the size of the 75th percentile firm would vary between 450 dollars and 8,000 dollars across destinations (by a factor of 18).

In order to capture the prevalent asymmetry in the distribution of log-sales we propose a novel distribution: the Exponentially Modified Gaussian (EMG) distribution. An EMG distribution is constructed as a convolution of two independent distributions: a Normal and an Exponential. As such, an EMG distribution exhibits a fat, Pareto-like right-tail and a Normal-like left-tail. Hence, the EMG distribution has the potential to capture, in a parsimonious manner, the asymmetric nature of the empirical log-sales distributions as well as generate very small sales in the left tail. We further demonstrate that the novel EMG distribution arises from models of learning such as Timoshenko (2015), where firms draw their productivity from a Pareto distribution and a demand shock from a Normal distribution.

We next demonstrate that an EMG distribution provides a superior fit to the log-sales data. We fit an EMG distribution to the empirical log-sales distributions across export destinations and compare the fit of the EMG to the fit of Normal and Exponential distributions. Across various goodness of fit measures, the EMG distribution provides a superior fit in a large majority of destinations. Hence, models that assume an EMG distribution will match the micro-data better than models that incorporate either a Normal or an Exponential distribution alone.

Distributional assumptions have important implications for the measured magnitudes of the aggregate trade elasticity. In the context of the Melitz (2003) model with monopolistic competition, heterogeneous firms, and CES preferences, the trade elasticity can be decomposed into the intensive and the extensive margin components. The extensive margin arises from the entry and exit of new firms into an export destination, while the intensive margin arises from changes in sales of existing firms. We further demonstrate that the *level* of trade elasticity depends on the elasticity of substitution across varieties. The *relative contribution* of the intensive margin to that level is always unity, while the relative contribution of the extensive margin depends on the entire shape of the log-sales distribution and the size of an average firm relative to the smallest firm.

Accordingly, we use the fitted distributions to evaluate the contribution of the extensive and intensive margins to the aggregate trade elasticity in the context of the Melitz (2003) model. When a log-sales distribution deviates from a Normal distribution and is more accurately characterized by an EMG distribution, the Normal distribution underestimates the extensive margin elasticity by a factor of 10 to 1,000. The more fat-tailed is the log-sales distribution, the more skewed is the fitted EMG distribution, and the larger is the extent of the bias generated by fitting a Normal distribution.

Finally we point out an important issue in estimating extensive margin elasticities using truncated export data. It is not uncommon in collecting trade and customs data 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 per firm-destination sales. While our data is not truncated, we create a counterfactual dataset that is, and study the effect of truncation on measured aggregate trade elasticities. We show that the extensive margin elasticities implied by the truncated data are overestimated by on average factor of 10^5. For example, when truncated data is used, the contribution of the intensive relative to the extensive margin is 1 to 0.1, while when a full sample is used the ratio is 1 to 10^{-5}. Therefore, truncated samples overestimate the contribution of the extensive margin to trade flows and, hence, to welfare.

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 the unity. In contrast, recent work however has argued that sales distributions are not well characterized by Zipf's law alone. 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.

This paper contributes to the firm size distribution literature by demonstrating that the Exponentially Modified Gaussian distribution fits the sales data better than either the Normal distribution or Zipf's law. Being a convolution of a Normal and an Exponential distribution, the EMG can simultaneously match power law distributed right tails and Normally distributed left tails of a standard log-sales distribution.

Further, in the context of Portuguese domestic manufacturing data, Cabral and Mata (2003) document that the firm-size distribution is positively skewed. Bastos and Dias (2013) extend this result to Portuguese exporting firms. In our work, we show that the asymmetric nature of the data is also pronounced in the distribution of export sales. In the Brazilian export data, the majority of destination-level exports are positively skewed and the degree of skewness varies across destination. We demonstrate that an 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.

This paper also contributes to the literature on the interaction between firm-level heterogeneity and the gains from trade. Melitz and Redding (2015) demonstrate that once firm-level heterogeneity is no longer governed by a Pareto distribution, the aggregate 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 Exponentially Modified Gaussian distribution and find that the data favors the EMG over either the Normal or Exponential (Pareto). This finding has implications for the aggregate trade elasticity, that we develop herein.

We further add to the theoretical literature that provides microfoundations for empirically observed distributions, such as power laws in the firm size distribution (see Gabaix (2009) for a extensive review). We can show that the EMG distribution arises from a broad set of model ingredients, such as Pareto distributed firm-level productivity and log-Normally distributed product demand (see Timoshenko (2015) and Arkolakis et al. (2015)), or such as Pareto distributed firm-level productivity and log-Normally distributed fixed costs (see Eaton et al. (2011)). Log-sales are EMG distributed whenever the log-revenue function is linear in random shocks, and those shocks consist of independent Normal and Exponential components. In fact, it could be shown that if log-productivity follows a Brownian motion and reflecting barriers are Normally distributed across firms, then the resulting steady state distribution is an Exponentially Modified Gaussian distribution. This result is reminiscent of Champernowne (1953), in which the reflecting barrier has a point mass.

In related work, Mrázová, Neary, and Parenti (2016) show that replacing standard CES preferences with a "Constant Revenue Elasticity of Marginal Revenue" demand system generates a type of equivalence between log-Normal or Pareto productivity distributions. While our papers share a focus on altering model ingredients in order to better account for empirical observations, our paper focuses on assumptions governing distributions instead of preferences.

The Exponentially Modified Gaussian distribution has also been used in various recent macroeconomic applications. Heathcote and Tsujiyama (2015) use the EMG to model idiosyncratic earnings in an incomplete markets model with optimal taxation. Badel and Huggett (2014) employ the EMG in a similar model to capture the skewness in log-earnings distributions, as the EMG fits the cross-sectional earnings distribution better than a conventionally used Normal distribution. In a somewhat different context, Toda and Walsh (2015) use a generalization of the 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 facts about the properties of log-sales distributions across markets. Section 3 constructs a novel distribution, the Exponentially Modified Gaussian distribution, and characterizes its properties. Section 4 fits theoretical distributions (EMG, Normal and Exponential) to empirical export log-sales distributions and evaluates goodness of fit. Section 5 examines the implications of the distributional assumptions for estimates of the aggregate trade elasticity. Section 6 concludes. The proofs to all propositions are included in Appendix A.

2 Empirical Facts

The data come from the Brazilian customs declarations collected by SECEX (*Secretaria de Comercio Exterior*).[1] 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.[2] 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.[3]

An observation is an entire distribution of log-sales for a given destination in a given year. The final sample, hence, consists of 847 destination-year distributions of log-sales. We compute various moments of a distribution for each destination-year observation and present a set of stylized facts describing how properties of distributions vary across destination-year observations.

Table 1 presents various measures summarizing the dispersion and skewness of the log-sales distribution across destination-year observations. Each row of Table 1 reports a mea-

[1]See Molinaz and Muendler (2013) for a detailed description of the dataset.

[2]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.

[3]The same threshold is also used in the work of Fernandes et al. (2015).

sure's mean, minimum and maximum values as well as its standard deviation. To provide a comprehensive overview of the differences in the properties of distributions we present moment based (Panel A) and equivalent percentile based (Panel B) characterizations.

We focus our subsequent analysis and discussion on the percentile based characterizations for two reasons. First, percentile based measures are less sensitive to outliers and extreme values. Second, for any theoretical distribution percentiles are always defined and are finite. Hence they can potentially be computed and compared to the data. That is not always the case for moment based characterizations of a distributions. For example, consider a Pareto distribution widely used in the literature. If the scale parameter of a Pareto distribution falls below the value of two, the variance (a moment based measure) is infinite while the interquartile range (an equivalent percentile based measure) is well defined. The moment based characterizations are reported for completeness. Table 1 indicates that log-sales distributions vary substantially across destination-year observations.

Fact 1 *The dispersion of log-sales varies by a factor of 3 across destination-year observations.*

Consider the dispersion of log-sales as measured by the standard deviation and the interquartile range (IQR). The IQR is defined as the difference between the value of the 75th and the 25th percentiles of the log-sales distribution. Intuitively the IQR is a measure of heterogeneity across firms in the data. It measures by how much larger a representative large firm (the 75th percentile) is relative to a representative small firm (the 25th percentile). In the data, we observe large variation in the IQR across destination-year observations, ranging from 1.50 to 4.44. Accordingly, the destination with the most dispersed sales exhibits nearly 3 times more heterogeneity than the least dispersed destination. Figure 1 presents histograms of the values of the standard deviation (left Panel) and the interquartile range (right Panel) across destination-year observations and provides further visual evidence that the dispersion in log-sales varies substantially across destination-year observations.

In order to understand the quantitative significance of this level of dispersion, consider the following example. Suppose the 25th percentile firm in a hypothetical market received \$100 in sales. In the destination-year observation with the maximum IQR (=4.44), the 75th percentile firm's sales were $100 \times e^{4.44}$ or approximately \$8,000. Compare that to the value of $100 \times e^{1.50} = \$450$ in sales for the minimum IQR or $100 \times e^{2.88} = \$1,600$ in sales for the average destination-year observation.

At the 90th percentile and above, however, we observe firms with much larger sales, typically upwards of one million dollars. In order to more precisely describe such asymmetry in sales, next we present measures of skewness.

Fact 2 *Across destinations, a large majority of log-sales distributions are not symmetric.*

We consider three different measures of skewness. The first is the standard moment-based measure of skewness. The second is a nonparametric skew defined as the difference between the mean and the median of a distribution divided by the standard deviation. The third is the 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.[4] As can be seen in Table 1, the average sample values of the skewness, nonparametric skew, and the Kelly skewness are positive. All three statistics have mean values that are greater than zero with a maximum p-value of 0.0003. Furthermore, according to these measures, 51% to 75% of log-sales distributions exhibit positive skew. Among the 847 destination-year observations, 54% have positive skewness, 71% have positive nonparametric skew, and 75% have positive Kelly skewness. Figure 2 presents histograms of the skewness measures across destination-year observations and provides further visual evidence that the majority of log-sales export distributions are positively skewed.

In order to better understand the important of skewness in the export sales distributions, consider the following explanation of the percentile-based Kelly skewness measure. Kelly skewness measures length of the left tail (defined as $P90 - P50$) relative to the length of the right tail (defined as $P50 - P10$) of a distribution. For example, using the definition in equation (1), a Kelly skewness of 1/3 implies that the 90th percentile is twice as far away from the 50th than the 10th percentile, i.e. a distribution is heavily skewed to the right. The sample average of the Kelly skewness of 0.04 implies that the 90th percentile is 8% further from the 50th than the 10th, which is a non-negligible positive skew. Furthermore, the histogram in Panel C of Figure 2 shows that the average Kelly skewness is not due to a large mass around 0.04, but instead an average of both large and positive and negative skewness values.

The asymmetric nature of the log-sales distributions is confirmed further through a standard test of Normality as described in D'Agostino et al. (1990). The test for normality of a distribution based on skewness alone is rejected in 31% of observations at the 10-percent significance level, in 24% at the 5-percent significance level, and in 16% at the 1-percent significance level. The test for normality based on skewness and kurtosis is rejected in 42%

[4]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.

of observations at the 10-percent significance level, in 32% at the 5-percent significance level, and in 20% the 1-percent significance level.

Fact 3 *Over the period in which Brazilian enacted trade reforms, log-sales distributions became more dispersed and skewed.*

During the 1990s, Brazil went through a period of liberalization reforms.[5] During this time, total exports grew by 6% per year on average and the number of exporters doubled between 1990 and 2001. In order to better understand the effect of these economic changes on Brazilian export sales, we study the following empirical relationships:

$$y_{it} = \beta\gamma_t + \alpha_i + u_{it}. \tag{2}$$

In equation (2) we regress a variable of interest y_{it} on a time trend γ_t and a set of destination fixed effects α_i. The results are presented in Tables 2 and 3. In Table 2, the dependent variables are the standard deviation of log-sales (column 1) and the interquartile range of log-sales (column 2). In Table 3, the dependent variables are the skewness of log-sales (column 1), the nonparametric skew of log-sales (column 2), and the Kelly skewness of log-sales (column 3). With the exception of the inter-quartile range, all measures are increasing with time within destination cells. Hence, over time, during the period of Brazilian economic reforms in the 90's, the distributions of log-sales exhibit an increase in dispersion and skewness.

An increase in dispersion is consistent with predictions from models of export market participation, such as Melitz (2003). As trade costs decline, more firms enter. New firms are less productive compared to incumbents and will receive lower export sales revenues. Therefore, the entry of small firms will increase the variance of the log-sales distribution.

In the next section we introduce and characterize a novel distribution, 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.

3 The Exponentially Modified Gaussian (EMG) Distribution

The EMG distribution is defined as a convolution of a Normal distribution and an Exponential distribution. As a result, one of the key properties of the distribution is its differential

[5]See Dix-Carneiro and Kovak (2015) for a detailed description of the reforms.

behavior in the right and left tails. The distribution exhibits a Normally-distributed left tail and a Exponentially-distributed right tail. Hence, it is particularly well suited to fit empirical regularities in the log-sales export data. In this section we derive some of the key properties characterizing the distribution including its behavior in the right and left tails.

Consider a random variable z defined as $z = x + y$, where x and y are two independent random variables. Assume $x \sim N(\mu, \sigma^2)$, and $y \sim Exp(\lambda)$. In this case, random variable z is a convolution of a Normal and an Exponential random variables and is said to follow an Exponential Modified Gaussian (EMG) distribution with parameters (μ, σ, λ). Proposition 1 below describes the EMG distribution with its cumulative distribution function, probability density function, and the moment generating function.

Proposition 1 *Let x and y be independent random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim Exp(\lambda)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma \geq 0$ and $\lambda \geq 0$. The random variable $z \equiv x + y$ has the 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:

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

The proofs to all propositions are included in Appendix A.

The EMG distribution generalizes both the Normal and Exponential distributions. Consider the variance of an Exponential distribution with scale parameter denoted by λ, which is λ^{-2}. By increasing the value of λ 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 EMG distribution can be transformed into a Normal distribution when its Exponential distribution has zero variance ($\lambda \to +\infty$) or transformed into a Exponential distribution when its Normal distribution has zero variance ($\sigma \to 0$). Furthermore, the Exponential distribution controls the mass in the right tail. We provide formal justification for these claims in the Proposition 2 below.

Proposition 2 *(Limiting Results) 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:

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

The conditional expectation of the EMG distribution is described in Proposition 3 below.

Proposition 3 *If z is an Exponentially Modified Gaussian distributed random variable on $(-\infty, +\infty)$ then the conditional first moment on $(z^*, +\infty)$ is*

$$\int_{z^*}^{+\infty} e^z g(z) dz = M_z(1) \left[1 - \Phi\left(\frac{z^* - \mu - \sigma^2}{\sigma}\right) + e^{-(\lambda-1)(z^*-\mu-\lambda\sigma^2) - \frac{1}{2}(\lambda-1)^2\sigma^2} \Phi\left(\frac{z^* - \mu - \lambda\sigma^2}{\sigma}\right) \right].$$

4 Theoretical Export Sales 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 functional fit of the Exponentially Modified Gaussian, Normal and Exponential distributions. We show that the Exponentially Modified Gaussian distribution has a superior fit to the data. 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 identify 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, non-linear regression is necessary for calibrating the parameters of the Exponentially Modified Gaussian distribution, since the Exponentially Modified Gaussian distribution does not admit a linear quantile function (as can be inferred from Proposition 1). For the Normal and Exponential distributions, our procedure recovers 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.[6] 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 Exponentially Modified Gaussian distribution, the parameter vector is $\Theta = (\mu, \sigma, \lambda)$ such that $q \sim f(q|\mu, \sigma, \lambda)$. However, the inverse of the Exponentially Modified Gaussian distribution does not admit a closed form expression. Therefore, the inverse of the Exponentially Modified Gaussian must be computed numerically.

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

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

the sum of the squared errors between empirical and theoretical quantiles:

$$\min_{\Theta} \sum_{i=1}^{n_q} \left(q_i^d - q_i(\Theta)\right)^2. \tag{3}$$

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 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 the same parameter estimates for the Normal and Exponential distributions as those authors' method.

4.2 Evaluating Distribution Fit

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

Result 1 *Across multiple goodness of fit statistics, the Exponentially Modified Gaussian distribution fits the data better than the Normal and Exponential distributions.*

We first argue that the Exponentially Modified Gaussian distribution fits the data better by examining fitted distribution functions versus their empirical counterparts. Panel A of Figure 3 shows the deviation of *cross sectional destination-year average sales* of theoretical from empirical distributions at percentiles 5 through 95 of the distribution. We observe that the Exponentially Modified Gaussian distribution deviates from the data less than the Normal distribution. This is especially true at the mid to upper percentiles. Panel B compares the left tail across the empirical, Exponentially Modified Gaussian and Normal distributions. Specifically, the plot shows the log of the cumulative distribution functions, which magnifies differences since the natural logarithm of a a small number is large and negative. We observe that the Normal and Exponentially Modified Gaussian distributions are very similar and only deviate from the empirical distribution at the zeroth percentile. Panel C compares the left tail across distributions. We observe that the Exponentially Modified Gaussian distribution barely deviates from the empirical distribution up to the 99th percentile. On the other hand, the Normal distribution is too thin relative to the data.

The second row in Figure 3 provides a stark and illustrative destination-year observation. Not only does the Exponentially Modified Gaussian distribution deviate little from the empirical distribution in the 5th through 95th percentiles, but it fits the left and right tails of the distribution very closely. Only at the zeroth percentile is there a sizable deviation. In comparison, the Normal distribution overpredicts the data in the middle of the distribution and underpredicts in tails. The Normal distribution's right tail is especially thin compared to the empirical right tail.

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 the Exponentially Modified Gaussian, Normal and Exponential distributions. Specifically, the Figure presents scatter plots of goodness of fit measures from the Normal distribution (top row) or the Exponential distribution (bottom row) on goodness of fit measures for the Exponentially Modified Gaussian distribution.

The first measure is the sum of squared errors, which is given by the objective criterion from the estimation procedure given in equation (3) when evaluated at the error-minimizing parameters. Panel A of Figure 4 shows that errors are larger for the Normal and Exponential distributions than the Exponentially Modified Gaussian distribution. This is unsurprising, since the Exponentially Modified Gaussian distribution nests both the Normal and Exponential distributions as limiting cases (see Proposition 2). More interesting is the fact that both of the scatter plots in the first column (Panels A and D) show that the errors are much larger, quantitatively speaking, for the Normal and Exponential distributions. However, the magnitude of the difference in errors is smaller for the Normal distribution.

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

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

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 & E) of Figure 4 shows that errors are larger for the Normal and Exponential distributions than the Exponentially Modified Gaussian distribution. Therefore, the Mean Absolute Error reinforces that the Exponentially Modified Gaussian 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.[7] 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 Exponentially Modified Gaussian 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 Exponentially Modified Gaussian distribution routinely fits the log-sales distributions better across destination-year observations, and that the Normal and Exponential distributions routinely fit the data worse in the tails of the distribution.

Result 2 *Only the Exponentially Modified Gaussian distribution can match the empirical dispersion in skewness.*

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. Figures 5, 6 and 7 plot theoretical moments from the estimated distributions against the empirically observed moments.

Figure 7 compares Kelly skewness in the Exponentially Modified Gaussian distribution (Panel A), Normal distribution (Panel B) and Exponential distribution (Panel C) to the data. It is immediately clear that the Exponentially Modified Gaussian distribution is the only distribution that exhibits variation in skewness across destination-year observations. Furthermore, it captures positive skewness well.[8]

---

[7] 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.

[8] However, because the Exponentially Modified Gaussian distribution cannot generate negative skewness, it cannot match all of the variation in the data. Our future work will relax this feature of the Exponentially Modified Gaussian distribution.

Figure 6 compares 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. Figure 5 shows the comparison with the median of the empirical and theoretical distributions. Only the Exponentially Modified Gaussian 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.

Result 3 *There is substantial parameter heterogeneity across destination-year observations.*

To round out the section, we characterize the dispersion in estimated parameters across destination-year observations. First consider Figure 8, which provides histograms for the Exponentially Modified Gaussian distribution parameters (in solid black) and the Normal distribution parameters (in dotted red). Note that we exclude the Normal distribution's μ parameter from the Panel A of Figure 8 since it will introduce spike at zero that will dwarf the scale of the μ parameter for the Exponentially Modified Gaussian distribution. In Panel A and B, we observe that there is a large dispersion in μ and σ parameters for the Exponentially Modified Gaussian distribution. Notably, in Panel A we observe a spike near zero, which reflects the mass of destination-year observations that the Exponentially Modified Gaussian distribution estimation recovers a Normal distribution. In the Panel B, we observe that the Normal distribution has similar dispersion in σ parameters, but that the point estimates are larger on average. Lastly, Panel C provides a histogram for the Exponentially Modified Gaussian distribution's tail parameter, λ. We observe a spike near one, which is Zipf's law and a large mass of estimates below one.[9] There is also a smaller mass of tail parameters above five, which are the mass of destination-year observations that the Exponentially Modified Gaussian distribution identifies as normally distributed.

Lastly, Figure 9 represents the differences between distribution parameters using a series of scatter plots. In Panel A we again see that the parameter μ does not vary in the Normal distribution, while μ is highly dispersed for the Exponentially Modified Gaussian distribution. Panel B shows that the parameter σ is similarly dispersed in the Normal and Exponentially Modified Gaussian distributions. However, we observe that σ under the Normal distribution is systematically larger, which reflects how the Normal distribution compensates for its inability to match skewness in the data by increasing its variance. Panel C shows

[9]It is well known that the as the tail parameter approaches Zipf's law, there no longer exist well defined first and higher order moments of the Pareto distribution. The Exponentially Modified Gaussian distribution inherits this issue from the Pareto distribution. Truncating the Exponentially Modified Gaussian distribution resolves this issue.

the parameter λ in the Exponential and Exponentially Modified Gaussian distributions, and shows that the parameter λ under the Exponential distribution is systematically smaller. This feature of the estimation helps the Exponential distribution match both middle and upper percentiles simultaneously, which underscores a tradeoff the Exponential distribution faces in fitting to the empirical distributions.

5 Aggregate Trade Elasticities

In this section we consider the standard workhorse model of trade and demonstrate how its implications are both theoretically and quantitatively affected by varying assumptions on fundamental sources of firm heterogeneity.

5.1 Theoretical Framework

We consider an economic environment in which 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). In this environment, the sales of a firm from country i to country j are given by

$$r_{ij}(z_{ij}) \;=\; \left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}(\tau_{ij}w_i)^{1-\epsilon}\,Y_j P_j^{\epsilon-1}e^{z_{ij}}, \tag{4}$$

where ϵ is the elasticity of substitution, τ_{ij} is iceberg transportation costs, w_i is the wage rate in country i, and Y_j and P_j are the income level and the price level in country j.[10] We will refer to z_{ij} as a firm's *profitability* in market j originating from country i. Notice, that when taking logs of both sides of equation (4), the distribution of log-sales, $\log(r_{ij})$, is entirely governed by the distribution of z_{ij} and is shifted by an origin-destination specific constant $C_{ij} \equiv \log\left(\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}(\tau_{ij}w_i)^{1-\epsilon}\,Y_j P_j^{\epsilon-1}\right)$:

$$\log(r_{ij}) = C_{ij} + z_{ij}. \tag{5}$$

The trade literature varies in terms of the assumptions made with regards to the sources and the nature of heterogeneity in the firms' profitability z_{ij}. The general representation in equation (4) encompasses the following three types of assumptions on fundamentals made in the literature.

Assumption 1: Melitz (2003) assumes that the underlying source of heterogeneity in prof-

[10] A complete description and exposition of the model is presented in Appendix B.

itability arises from heterogeneity in labor productivity across firms, denoted by φ. Chaney (2008) further assumes that firm-level labor productivity is drawn from a Pareto distribution with a shape parameter denoted by ξ. In this case, $e^{z_{ij}}$ is equal to $\varphi^{\epsilon-1}$, and by a change of variables z_{ij} is distributed according to an Exponential distribution with a shape parameter $\lambda = \xi/(\epsilon - 1)$. One can fit an Exponential distribution to log-sales data to get an estimate for λ, and then recover the underlying ξ for a given value of the elasticity of substitution.

Assumption 2: 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, $\log N(m, \upsilon^2)$. In this case, $e^{z_{ij}}$ also equals $\varphi^{\epsilon-1}$, and z_{ij} follows a Normal distribution, $N(\mu, \sigma^2)$ where $\mu = m(\epsilon-1)$ and $\sigma^2 = \upsilon^2(\epsilon-1)^2$. Similarly, one can fit a Normal distribution to log-sales data to get an estimate of μ and σ^2, and use the above equation and an assumed value for the elasticity of substitution to recover the underlying m and υ^2.[11]

Assumption 3: A third set of assumptions originates from the literature on learning and firm-level export decisions. Timoshenko (2015) assumes that there are two separate sources of heterogeneity in a firm's profitability: a labor productivity φ drawn from a Pareto distribution with a shape parameter ξ, and a demand shock e^{θ}, where θ is drawn from a Normal distribution $N(m, \upsilon^2)$. 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 demand shock, and that the demand shock does not vary over time (see Appendix B). In this case, $e^{z_{ij}}$ equals $e^{\theta}\varphi^{\epsilon-1}$. Notice that when taking logs we obtain:

$$z_{ij} = \theta + \log\left(\varphi^{\epsilon-1}\right). \tag{6}$$

In this simplified learning model, a firm's profitability is a sum of a Normal and an Exponential random variable. Hence, z_{ij} is a random variable that is EMG distributed with parameters (μ, σ, λ), where $\mu = m + (\epsilon - 1)/\xi$, $\sigma^2 = \upsilon^2$, and $\lambda = \xi/\epsilon$. One can fit an EMG distribution to log-sales data to estimate μ, σ^2, and λ, and use an assumed value of the elasticity of substitution to recover the fundamental parameters.[11]

The learning model described above, along the lines of Timoshenko (2015), provides a micro-foundation for an EMG distribution governing log-sales. An EMG distribution arises from environments in which two sources of fundamental uncertainty (that are distributed according to a Normal and Exponential) additively determine the variation in firm-level log-revenues.

[11] Given that the log-sales distribution is shifted by C_{ij} (see equation (5)), the means are recovered up to a constant. The second and higher order moments are not affected by such transformation.

5.2 Trade Elasticity

The aggregate trade flow from country i to country j is given by

$$X_{ij} = M_{ij} \int_{z^*_{ij}}^{+\infty} r_{ij}(z) \frac{g_{ij}(z)}{1 - G_{ij}(z)} dz, \tag{7}$$

where M_{ij} is the mass of firms exporting from country i to j, z^*_{ij} is the profitability entry threshold, $G_{ij}(z)$ is the cumulative distribution function and $g_{ij}(z)$ is the probability density function governing firm profitability. The aggregate trade elasticity, defined as the percent-change in trade flows between i and j as a result of a percent-change in variable trade costs τ_{ij}, can be expressed as:

$$\frac{\partial \log X_{ij}}{\partial \log \tau_{ij}} = (1 - \epsilon)(1 + \gamma_{ij}), \tag{8}$$

where the contribution of the extensive margin to the elasticity is governed by the parameter γ_{ij} defined as:

$$\gamma_{ij} \equiv \frac{g_{ij}(z^*_{ij})}{(1 - G_{ij}(z^*_{ij}))} \cdot \frac{e^{z^*_{ij}}}{E_{ij}(e^z | z > z^*_{ij})}, \tag{9}$$

where $E_{ij}(\cdot | z > z^*_{ij})$ is a conditional expectation over profitability.

Implication 1 *The elasticity of substitution affects only the magnitude of the overall trade elasticity, but not the relative contribution of the intensive versus extensive margins to the elasticity.*

Notice from equations (8) and (9) that the elasticity of substitution ϵ only affects the magnitude of the overall trade elasticity, but not the relative contribution of the intensive versus extensive margins to the elasticity. A conventional way to write equation (8) is

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

where the first term $(1 - \epsilon)$ is the contribution of the intensive margin to the aggregate elasticity, and the second term $(1 - \epsilon)\gamma_{ij}$ is the contribution of the extensive margin.[12] However, the ratio of the intensive to extensive margin is independent from ϵ and is given by γ_{ij}. Therefore, γ_{ij} is a sufficient statistic for the importance of the extensive margin.

[12] This decomposition was first suggested by Chaney (2008) in the context of a Pareto distribution, Melitz and Redding (2015) provide a generalization consistent with equation (8).

To situate the intensive and extensive margins in the context of the trade literature, consider applying Assumption 1 to equation (8) so that the source of firm-level heterogeneity is a Pareto distributed productivity shock. In this case, equation (8) simplifies to

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

where λ is estimated from the log-sales distribution. The empirical distribution frequently exhibits a fat-tail, implying an estimate of λ close to one.[13] As a result, the contribution of the extensive margin ($=\lambda-1$) relative to the intensive margin ($=1$) is approximately zero. This is a well known result in the trade literature.

Implication 2 *The contribution of the extensive margin to the trade elasticity depends only on the properties of the log-sales distribution.*

In equation (9), the extensive margin elasticity γ_{ij} is determined by two two objects: (i) the entire shape of the log-sales distribution, which is given by the probability density and the cumulative distribution functions, $g_{ij}(\cdot)$ and $G_{ij}(\cdot)$ respectively, and (ii) the entry profitability threshold, z_{ij}^*.

In order to empirically discipline the extensive margin elasticity we will use empirical log-sales distributions. First, parameters of the distribution $G_{ij}(\cdot)$ can be recovered from the micro-data on the log-sales distribution by applying the estimation procedure in Section 4.1. Given estimated parameters we can recover the threshold z_{ij}^*. As noted in Bas, Mayer, and Thoenig (2015), z_{ij}^* is determined by the average-to-minimum ratio of a sales distribution. Using equation (4), we can express the theoretical average-to-minimum ratio as a function of z_{ij}^* alone:

$$\frac{E_{ij}(r_{ij}(z_{ij})|z_{ij}>z_{ij}^*)}{r_{ij}(z_{ij}^*)} = \frac{E_{ij}(e^{z_{ij}}|z_{ij}>z_{ij}^*)}{e^{z_{ij}^*}}. \tag{10}$$

Therefore, we can choose z_{ij}^* so that expression (10) equals the empirical average-to-minimum ratio. Importantly, computing z_{ij}^* does not require any knowledge of the demand side elasticity of substitution across varieties, ϵ.

As a result, in order to obtain an accurate estimate of the extensive margin elasticity, it is crucial that the theoretical distribution matches the data as accurately as possible. As we argued in Section 4, the Exponentially Modified Gaussian distribution provides a close fit to the log-sales data and its fit is superior to other distributions commonly used in the trade literature.

[13]For the estimates of the tail parameter see di Giovanni and Levchenko (2013).

In the next section we apply the described methodology based on micro-data for estimating the extensive margin contribution to trade elasticity to evaluate its magnitude, and compute the bias implied by assuming a Normal distribution.

5.3 Estimates of Elasticities

Using equation (9) we compute the contribution of the extensive margin to trade elasticity for each destination-year. We use destination-year specific estimates of the parameters of distribution γ_{ij} for two distributional assumptions (EMG versus Normal) as estimated in Section 4.

Result 1 *Overall, the extensive margin elasticity implied by EMG distributed log-sales is larger than the elasticity implied by Normally distributed log-sales.*

Panel A of Figure 10 plots estimates of the extensive margin elasticity γ_{ij}. Each point corresponds to an elasticity estimate for a given destination-year. The x-coordinate corresponds to the elasticity implied by a Normal fit to the log-sales distribution, while the y-coordinate corresponds to the elasticity implied by an EMG fit. The red line is the 45-degree line.

If asymmetry in log-sales distributions played no role in magnitude of the extensive margin elasticity, then the EMG extensive margin elasticity estimates would reduce to a symmetric Normal distribution estimates and all points in Panel A would lie on the 45-degree line. However, as is clear from Panel A of Figure 10, that is not the case. The deviations from the 45-degree line are the most pronounced for the destination-year observations exhibiting fatter right-tails compared to those implied by a Normal distribution. Those are the distributions with an estimate of the EMG parameter λ being below 8, and are denoted by stars in Panel A of Figure 10. Hence, as can be seen from Panel A of Figure 10 destination-year observations that exhibit a fatter right-tail have much larger extensive margin elasticities than would be implied by a Normal distribution.

Furthermore, as can be seen from Panel A of Figure 10, the magnitude of the extensive margin elasticity varies between 10^{-12} to 10^{-2} with majority of the values being concentrated around 10^{-5}.[14]

Result 2 *When log-sales distributions are fat-tailed, the Normal distribution generates extensive margin elasticities that under-predict magnitudes by a factor of 10 to 1,000.*

Panel B of Figure 10 demonstrates the extent of the bias in the extensive margin elasticities generated by the Normal distribution. For a subset of observations in which $\lambda < 2$, Panel

[14] The sample average extensive margin elasticity implied by the EMG distribution is 1.6×10^{-5}; by Normal 8.2×10^{-6}.

B of Figure 10 plots the ratio of the EMG relative to Normal extensive margin elasticity estimates against the estimate of λ. As is clear from the Figure, the smaller is the estimated value of λ, the more fat-tailed is the log-sales distribution and the larger is the bias implied by the Normal distribution. For example, if the estimated λ is 1.05 or smaller, then the EMG implied extensive margin elasticity is 10 to 1,000 larger than the elasticity generated by a Normal distribution for the same destination-year observation. Furthermore, the bias is pervasive, as approximately 65% of all observations have an estimated λ less than 2.

Result 3 *Even a small data truncation of all export sales lower than $1,000 generates an upward bias in the estimates of the extensive margin elasticity of the order of magnitude* 10^5 *on average.*

The value of the extensive margin contribution to the trade elasticity γ_{ij} is highly sensitive to the value of the the average-to-minimum ratio. To demonstrate the nature of the sensitivity, Panel A of Figure 11 plots the value of γ_{ij} for parameters of an EMG distribution fitted to match Brazilian exports to Germany in 2001 against counterfactual values of the average-to-minimum ratio. Both the counterfactual values of the average-to-minimum ratio and the counterfactual values of the extensive margin elasticity are normalized by their respective true values in the data. We hold the fitted distributional parameters fixed but re-compute the value of z^*_{ij} using the average-to-minimum ratios between 50% below the true value and 50% above the true value. This allows us to re-compute the extensive margin elasticity γ_{ij} with truncated data.

As can be seen from Panel A of Figure 11, underestimation of the average-to-minimum ratio by 50%, for example, leads to overestimation of the extensive margin elasticity by a factor of 4.5. Similarly, the lower is the average-to-minimum ratio, the greater is the extent of the bias.

Because many customs-level data sets are truncated, the example in Panel A of Figure 11 demonstrates an important issue in computing the extensive margin elasticity. For example, Bas et al. (2015) use French trade data, where 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. Omitting the smallest firms from the sample leads to underestimated average-to-minimum export sales ratios, and therefore leads to an overestimation of the contribution of the extensive margin to trade.

We infer the magnitude of the bias created by the exogenous data truncation by conducting the following counterfactual experiment. We take the original data and drop all firm-destination-year observation with a value of exports below $1,000. We then re-fit an EMG distribution to these truncated data, recompute the average-to-minimum ratio based on the truncated sample, and finally recompute the extensive margin elasticities. In Panel B of Figure 11 we plot the counterfactual elasticities relative the true ones implied by the non-truncated data (as computed in Section 4). Each point in the scatter plot represents a destination-year observation. The counterfactual value of the average-to-minimum ratio is measured on x-axis and the counterfactual value of the extensive margin elasticity is measured on y-axis. Both values are normalized by their respective true value for the corresponding destination-year observation.

As can be seen from Panel B of Figure 11, a relatively small truncation value of $1,000 results in an average-to-minimum ratio that is 21% as large as and as small as 0.2% of its non-truncated value. Truncation results in an overestimation of the extensive margin elasticity by a factor ranging from 0.8 to 10^7 with an average value of 10^5. For further comparison, Panel C of Figure 11 plots a histogram of the extensive margin elasticities for the full versus the truncated sample. Panel C shows that when the data is truncated, the entire distribution of the extensive margin elasticities is shifted to the right by a few orders of magnitude.

6 Conclusion

In this paper, we introduced a novel distribution to the growing literature on the interaction between firm-level heterogeneity and the gains from trade. The Exponentially Modified Gaussian (EMG) distribution parsimoniously captures the salient features of log-sales distributions of exports, particularly a fat right tail and a Normal-like left tail. Furthermore, we document high degrees of variation in sales dispersion and asymmetry across destination markets. The EMG, once fit to log-sales distributions of exports across destination markets, not only characterizes the data well but also fits the data better than either the Normal or Exponential distributions that are used in most international trade research. Using these fitted distributions, we compute aggregate trade elasticities and show that the Normal distribution overstates the magnitude of the extensive margin elasticity. Furthermore we show that when export sales data are truncated, as is true for other work that find high aggregate trade elasticities due to a large contribution of the extensive margin, aggregate trade elasticities become larger than their true values computed from untruncated data.

References

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. 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.

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

Champernowne, D. G. (1953): "A model of income distribution," *Economic Journal*, 63, 318–351.

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 J. R. B. D'Agostino (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.

Dix-Carneiro, R. and B. K. Kovak (2015): "Trade reform and regional dynamics: evidence from 25 years of Brazilian matched employer-employee data," Unpublished.

Eaton, J., S. Kortum, and F. Kramarz (2011): "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 79, 1453–1498.

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.

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

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.

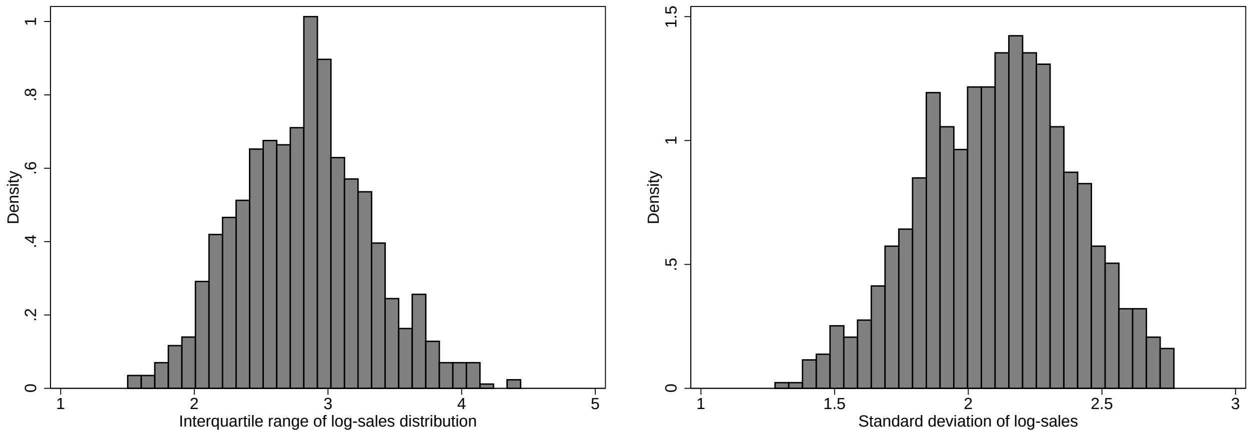
Molinaz, D. and M.-A. Muendler (2013): "Preparing to Export," NBER Working Paper No. 18962.

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

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

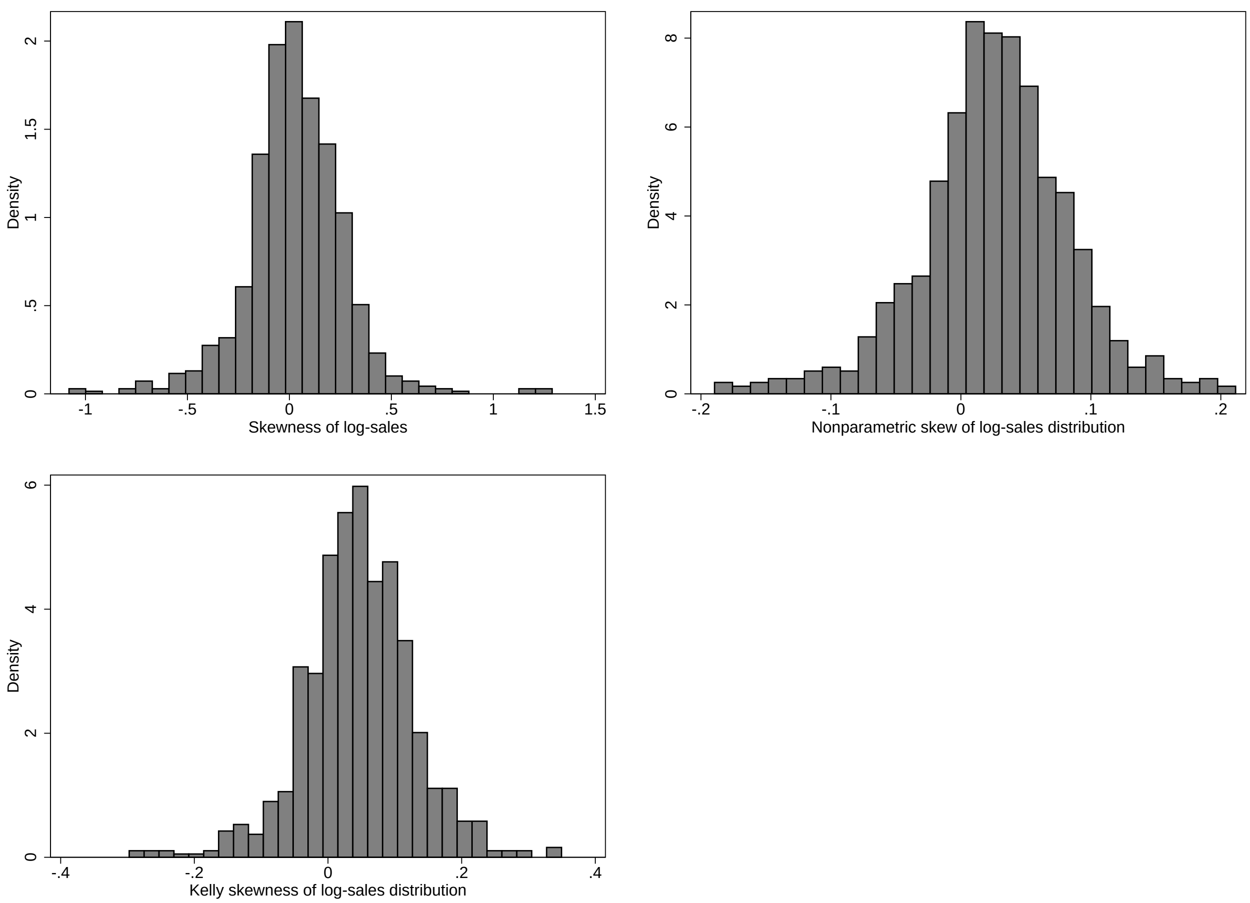
Toda, A. A. and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123, 1177 – 1200.

Figure 1: Heterogeneity in the dispersion of log-sales across export destinations.



Notes: The figure depicts the distribution of the values of the interquartile range and the standard deviation of log-sales across 847 destination-year observations where at least 100 firms export.

Figure 2: Heterogeneity in the skewness of log-sales across export destinations.



Notes: The figure depicts the distribution of the values of the various measures of skewness of log-sales across 847 destination-year observations where at least 100 firms export.

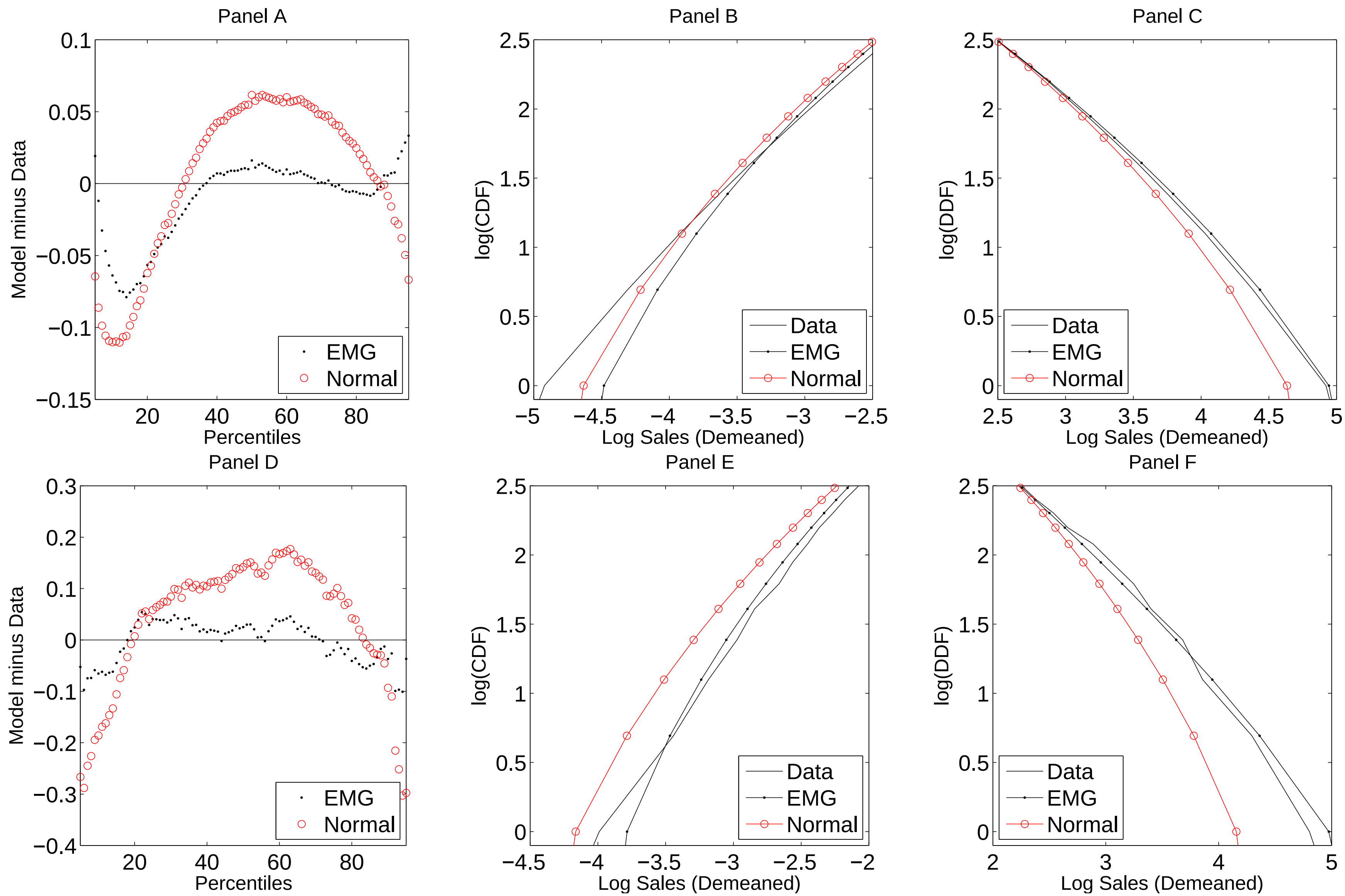


Figure 3: Model errors. The top row shows model errors averaged over each destination-year pair. The bottom row shows model errors for a particular destination-year.

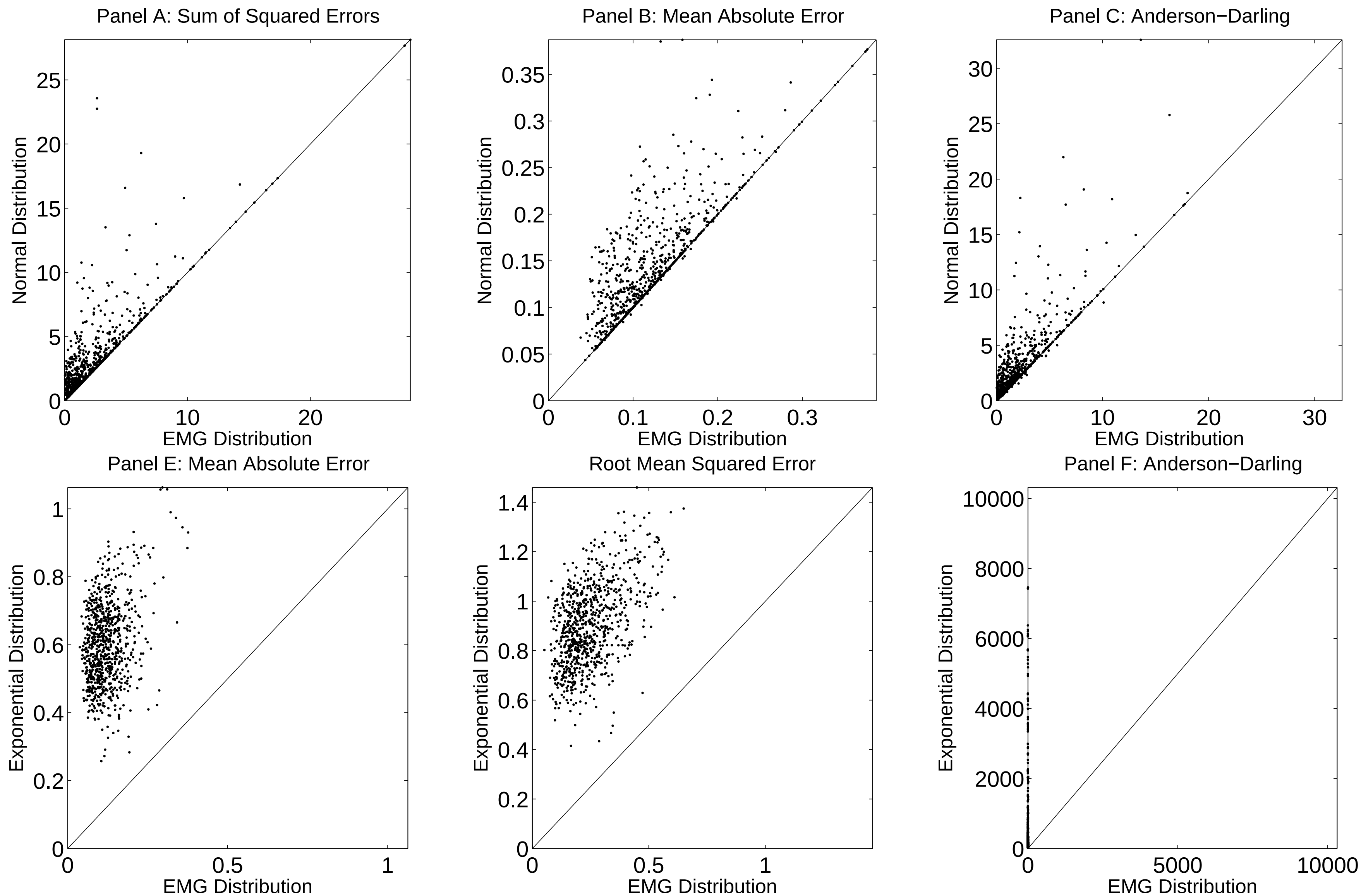


Figure 4: Goodness of fit statistics across each destination-year pair. Scatter plots show the goodness of fit statistic for the Exponentially Modified distribution (x-axis) versus an alternative distribution (y-axis). The Normal distribution is the alternative in the top row, while the Exponential distribution is the alternative in the bottom row.

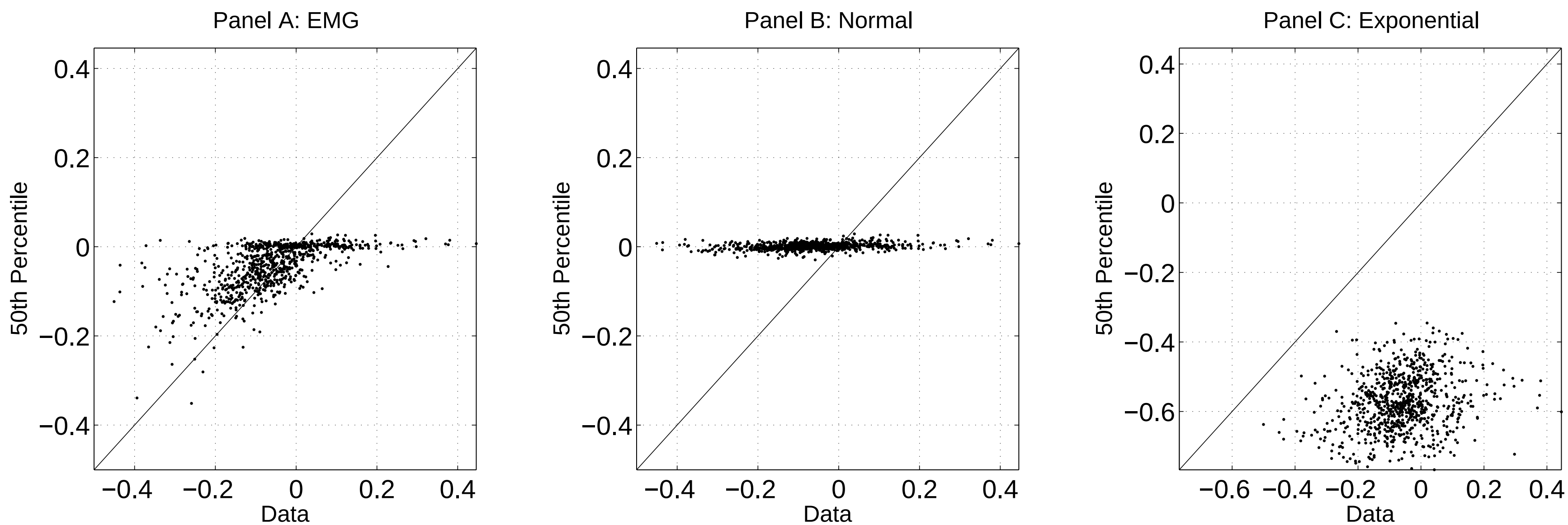


Figure 5: Scatter plots show model generated medians (y-axis) against the empirically observed medians (x-axis) across destination-year pairs. Red points indicate that estimated parameters for the Exponentially Modified Gaussian distribution recover a Normal distribution.

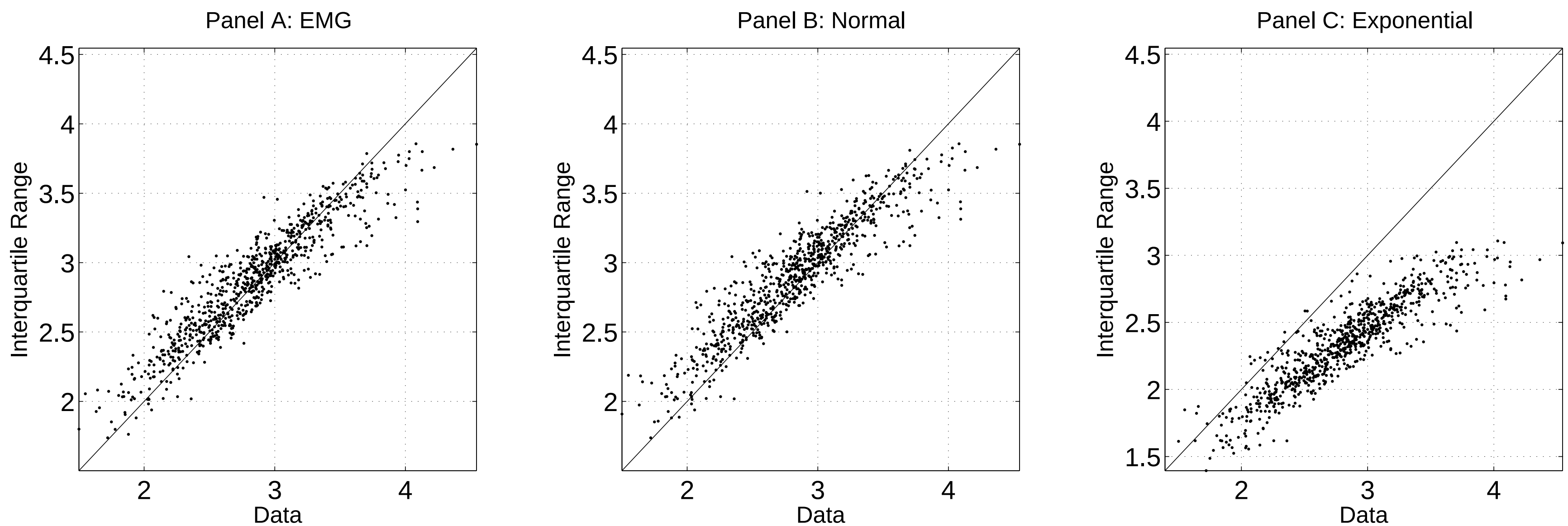


Figure 6: Scatter plots show model generated interquartile ranges (y-axis) against the empirically observed interquartile ranges (x-axis) across destination-year pairs. Red points indicate that estimated parameters for the Exponentially Modified Gaussian distribution recover a Normal distribution.

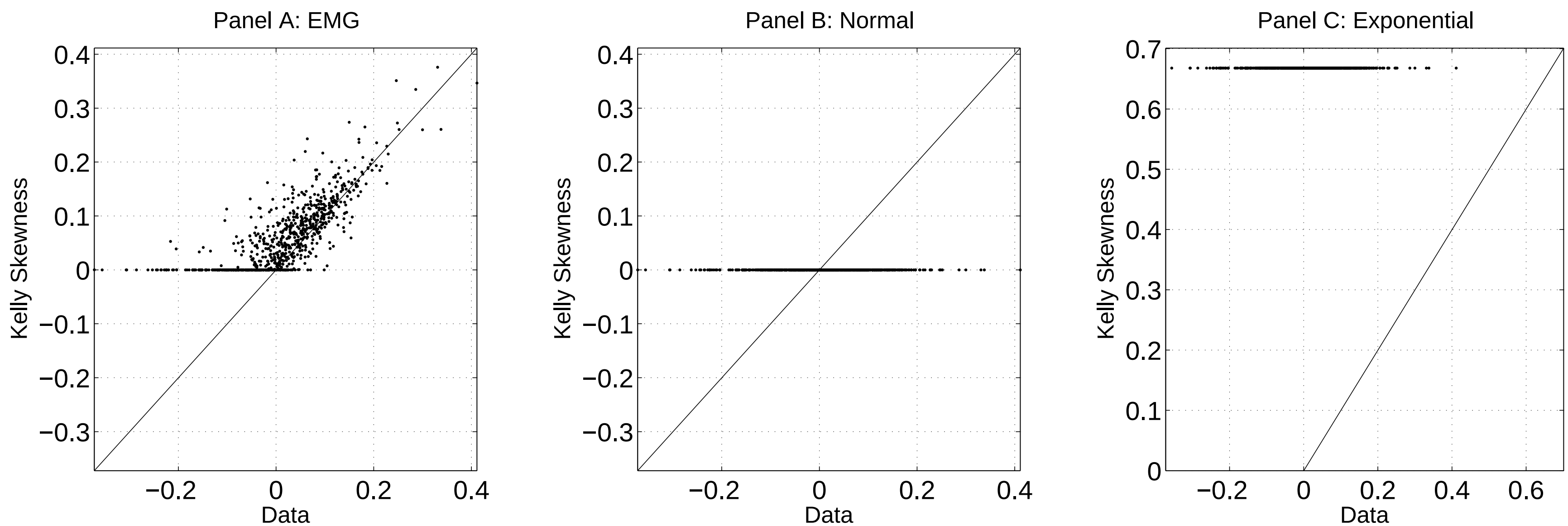


Figure 7: Scatter plots show model generated Kelly skewness (y-axis) against the empirically observed Kelly skewness (x-axis) across destination-year pairs. Red points indicate that estimated parameters for the Exponentially Modified Gaussian distribution recover a Normal distribution.

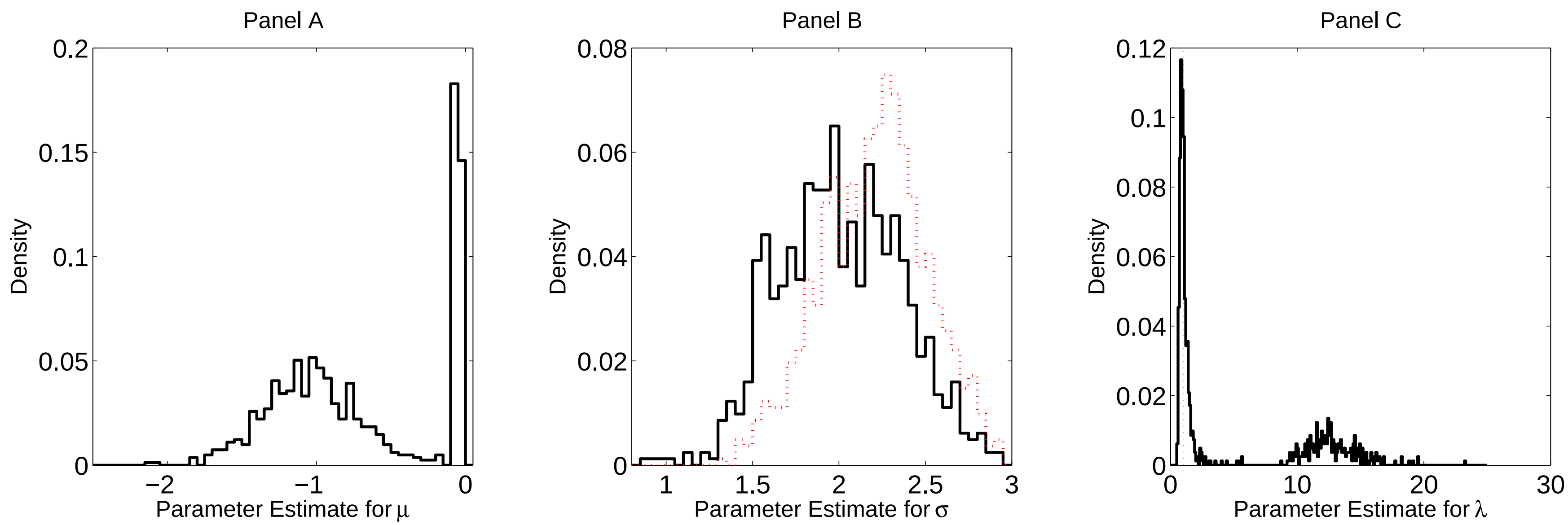


Figure 8: Histograms for parameter estimates across destination-year pairs in the data. Solid black lines indicate the histogram for the Exponentially Modified Distribution. Dotted red lines indicate the histogram for the Normal distribution.

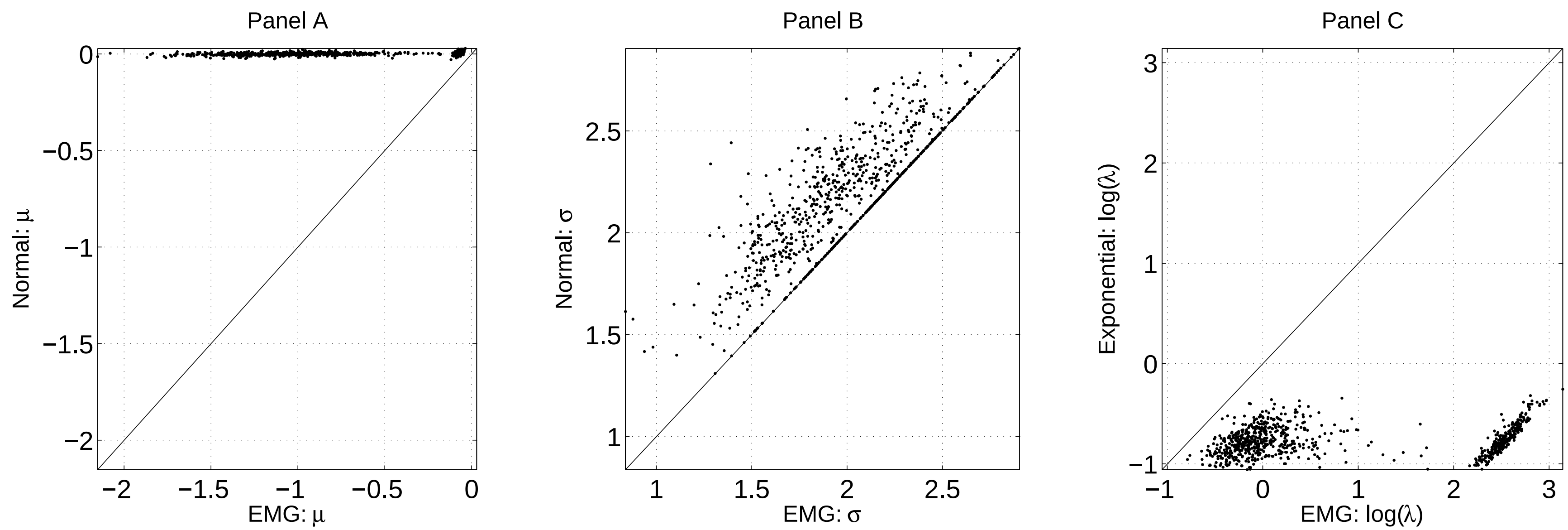
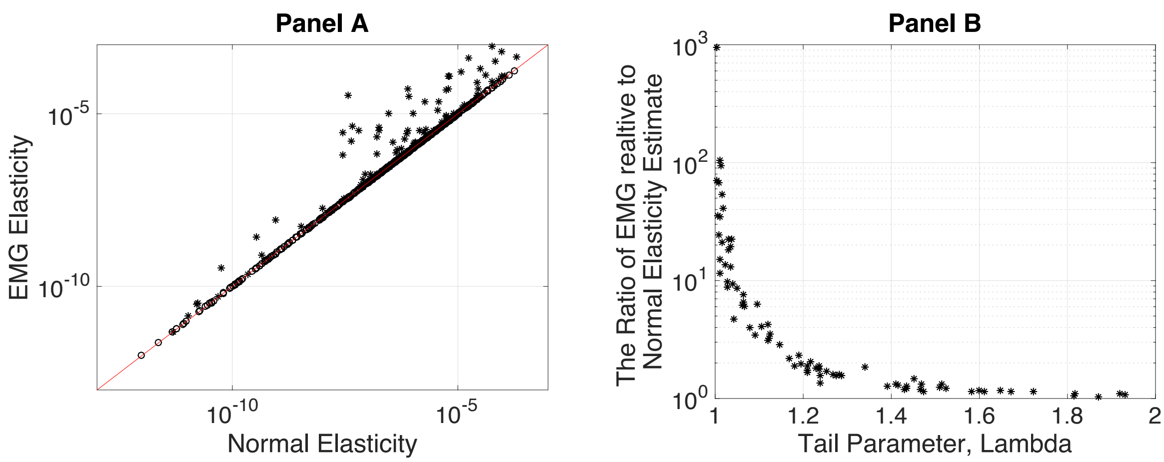


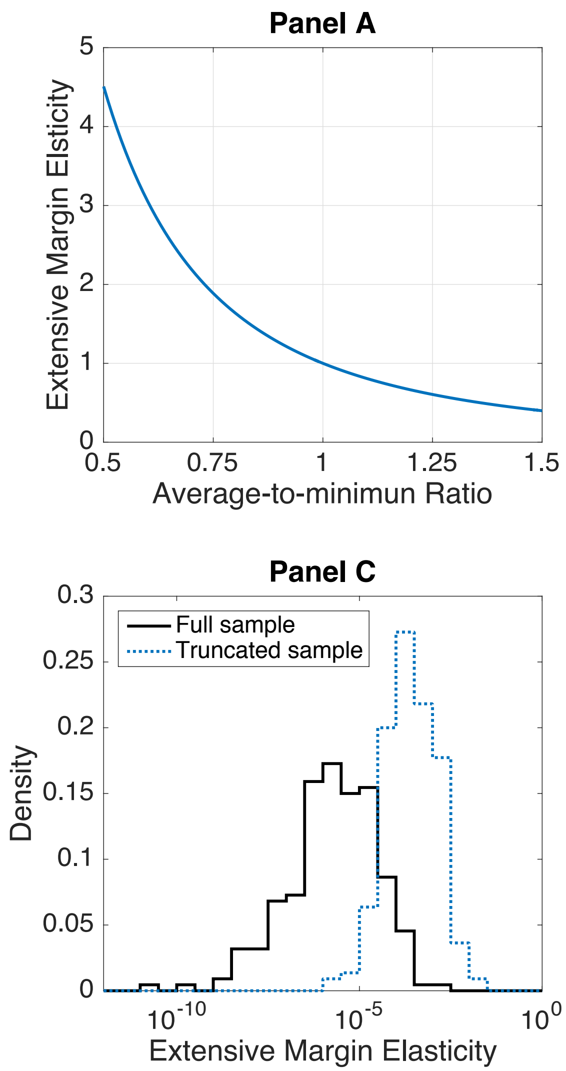
Figure 9: Scatter plots compare estimated model parameters. Red points indicate that estimated parameters for the Exponentially Modified Gaussian distribution recover a Normal distribution.

Figure 10: Extensive Margin Elasticity Estimates



Notes: All panels in the figure depict the estimates of the extensive margin elasticity for destination-year observations with an EMG estimate of the tail parameter $\lambda > 1$. The elasticity is not defined for $\lambda \leq 1$. In Panel A stars denote destination-year observations with the estimated parameter $1 < \lambda < 8$; circles denote destination-year observations with the estimated parameter $\lambda > 8$.

Figure 11: Extensive Margin Elasticity: Sample Truncation Bias



Notes: Panel A plots the values of the extensive margin elasticity computed using a fitted EMG distribution to Brazilian exports to Germany in 2001 against a counterfactual value of the average-to-minimum ratio. The conterfactual value of the average-to-minimum ratio measured on x-axis is normalized by the true value. The counterfactual value of the extensive margin elasticity measured on y-axis is normalized by the true value. Panel B plots the values of the extensive margin elasticity computed using a fitted EMG distribution. Each dot is a destination-year observation. The counterfactual value of the average-to-minimum ratio measured on x-axis is normalized by the true value for the corresponding observation. The counterfactual value of the extensive margin elasticity measured on y-axis is normalized by the true value for the corresponding observation.

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

Statistic	Mean	Standard Deviation	Min	Max
Panel A: Moment based characterization of a distribution				
The standard deviation of log-sales	2.11	0.28	1.28	2.77
The skewness of log-sales	0.03	0.24	-1.08	1.29
The nonparametric skew	0.03	0.06	-0.19	0.21
Panel B: percentile based characterization of a distribution				
The interquartile range	2.82	0.49	1.50	4.44
The Kelly skewness	0.04	0.08	-0.30	0.35

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

Table 2: OLS regressions of log-sales dispersion on a time trend

Dep. variable: dispersion	st. dev. (1)	IQR (2)
Time trend	0.361***	-0.358*
	(0.097)	(0.210)
R^2	0.87	0.80
No. obs.	845	845
Dest. FE	yes	yes

Note: the dependent variable across columns is the inidcated measure of dispersion of the log-sales disitribution across destination-year observations.
***, * Statistically significant at 1%, 10% level.

Table 3: OLS regressions of log-sales skewness on a time trend

Dep. variable: skewness	sample	nonparametric	kelly
	(1)	(2)	(3)
Time trend	0.845***	0.245***	0.332***
	(0.161)	(0.044)	(0.062)
R^2	0.53	0.40	0.39
No. obs.	845	845	845
Dest. FE	yes	yes	yes

Note: the dependent variable across columns is the inidcated measure of skewness of the log-sales disitribution across destination-year observations.

*** Statistically significant at 1% level.

A Proofs of Propositions

Proof of Proposition 1

Let x and y be random variables such that $x \sim \mathcal{N}(\mu, \sigma^2)$, $y \sim Exp(\lambda)$ and parameters satisfy $\mu \in \mathbb{R}$, $\sigma \geq 0$ and $\lambda \geq 0$. 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:

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

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:

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

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:

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

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

Proof of Proposition 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. ■

Proof of Proposition 3

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

$$\begin{aligned}\int_{z^*}^{+\infty} e^z dG(z) &= \int\int_{x+y>z^*} e^{x+y}\lambda e^{-\lambda y} f(x)dxdy \\ &= \int_{z^*}^{+\infty} e^x f(x)dx \int_0^{+\infty} e^y \lambda e^{-\lambda y}dy + \int_{-\infty}^{z^*} e^x \left(\int_{z^*-x}^{+\infty} e^y\lambda e^{-\lambda y}dy\right) f(x)dx\end{aligned}$$

Take the first integral in the expression. This can be simplified to:

$$\int_{z^*}^{+\infty} e^x f(x)dx \int_0^{+\infty} e^y\lambda e^{-\lambda y}dy = \frac{\lambda}{\lambda-1} \cdot e^{\mu+\frac{1}{2}\sigma^2}\left(1-\Phi\left(\frac{z^*-\mu-\sigma^2}{\sigma}\right)\right)$$

The second integral in the expression requires more work:

$$
\begin{aligned}
\int_{-\infty}^{z^*} e^x \left(\int_{z^*-x}^{+\infty} e^y \lambda e^{-\lambda y} dy \right) f(x)dx &= \frac{\lambda}{\lambda-1} e^{-(\lambda-1)z^*} \int_{-\infty}^{z^*} e^{\lambda x} f(x)dx \\
&= \frac{\lambda}{\lambda-1} e^{-(\lambda-1)z^*} \int_{-\infty}^{\lambda z^*} e^{\lambda x} f(\lambda x)d(\lambda x) \\
&= \frac{\lambda}{\lambda-1} e^{-(\lambda-1)z^*} e^{\lambda\mu+\frac{1}{2}\lambda^2\sigma^2} \Phi\left(\frac{z^*-\mu-\lambda\sigma^2}{\sigma}\right)
\end{aligned}
$$

Therefore, by summing the two integrals we obtain the conditional expectation:

$$
\int_{z^*}^{+\infty} e^z dG(z) = M_z(1) \left[1 - \Phi\left(\frac{z^*-\mu-\sigma^2}{\sigma}\right) + e^{-(\lambda-1)(z^*-\mu-\lambda\sigma^2)-\frac{1}{2}(\lambda-1)^2\sigma^2} \Phi\left(\frac{z^*-\mu-\lambda\sigma^2}{\sigma}\right) \right]
$$

as desired. ■

B The model

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

$$
U_j = \left(\sum_{i=1}^{N} \int_{\omega \in \Omega_{ij}} \left(e^{\theta_{ij}(\omega)}\right)^{\frac{1}{\epsilon}} c_{ij}(\omega)^{\frac{\epsilon-1}{\epsilon}} d\omega \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{11}
$$

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 $\omega \in \Omega_{ij}$.[15]

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

$$
c_{ij}(\omega) = e^{\theta_{ij}(\omega)} p_{ij}(\omega)^{-\epsilon} Y_j P_j^{\epsilon-1}, \tag{12}
$$

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

[15] 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.

price index in country j. The aggregate price index is given by

$$P_j^{1-\epsilon} = \sum_{i=1}^{N} \int_{\omega \in \Omega_{ij}} e^{\theta_{ij}(\omega)} p_{ij}(\omega)^{1-\epsilon} d\omega. \tag{13}$$

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 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 (12) 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

$$\begin{aligned} r_{ij}(\theta_{ij}, \varphi) &= \left(\frac{\epsilon - 1}{\epsilon}\right)^{\epsilon - 1} (\tau_{ij} w_i)^{1-\epsilon} Y_j P_j^{\epsilon - 1} e^{\theta_{ij}} \varphi^{\epsilon - 1}, \qquad (14) \\ \pi_{ij}(\theta_{ij}, \varphi) &= \frac{r_{ij}(\theta_{ij}, \varphi)}{\epsilon} - w_i f_{ij}. \qquad (15) \end{aligned}$$

Notice from equations (14) and (15) 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

$$z_{ij} = \theta_{ij} + \log\left(\varphi^{\epsilon - 1}\right). \tag{16}$$

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

$$r_{ij}(z_{ij}) = \left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}(\tau_{ij}w_i)^{1-\epsilon}Y_jP_j^{\epsilon-1}e^{z_{ij}}. \tag{17}$$

$$\pi_{ij}(z_{ij}) = \frac{r_{ij}(z_{ij})}{\epsilon} - w_i f_{ij}. \tag{18}$$

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

$$e^{z^*_{ij}} = \frac{\epsilon w_i f_{ij}(w_i\tau_{ij})^{\epsilon-1}}{\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}Y_jP_j^{\epsilon-1}}. \tag{19}$$

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

$$r_{ij}(z_{ij}) = \epsilon w_i f_{ij}\frac{e^{z_{ij}}}{e^{z^*_{ij}}}. \tag{20}$$

B.3 Trade Elasticity

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

$$X_{ij} = M_{ij}\int_{z^*_{ij}}^{+\infty} r_{ij}(z)\frac{g_{ij}(z)}{1-G_{ij}(z)}dz, \tag{21}$$

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

$$M_{ij} = J_i(1-G_{ij}(z^*_{ij})). \tag{22}$$

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

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

Proof: Substitute equations (20) and (22) into equation (21) 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 (19):

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

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:

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

■