A Tale of Two Tails: Productivity Distribution and the Gains from Trade

Sergey Nigai[*]

ETH Zurich, KOF and CESifo

This draft: March 2016

Abstract

I use firm-level data to show that neither the Log-normal nor the Pareto distribution can approximate the shape of the productivity distribution along the entire support. While the former underpredicts the thickness of the right tail, the latter does not capture the shape of the left one. Using empirical distribution as a benchmark, I show that such inaccuracies lead to sizable errors in the estimates of the gains from trade in models featuring firm selection. I propose using a mixed distribution which models the left tail as Log-normal and right tail as Pareto and produces negligible errors in quantitative analysis.

Keywords: Productivity distribution, Welfare gains, International Trade, Pareto tail

JEL-codes: F1; F10; F12.

1 Introduction

In trade models where firms matter, selection effects largely determine how falling trade barriers affect the number of goods and prices available to consumers.[1] Quantifying these effects depends crucially on the shape of the productivity distribution, which is most often assumed to follow either a Pareto or Log-normal distribution. I explore how well these assumptions match available micro-level data and, more importantly, what errors they bring to the estimates of the gains from trade.

It has been well documented in the empirical and theoretical literature that firm-specific characteristics such as size and productivity often follow a Pareto distribution at least in the upper-right tail (see Axtell, 2001; Gabaix, 2008; Levchenko and di Giovanni, 2012; Arkolakis, 2015).[2]Due to the consistency with the data as well as analytic tractability, Pareto distribution has been the most popular choice for modelling heterogeneity parameters in different variants of Melitz (2003).[3]Recently, however, the plausability of the Pareto assumption has been challenged on the grounds of the available micro data on firms' sales, e.g., Head, Mayer and Thoenig (2014) and Freund and Pierola (2015), specifically by emphasizing that Log-normal provides a closer fit to the data when the entire distribution of sales is considered.[4]This debate is not unique to international trade and arises whenever the choice between Log-normal and Pareto is unclear, e.g., Eeckhout (2004) argues that Log-normal dominates Pareto in matching the city size distribution when the entire distribution (not just the upper-tail) is considered.

I combine these seemingly conflicting arguments by suggesting that while Log-normal distribution provides a closer fit to the data on measures of efficiency for a vast part of the support, the upper-right tail is better approximated by Pareto which calls for a mixed distribution. This is easy to see in Figure 1, where I plot the empirical probability density function of a productivity measure consistent with Melitz (2003) for almost one million French firms in 2012 along with the best fitting Log-normal (red dash-dotted line) and Pareto (green dashed line) models when fitted separately on the bottom 95 percent (left panel) and top 5 percent (right panel) of the distribution. Looking at the bottom 95 percent of the observations, one may conclude that Log-normal is a clearly better approximation. However, the results reverse completely for the top 5 percent where Pareto clearly dominates.

---

[1]For example, using data on French firms Eaton, Kortum and Kramarz (2011) show that the selection effects account for more than 50 percent of firm entry in different markets.

[2]Also see Simon and Bonini (1958), Luttmer (2007), Levchenko, di Giovanni, and Ranciere (2011), Levchenko and di Giovanni (2013).

[3]Following Baldwin (2005) and Chaney (2008), hundreds of papers assumed either unbounded or bounded Pareto distribution of productivities. A non-exhaustive list of seminal works in international trade includes Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008), Helpman, Melitz and Rubinstein (2008), Melitz and Ottaviano (2008), Arkolakis, Costinot and Rodríguez-Clare (2012), Melitz and Redding (2014) and many others.

[4]Yang (2014) and Bas, Mayer and Thoenig (2015), and Fernandes, Klenow, Meleshchuk, Pierola, and Rodríguez-Clare (2015) also use Log-normal distribution in the Melitz (2003) framework. For theoretical treatment on how assumptions on technology and demand affect the distribution of sales see Mrazova, Neary and Parenti (2015).
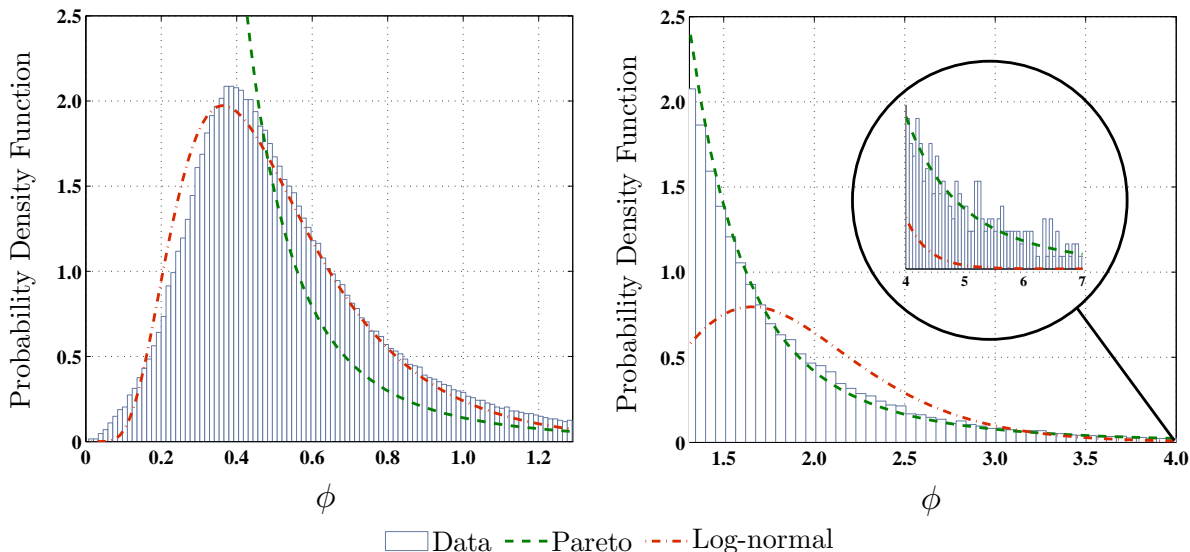
Figure 1: Empirical and Parametric Probability Density Functions

Perhaps, the most striking implication of Figure 1 is that neither Log-normal nor Pareto is able to simultaneously match both tails of the empirical distribution which, as it turns out, is extremely important for the correct calculation of different trade outcomes such as the gains from trade. I use a workhorse general equilibrium model of trade with heterogeneous firms to show that the assumptions of Log-normal and (un-)bounded Pareto can generate significant errors in the estimates of the gains from trade with magnitudes on par with the total gains implied by the empirical benchmark.

I propose using an alternative distribution model that amalgamates the left tail of Log-normal and the right tail of Pareto with an endogenous threshold point. I show that this distribution fits the data well in both tails while still offering the advantages of a well-behaved parametric equation. In the baseline estimation, I use data on domestic sales of almost a million French firms in 2012 and show that the bottom 94 percent of observations follow Log-normal and top 6 percent follow Pareto. I show that the mixed distribution outperforms Log-normal and (un-)bounded Pareto in matching the data and correctly predicting different trade outcomes relative to the empirical distribution.

The results prove to be robust in a number of dimensions. The proposed distribution dominates more general classes of Pareto and Log-normal models that feature higher number of parameters. The results are not sensitive to (i) truncating the sample from the right or from the left, (ii) using data from different countries and (iii) using different measures of productivity. The proposed mixed distribution also squares well with the data in an out-of-sample validity check, which suggests that it may also be of value in other areas of economics when the choice between (un-)bounded Pareto and Log-normal is not obvious. For example, research on the city size distribution (see Gabaix,

1999; Eeckhout, 2004, 2009; Levy, 2009) often involves debates about whether the upper-tail follows Pareto. I employ data from Eeckhout (2009) and show that the mixture distribution outperforms both Pareto and Lognormal, and find estimates suggesting that around the top 2 percent of the cities in the United States follow Pareto.

This paper relates to a broad body of work on heterogeneous firms models and the aggregate gains from trade (Melitz, 2003; Chaney, 2008; Arkolakis, Costinot and Rodríguez-Clare, 2012; Melitz and Redding, 2014; Feenstra, 2014 and many others). I quantify the magnitude of potential errors in the estimates of the gains from trade that arise in such trade models when most common parametric assumptions about the shape of the efficiency distribution are employed. This paper is also in the spirit of the literature that uses micro-level data to estimate the parameters of the heterogeneous trade models (Bernard, Eaton, Jensen and Kortum, 2003; Arkolakis, 2010; Eaton, Kortum and Kramarz, 2011) with a particular focus on employing such data for the parameterization of the productivity distribution. I also relate to Arkolakis (2015) who shows that a two-piece distribution of productivities can arise as a result of firm selection and growth as well as to Mrazova, Neary and Parenti (2015) who discuss how interactions between assumptions on demand and technology shape the distribution of firm-specific outcomes. However, the focus of this work is largely different, i.e., putting forward and testing the performance of a two-piece distribution in terms of matching the data on efficiency measures in workhorse models of international trade featuring firm selection.

The remainder of the paper is organized as follows. The next section presents and discusses properties of a two-piece distribution that mixes the left tail of Log-normal and right tail of Pareto. In Section 3, I estimate the proposed distribution together with the three alternative models most frequently encountered in the literature and compare their performance across different dimensions. I sketch a model of trade with heterogeneous firms in Section 4 and compare the predictions of the welfare gains from trade in counterfactual experiments implied by different parametric distributions to the numerical benchmark. Section 5 provides sensitivity analysis, discusses possible extensions and shows how the proposed distribution and estimation approach can be applied to out-of-sample data. The last section offers a brief conclusion.

2 Two-piece Distribution: Log-normal meets Pareto

A two-piece probability distribution combines standard Log-normal and Pareto distributions with the following probability density functions:[5]

$$f_L(\phi) = \frac{1}{\sqrt{2\pi}s\phi}e^{-\frac{1}{2}\left(\frac{\ln\phi-\mu}{s}\right)^2} \ \text{ and } \ f_P(\phi) = \frac{\alpha\theta^\alpha}{\phi^{\alpha+1}}. \tag{1}$$

I mix the distributions such that the left tail up to a threshold value θ is distributed according

---

[5]Two-piece probability distributions that mix Log-normal and Pareto were originally developed in Cooray and Ananda (2005) and Scollnik (2005). Here, I build on and extend a version originally derived in Scollnik (2005).

4

to Log-normal, whereas the right tail beyond θ is distributed Pareto. Under the assumptions of continuity and differentiability of the resulting cumulative distribution function (c.d.f.) and probability density function (p.d.f.), I derive a mixture dubbed *Two-piece* with shape parameter, α, and two scale parameters, θ and ρ, with values determined by the original parameters in f_L and f_P:

$$f(\phi) = \begin{cases} \dfrac{\rho}{\Phi\big(\alpha s(\alpha, \rho)\big)} \dfrac{1}{\sqrt{2\pi} s(\alpha, \rho)\phi} e^{-\frac{1}{2}\left(\alpha s(\alpha,\rho) - \frac{\ln\theta - \ln\phi}{s(\alpha,\rho)}\right)^2} & \text{for } \phi \in (0, \theta] \\[4mm] (1-\rho)\dfrac{\alpha\theta^\alpha}{\phi^{\alpha+1}} & \text{for } \phi \in [\theta, \infty), \end{cases} \tag{2}$$

and:

$$F(\phi) = \begin{cases} \dfrac{\rho}{\Phi\big(\alpha s(\alpha, \rho)\big)} \Phi\left(\alpha s(\alpha, \rho) + \dfrac{\ln\phi - \ln\theta}{s(\alpha, \rho)}\right) & \text{for } \phi \in (0, \theta] \\[4mm] 1 - (1-\rho)\dfrac{\theta^\alpha}{\phi^\alpha} & \text{for } \phi \in [\theta, \infty), \end{cases} \tag{3}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal and $s(\rho, \alpha)$ is an implicit function which defines s given ρ and α according to:

$$\Phi\big(\alpha s(\alpha, \rho)\big)\sqrt{2\pi}\big(\alpha s(\alpha, \rho)\big)e^{\frac{1}{2}[\alpha s(\alpha,\rho)]^2} = \frac{\rho}{1-\rho}. \tag{4}$$

I provide detailed derivations of the Two-piece c.d.f. and p.d.f. in the Appendix. At this point, $f(x)$ and $F(x)$ are well-behaved functions that satisfy necessary properties and feature two scale parameters, θ and ρ, and one shape parameter, α. The first scale parameter indicates the threshold value of the random variable which splits the distribution into two tails. The second scale parameter, ρ, has a straightforward interpretation and indicates the share of random variables that are distributed according to Log-normal. For example, if $\rho = 0.95$ then the bottom 95 percent of the observations are distributed according to Log-normal and the top 5 percent – according to Pareto.

For illustrative purposes, I plot the c.d.f. and p.d.f. of a parameterized version of the Two-piece distribution (blue solid line), where I arbitrarily set $\theta = 1$, $\rho = 0.95$ and $\alpha = 3$ in the left and right panels of Figure 2, respectively. Here, governed by the assumed values of the parameters, the bottom 95 percent of observations are distributed according to Log-normal up to a threshold value of unity, and the top 5 percent follow Pareto. For comparison, I also plot c.d.f.s and p.d.f.'s of Log-normal (red dash-dotted line) and Pareto (green dashed line) that match the first two moments of the Two-piece such that the mean and variance of ϕ are identical in all three distributions.

The figure suggests several interesting differences between Two-piece, Log-normal and Pareto distributions. On the one hand, relative to the Log-normal distribution, the Two-piece distribution converges to unity at a slower rate (left panel), which translates into a thicker right tail (right panel). Intuitively, this would mean that when ϕ is interpreted as a measure of productivity, the
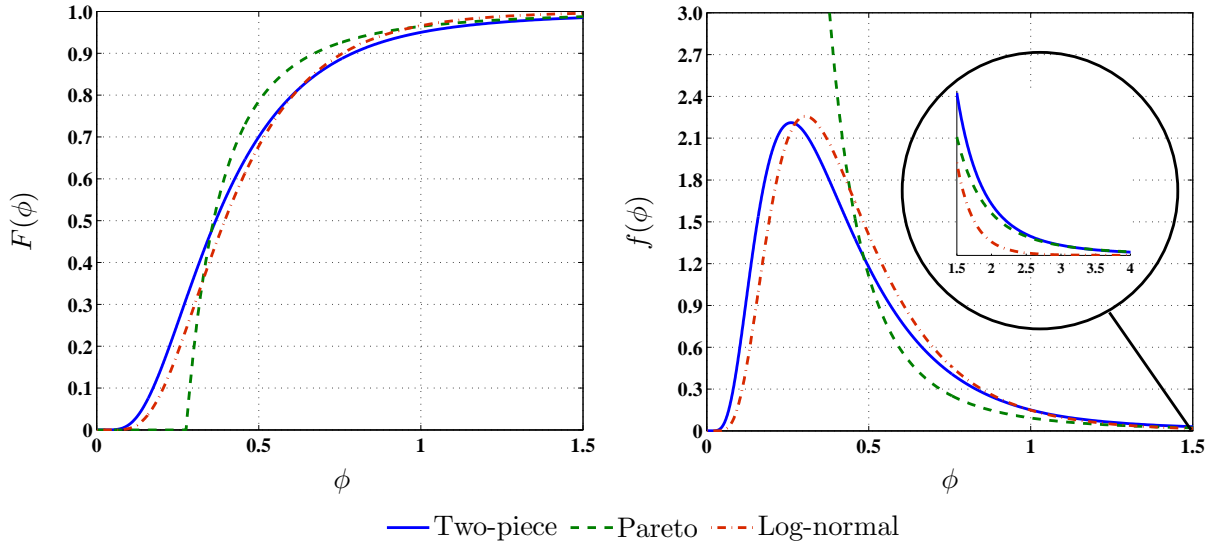
5

Figure 2: Two-piece, Log-normal and Pareto distributions with identical first two moments

Two-piece would have a larger mass of firms with relatively high productivities. On the other hand, in comparison to Pareto, Two-piece has a bell-shaped left tail indicating a larger mass of firms with relatively low productivity. Hence, the Two-piece distribution can be viewed as a compromise between the two most popular models of productivity distribution as it is able to capture the bell-shaped left tail while still having a relatively fat right tail. As it turns out, this feature helps fit the data much better in comparison to pure (un-)bounded Pareto and/or Lognormal models.

3 Empirical Application

In this section, I use firm-level data to highlight the empirical relevance of the Two-piece distribution in comparison to more popular alternatives such as Log-normal, unbounded Pareto and bounded Pareto distributions.[6] The analysis here allows comparing the four distributions along several important dimensions: (i) relative size of the residuals across different slices of the data, (ii) distance of the predicted to the observed quantiles and (iii) distance of the predicted to the observed densities.

The dependant variable is calculated from the raw data on domestic sales of French firms in 2012 and consists of 928,569 observations. To translate these observations into a meaningful measure of efficiency along the lines of Melitz (2003), I demean the data and take them to the power of 0.33 which corresponds to the value of the elasticity of the substitution parameter of 4 from Bernard, Eaton, Jensen and Kortum (2003). I provide more information on how this normalization allows

---

[6]Bounded Pareto is employed in Helpman, Melitz, Rubinstein (2007), Feenstra (2014), and Melitz and Redding (2014).

6

me to recover firm-specific productivity measures in the next section. The description and sources of the raw data are in the Appendix.

I employ a QQ-estimator that minimizes the sum of the squared distance between (log) observed quantiles of the data and (log) predicted quantiles by each of the four models considered. The estimator solves the following:[7]

$$\min_{\Theta_\ell} \left\{ \sum_q \left(\ln\left[Q_e(q)\right] - \ln\left[Q_\ell(q|\Theta_\ell)\right] \right)^2 \right\}, \tag{5}$$

where Q_e is the empirical quantile function and Q_ℓ is its parametric counterpart with ℓ denoting the Two-piece, Log-normal, Pareto, or Bounded-Pareto models. For example, in the case of the Pareto distribution with shape parameter α and scale parameter x_m such that $\Theta_{Pareto} = \{\alpha, x_m\}$, the theoretical quantile function is $Q_{Pareto} = x_m(1-q)^{-1/\alpha}$ which in logs reduces the estimator in equation 3 to a simple linear regression. For computational purposes, I produce 100,000 observations for Q_e using a 100,000 point grid on the empirical c.d.f. of the original data and their corresponding values. Since the grid defines the size of increments on (0,1), increasing it further, though feasible, wouldn't change the results but simply slow the optimization algorithm.[8]Standard errors are bootstrapped using 1,000 draws. I provide full details on the exact functional forms of all Q_ℓ in the Appendix. The estimated parameters along with the standard errors and root mean squared errors (RMSE) across different slices of the data are reported in Table 1.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	3.033	1.185	0.938	0.058	0.465	0.221	0.026	0.033
	(0.006)	(0.005)	(0.001)					
Log-normal	0.569	-0.701		0.069	0.415	0.194	0.156	0.304
	(0.001)	(0.001)						
Pareto	1.914	0.294		0.236	1.405	0.840	0.344	0.648
	(0.005)	(0.001)						
Bounded Pareto	0.372	0.214	1.438	0.183	1.105	0.582	0.406	0.799
	(0.026)	(0.001)	(0.015)					

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in the case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 1: ESTIMATION RESULTS

Table 1 suggests that parameters of the four models are estimated with good precision and that they fit the data relatively well. However, it is important to note that the Two-piece distribution

---

[7]A similar estimator is employed in Head, Mayer and Thoenig (2014). The results are robust to using alternative estimation methods.

[8]I have experimented with increasing the grid to the exact number of observations with no changes to the estimates.

dominates the other three models in terms of fitting the data when the entire support is considered. The estimates suggest that about the top 6 percent of the data follow Pareto and that the threshold value is equal to 1.18. The overall value of RMSE is the lowest among the four models and is equal to 0.058. The Two-piece distribution also fits the data considerably better in the right tail of the distribution. The only instance when it is dominated by one of the alternatives occurs in the bottom 5 percent of observations where Log-normal has a slight edge reflected in marginally lower RMSE on that interval. Unbounded and bounded Pareto perform significantly worse than Two-piece and Log-normal in every dimension.
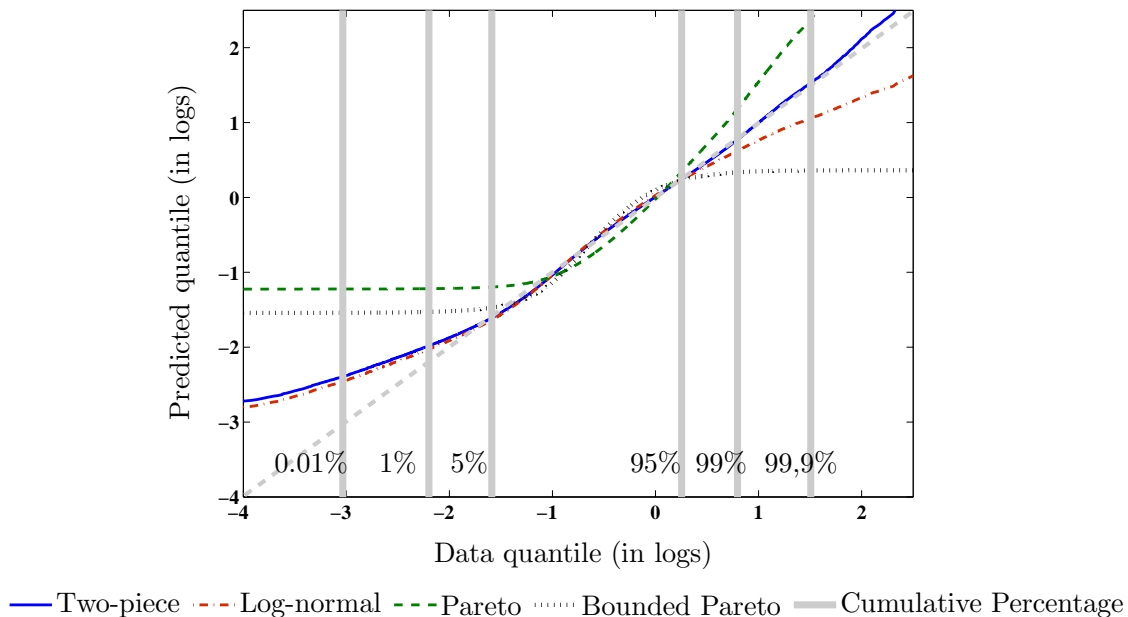

Figure 3: QQ Plot of Two-piece, Log-normal and Pareto vs. Data

The results of Table 1 are confirmed in a QQ-plot in Figure 3 where I plot empirical quantiles against their predicted counterparts for the four parametric models. The Two-piece distribution (blue solid line) follows closely the 45-degree line from the top to the bottom 5 percent where it slowly starts to diverge. However, it performs much better than Log-Normal (red dash-dotted line) in the upper 5 percent of the distribution and only slightly worse in the bottom 5 percent. The Two-piece distribution also dominates bounded (green dashed line) and unbounded (black dotted line) Pareto, which deviates from the data substantially in the left and right tails. Unbounded Pareto seems to outperform its truncated counterpart in the upper-tail but falls short everywhere else.

Finally, I compare the predictions of the four models in terms of fitting the empirical probability density in Figure 4. In the left panel, I plot the predicted density for the bottom 94 percent of the distribution. Here, both Two-piece and Log-normal distributions closely fit the data whereas unbounded and bounded Pareto deviate substantially. However, only the Two-piece distribution is able to match the top 5 percent of the data as suggested by the right panel of Figure 4 where

I plot the right tail of the empirical probability density function. Log-normal (unbounded Pareto) tends to underpredict (overpredict) observed frequencies in the right tail.
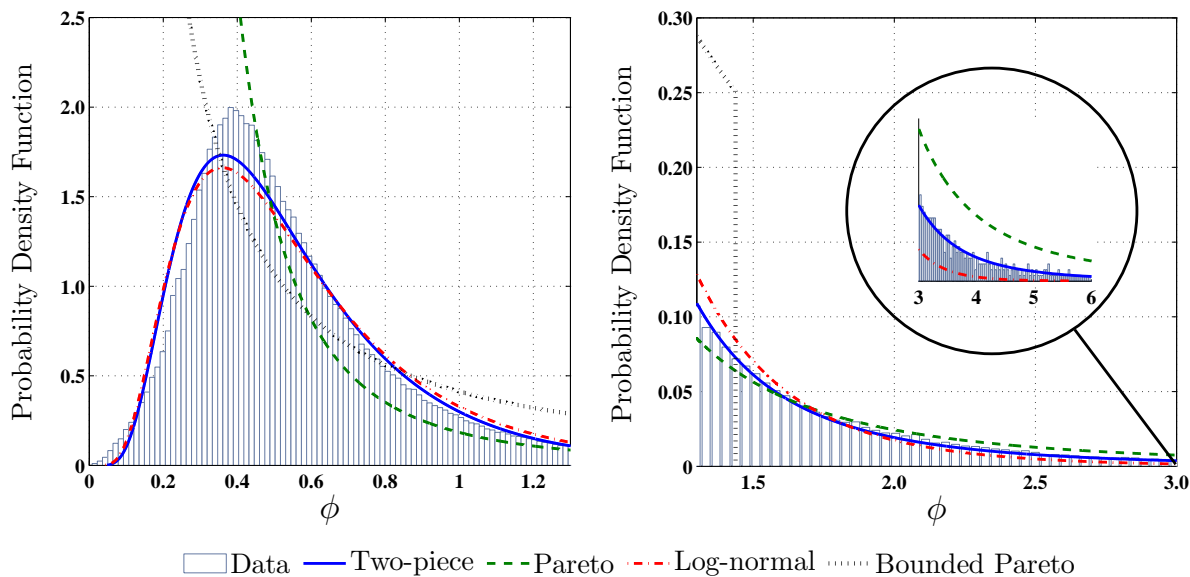

Figure 4: DENSITY OF TWO-PIECE, LOG-NORMAL AND PARETO VS. DATA

Overall, I conclude that while the Two-piece distribution serves as a good approximation of the empirical c.d.f. and p.d.f., Log-Normal and (un-)bounded Pareto exhibit substantial deviations from the data especially in the right tail of the distribution. Such deviations may entail non-trivial errors in the predictions of the welfare gains from trade in a heterogeneous firms trade model where the selection mechanism (into operating and exporting) depends crucially on the shape and location of the productivity distribution. I quantify these errors in a standard general equilibrium model of trade in the next section.

4 Workhorse Heterogenous Firm Trade Model

In this section, I sketch out a version of the Melitz (2003) model of international trade with heterogeneous firms. The model is standard and does not notably deviate from the main workhorse versions popular in the literature. The setup closely follows Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008) and only slightly deviates from Arkolakis, Costinot and Rodríguez-Clare (2012), and Melitz and Redding (2014).[9]

There are J countries in the world, each country $j \in J$ is populated by the L_j measure of homogeneous consumers that maximize utility according to the usual CES-type function by consuming different varieties denoted by ϕ:

---

[9]The version here assumes that the fixed cost of exporting is paid in terms of the labor in the importing country, whereas it is paid in terms of the domestic labor in Melitz and Redding (2014), and in terms of both in Arkolakis, Costinot and Rodríguez-Clare (2012). The results do not rely on this assumption.

$$U_j = \left(\sum_{i \in J} \int_{\Omega_{ij}} q_{ij}(\phi)^{\frac{\sigma-1}{\sigma}} d\phi \right)^{\frac{\sigma}{\sigma-1}}.$$

where Ω_{ij} is the set of goods from i available in j, and σ is the usual elasticity of the substitution parameter. Consumer optimization leads to the following expressions for the demand for each variety and the CES price index:

$$x_{ij}(\phi) = \frac{1}{p_{ij}(\phi)} \left(\frac{p_{ij}(\phi)}{P_j} \right)^{1-\sigma} L_j w_j \;\; \text{and} \;\; P_j^{1-\sigma} = \sum_{i \in N} \int_0^{\bar{\phi}} p_{ij}(\phi)^{1-\sigma} d\phi.$$

Firms are heterogeneous in terms of the productivity parameter $\phi \in (0, \bar{\phi})$ where $\bar{\phi}$ is infinity when the productivity distribution is unbounded from the right, and a positive constant otherwise. They employ domestic labor for production and entry cost, and foreign labor for the fixed cost of exporting and pay wages w_i and w_j per unit of labor, respectively. With a slight derivation of the notation, let me also use ϕ to denote a productivity parameter such that each variety is associated with a certain productivity level. Then, firms from i maximize their profit in market j according to the following function:

$$\pi_{ij}(\phi) = \left(\frac{p_{ij}(\phi)}{P_j} \right)^{1-\sigma} L_j w_j - \frac{w_i}{\phi} p_{ij}(\phi)^{-\sigma} \tau_{ij} P_j^{\sigma-1} L_j w_j - w_j f_{ij},$$

where f_{ij} is the fixed cost of exporting from i to j in terms of L_j. Taking the derivative with respect to $p_{ij}(\phi)$ leads to the usual pricing equation:

$$p_{ij}(\phi) = \frac{\sigma}{\sigma-1} \frac{w_i}{\phi} \tau_{ij}.$$

Not all firms in i will choose to export to j but only those that have productivity higher than the cut-off defined as:

$$\phi_{ij}^* = \left(\frac{\sigma}{L_j} \right)^{\frac{1}{\sigma-1}} \left(\frac{\sigma}{\sigma-1} w_i \tau_{ij} \right) P_j^{-1} f_{ij}^{\frac{1}{\sigma-1}}. \tag{6}$$

Note that in the empirical section, I make use of the expression for revenues of all firms from i in their domestic market:

$$r_{ii}(\phi) = x_{ii}(\phi) p_{ii}(\phi) = \left(\frac{\sigma}{\sigma-1} \frac{w_i}{\phi} \tau_{ii} \right)^{1-\sigma} P_i^{\sigma-1} L_i w_i. \tag{7}$$

Without loss of generality, let me normalize revenues of an average firm such that its productivity parameter is unity. Then, dividing equation (7) by the sample average and taking it to the power $\frac{1}{\sigma-1}$ allows calculating efficiency. This is the measure that I use in the empirical section of the paper.

Upon paying a fixed entry cost, f_i^e, firms can draw the productivity parameter and decide on whether to produce and serve certain markets or exit. In equilibrium, the expected profits must be zero such that the expected revenues exactly cover the entry cost:

$$\sum_{j \in J} \left(\int_{\phi_{ij}^*}^{\bar{\phi}} w_j f_{ij} (\phi_{ij}^*)^{1-\sigma} \phi^{\sigma-1} f(\phi) d\phi - \int_{\phi_{ij}^*}^{\bar{\phi}} w_j f_{ij} f(\phi) d\phi \right) = w_i f_i^e, \tag{8}$$

where $F(\phi)$ and $f(\phi)$ denote c.d.f and p.d.f. of the productivity parameters. Finally, there is a labor market clearing condition which says that domestic labor is used up in domestic production, paying the entry costs and fixed costs by foreign firms:

$$\frac{N_i}{1 - F(\phi_{ii}^*)} \sum_{j \in J} \left(\frac{(\sigma-1)w_j}{w_i} f_{ij} (\phi_{ij}^*)^{1-\sigma} \int_{\phi_{ij}^*}^{\bar{\phi}} \phi^{\sigma-1} f(\phi) d\phi + f_i^e \right) + \sum_{j \in J} \frac{N_j}{1 - F(\phi_{jj}^*)} f_{ji} \int_{\phi_{ji}^*}^{\bar{\phi}} f(\phi) d\phi = L_i. \tag{9}$$

Upon the choice of the numéraire , $w_1 = 1$, equations (6), (8) and (9) solve the model. Then, the welfare of consumers in i can be measured as the ratio of wages to the price index, w_i/P_i. In the counterfactual exercises that follow, I exogenously change τ_{ij} to some new values τ_{ij}' and express change in consumer welfare as:

$$\text{Welfare Gains} = 100\% \times \left(\frac{w_i(\tau)}{P_i(\tau)} \frac{P_i(\tau')}{w_i(\tau')} - 1 \right), \tag{10}$$

where τ and τ' are $J \times J$ matrices of the initial and counterfactual levels of variable trade costs, respectively.

Note that in terms of the shape and location of the productivity distribution, the model solution involves two important selection statistics: (i) $1 - F(\phi_{ij}^*)$, which measures the probability of firms from i being active in j and (ii) $\int_{\phi_{ij}^*}^{\bar{\phi}} \phi^{\sigma-1} f(\phi) d\phi$, which is required to calculate total revenues of firms from i in market j. Note that the third necessary statistics, $\int_{\phi_{ij}^*}^{\bar{\phi}} f(\phi) d\phi$, which enters equations (8) and (9) is identical to (i) due to the following:

$$\int_{\phi_{ij}^*}^{\bar{\phi}} f(\phi) d\phi = \int_0^{\bar{\phi}} f(\phi) d\phi - \int_0^{\phi_{ij}^*} f(\phi) d\phi = 1 - F(\phi_{ij}^*).$$

As selection statistics (i) and (ii) determine the equilibrium outcome, the shapes of $F(\cdot)$ and $f(\cdot)$ are central in determining the size of the gains from trade under a hypothetical reduction in variable trade costs. I provide explicit expressions for (i) and (ii) under four considered parametric distributions as well as full derivation details in the Appendix.

To demonstrate the virtues of the Two-piece distribution relative to the alternatives, I design and conduct two counterfactual experiments using the described theoretical model. The two exercises are specifically designed such as to flesh out errors in the estimates of the welfare gains from trade caused by deviations in the parametric distributions from the actual one in the simplest way. I will show that, though depending on the underlying economic primitives, Log-normal and (un-)bounded

11

Pareto can vary in the magnitudes of the associated errors between each other; they always produce larger errors than the Two-piece distribution. Without loss of generality, the model's primitives are chosen as follows:

Parameter	J	L_1	L_2	f_1^e	f_2^e	σ
Value	2	100	50	1	1	4

Table 2: PRIMITIVES OF THE MODEL

The parameterization is rather stylized and intentionally so. While extending the analysis to a multi-sector/multi-country model would be straightforward and would magnify the results consistent with the argument in Ossa (2015), using this simplistic version allows carrying through the main point in a clear and concise way. The parameters of the Two-piece, Log-normal and bounded Pareto productivity distributions are taken directly from Table 1. The only exception is unbounded Pareto as its parameters in Table 1 are such that the shape parameter is lower than $\sigma - 1$ which violates the assumptions of the underlying model. I parameterize it by setting the shape and scale parameters to 3.2 and 0.001, respectively.

In both experiments, I gradually reduce the level of international variable trade costs, τ_{12} and τ_{21}, from 3 to unity while keeping intra-trade costs at unity. For every reduction in variable trade costs I calculate the *true* welfare gains given by equation (10) by using selection statistics calculated directly from the data via numerical methods. The first selection statistics, $1 - F(\phi_{ij}^*)$, are calculated using the function of an empirical c.d.f. which is readily available in any statistical package. The second selection statistics, $\int_{\phi_{ij}^*}^{\bar{\phi}} \phi^{\sigma-1} f(\phi) d\phi$, are calculated using trapezoidal numerical integration given observations on $\phi^{\sigma-1}$ and the cut-offs. Given *true* welfare gains, I can calculate errors implied by each of the parametric assumptions as the difference between the *true* and predicted gains.

The difference between the two experiments is in the level of the cost of exporting f_{ij} relative to other primitives of the model. These costs will determine the relative location of the cut-off firms on the support of ϕ. Because parametric distributions deviate from the data differently at different points on the support, e.g., Log-normal approximates the data better than Pareto in the lower tail and vice versa, the cost of exporting will govern the relative size of the errors implied by each distribution.

Experiment 1: Falling variable costs and low fixed export costs

In this experiment, I set $f_{ii} = 0.001$ and $f_{ij} = 0.25$ for $i \neq j$. I plot true welfare gains $(- \circ -)$ for the large (country 1) and small (country 2) economies in the left and right panels of Figure 5, respectively. As usual, the smaller country gains relatively more – a reduction in variable trade costs of 65 percent increases its welfare by 35 percent in comparison to 12 percent in the large economy. Next, I plot errors in the welfare gains implied by the Two-piece distribution (blue solid line), Log-normal (red dash-dotted line), unbounded Pareto (green dashed line) and bounded Pareto (black dotted line) for both countries.

First, note that the error term under the Two-piece distribution is negligible for both economies and that it is not the case under the other three distributions. At a relatively low cost of exporting, the cut-off value of ϕ varies along relatively low values when many firms choose to export. Naturally, as Figure 3 suggested in that interval of the support, the Log-normal and bounded Pareto distributions fit the data well, whereas (un-)bounded Pareto does not. Hence, the magnitude of the error term under Log-normal and bounded Pareto is relatively lower, i.e., in the case of the large economy it amounts to 2.1 and 2.4 percentage points, respectively. This is nearly one fifth of the total gains from trade predicted by the empirical benchmark. Since the errors are defined as the difference between the true gains and their predictions, all three distributions significantly underpredict the actual gains from trade.
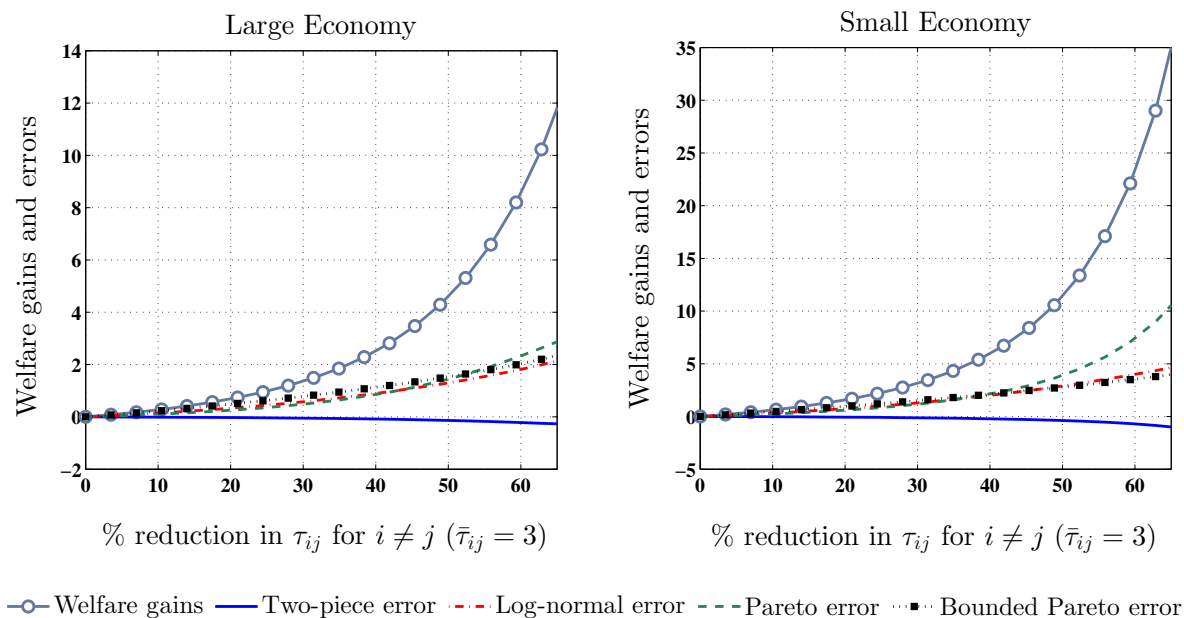
Figure 5: BENCHMARK WELFARE GAINS AND ERRORS: EXPERIMENT 1

The error implied by unbounded Pareto is about 2.9 percentage points. Moreover, in the case of the small economy the magnitude of the errors is higher. For that country, at 65 percent reduction in variable trade costs, assumptions of Log-normal and bounded Pareto entail errors of around 4.6 and 3.9 percentage points, respectively, whereas unbounded Pareto performs even worse with an error of about 10.4 percentage points. Overall, the results of this experiment suggest that at low fixed export costs, all distributions but Two-piece produce sizable errors in the predictions of the gains from trade.

Experiment 2: Falling variable costs and high fixed export costs

In the second experiment, I set $f_{ii} = 0.1$ and $f_{ij} = 1$ for $i \neq j$ such that the fixed exporting costs are relatively high. Hence, relative to the first experiment, all exporters will operate on the interval of the support closer to the right-tail. As before, I plot the results for the large and small economy in the left and right panels, respectively.

Figure 6: BENCHMARK WELFARE GAINS AND ERRORS: EXPERIMENT 2

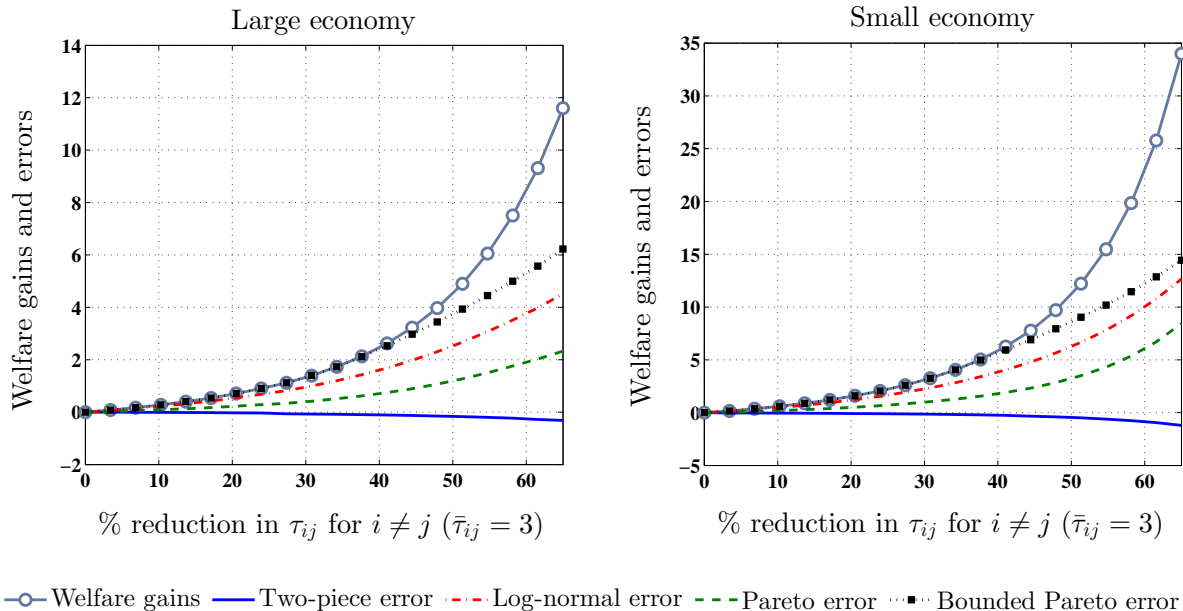In this experiment, the Two-piece distribution again performs well and produces negligible errors for both countries. The other three distributions, however, produce considerable errors. In terms of the gains of the large country, unbounded Pareto now produces smaller errors in comparison to both Log-normal and bounded Pareto. The reason for this, is that firms now operate on the interval of the support close to the right tail where Pareto fits relatively better. However, one should note that consistent with the results in Arkolakis, Costinot and Rodríguez-Clare (2012), Pareto predictions are invariant to changes in f_{ij} such that the differences in the error term stem from different predictions of the true gains from trade. The bounded Pareto performs particularly badly due to the truncated right tail, i.e., at extremely high levels of cut-offs no firms find it profitable to export such that no gains from trade are realized.[10]The assumptions of the Log-normal, bounded Pareto and unbounded Pareto distributions entail errors of 4.5, 6.2 and 2.3 percentage points, respectively. The errors are even larger in the case of the small economy. Again, all three distributions produce errors of significant magnitude relative to the total size of the gains from trade and underestimate the true gains by a sizable (sometimes by one half) margin.

The Two-piece distribution also performs favorably relative to the alternatives in terms of predicting other trade outcomes in both experiments. I measure its performance in terms of two additional trade outcomes: share of intra-trade and share of exporters denoted as λ_{ii} and χ_i, respectively. These two measures encompass intensive and extensive margins of trade that are often of interest (for example see Hummels and Klenow, 2005). I define root mean squared errors as follows:.

---

[10]This is reminiscent of the argument in Helpman, Melitz and Rubinstein (2008).

$$MSE_\ell(\lambda) = \sqrt{\frac{1}{J} \sum_j (\lambda_{jj} - \lambda_{jj,\ell})^2}; \quad MSE_\ell(\chi) = \sqrt{\frac{1}{J} \sum_j (\chi_j - \chi_{j,\ell})^2}, \tag{11}$$

where λ_{jj} and χ_j are *true* trade outcomes implied by the empirical benchmark, whereas $\lambda_{jj,\ell}$ and $\chi_{j,\ell}$ are their counterparts implied by the ℓ parametric distribution. I report calculated mean squared errors for the two outcomes in both experiments in Table 3. For brevity, the results are reported for four values of variable trade costs ranging between 3 and 1.2. In both experiments and in terms of both trade outcomes, the Two-piece distribution produces negligible errors in comparison to the other three distributions.

	Variable	Share of intratrade				Share of Exporters			
	τ_{ij} for $i \neq j$	3.0	2.4	1.8	1.2	3.0	2.4	1.8	1.2
Exp. 1	Two-piece	1.57	2.69	5.02	7.53	0.04	0.08	0.20	1.85
	Log-normal	31.38	51.16	78.91	81.65	2.29	7.80	27.77	93.02
	Bounded Pareto	12.72	22.06	40.39	63.01	0.51	1.02	2.25	7.20
	Pareto	37.30	71.05	121.29	75.83	0.75	1.51	26.44	155.79
Exp. 2	Two-piece	1.49	2.59	4.57	6.67	0.27	0.76	2.41	6.32
	Log-normal	21.61	29.11	33.42	22.13	23.66	54.58	126.29	265.54
	Bounded Pareto	14.79	25.99	48.96	80.46	1.92	3.88	9.25	23.73
	Pareto	35.13	34.51	29.28	16.23	8.67	91.78	197.69	326.43

Table notes: For expositional purposes, due to the fractional nature of the variables the results are reported in one thousandths.

Table 3: MEAN SQUARED ERRORS IN THE SHARE OF INTRATRADE AND EXPORTERS

Overall, quantitative evidence analyzed in this section suggests that the Two-piece distribution accurately approximates the empirical distribution and selection statistics, which allows making virtually no errors when calculating counterfactual trade outcomes. On the other hand, alternative distributions such as Log-normal and (un-)bounded Pareto can produce misleading results.

5 Sensitivity Analysis and Extensions

In this section, I test the robustness of the main results in several important dimensions. First, I ask whether more general specifications of Pareto and Log-normal such as Generalized Pareto and Three-parameter Log-normal can outperform the proposed Two-piece model. Second, I analyze the robustness of the results with respect to using alternative data and/or measures of productivity. Specifically, I ask if the results are driven by (i) a particular choice of the country, (ii) extreme outliers in the right or left tail of the data, (iii) a particular measure of productivity. As it turns out the answer to all three questions is no. I also suggest several theoretical and empirical avenues for extending the proposed approach.

5.1 Generalized Pareto and Three-parameter Log-normal

So far, I have compared the proposed Two-piece distribution to conventional Pareto and Log-normal two-parameter distributions. Admittedly, the Two-piece distribution features an additional parameter in comparison to these two distributions and equal number of parameters relative to the bounded Pareto case. To check whether the main results are driven by these restrictions, here I examine the fit of more general parametric models that are infrequently used but have the same number of parameters as the Two-piece distribution: Generalized Pareto and Three-parameter Log-normal with the following c.d.f.s:

$$F_{GP}(\phi) = 1 - \left(1 + \frac{\eta(\phi - \psi)}{\xi}\right)^{-1/\eta} \text{ and } F_{TLN}(\phi) = \Phi\left(\frac{\ln(\phi - \nu) - \mu}{\delta}\right), \tag{12}$$

Given expressions for the two c.d.f.s, I derive the associated quantile functions and apply the QQ-estimator. The results are reported in Table 4. Relative to its two-parameter counterpart, the Generalized Pareto distribution fits the data significantly better on the entire support as well as on different intervals of the support. For example, the total RMSE goes down from 0.236 to 0.144. However, it still yields to the Two-piece distribution everywhere and especially so in the bottom and top 5 percent of the data. The same is largely true for the Three-parameter Log-normal distribution, i.e., relative to the conventional Log-normal, Three-parameter Log-normal fits the data better when the entire support is considered and in the bottom 5 percent; however, it performs worse for the top 5 percent of the data. Three-parameter Log-normal is still unable to outperform the Two-piece distribution on the entire support and in the right tail.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Generalized Pareto	5.399	-0.081	0.424	0.144	1.011	0.512	0.196	0.409
	(0.354)	(0.006)	(0.003)					
3 Param. Log-Normal	0.518	-0.608	-0.042	0.067	0.252	0.120	0.195	0.367
	(0.004)	(0.009)	(0.004)					

Table notes: (I), (II) and (III) refer to the shape, location and scale parameters. All parameters are estimated using 100,000 quantile data points.

Table 4: Estimation Results (Alternative Parametric Distributions)

Overall, the results in Table 4 suggest that versions of Pareto and Log-normal distributions with the same number of parameters as the Two-piece distribution are still unable to fit the empirical distribution well especially in the right tail. Hence, the main results of the paper are also applicable to the class of three-parameter distributions.

5.2 Are the results sensitive to the choice of country?

Up to now, I have used data on French firms. However, it is important to check whether the data from other countries exhibits similar patterns. The data on France is a good starting point as it is the only country in the ORBIS dataset that has sufficient observations on both domestic sales and export revenues which allows obtaining clean measures of domestic sales necessary for calculating measures of efficiency consistent with Melitz (2003). This, however, is problematic for other countries as observations on export revenues are generally not available. Hence, results reported in this section are based on total sales (gross of export revenues), which arguably lead to a noisier measure of efficiency.

Country	Parameters			Root Mean Squared Error			
	(I)	(II)	(III)	Two-piece	Log-normal	Pareto	Bounded Pareto
France	2.880	1.057	0.923	0.053	0.070	0.228	0.181
	(0.006)	(0.004)	(0.001)				
Hungary	2.710	1.243	0.954	0.048	0.059	0.275	0.192
	(0.010)	(0.008)	(0.001)				
Italy	3.820	2.463	0.993	0.068	0.068	0.308	0.191
	(0.018)	(0.025)	(0.001)				
Japan	2.723	1.143	0.921	0.043	0.060	0.215	0.143
	(0.014)	(0.009)	(0.002)				
Norway	3.842	3.088	0.997	0.064	0.064	0.345	0.182
	(0.025)	(0.042)	(0.000)				
Portugal	2.637	0.925	0.885	0.038	0.070	0.198	0.154
	(0.008)	(0.004)	(0.001)				
Romania	2.672	1.260	0.954	0.048	0.059	0.283	0.202
	(0.010)	(0.008)	(0.001)				
Spain	3.049	1.308	0.952	0.053	0.061	0.247	0.181
	(0.008)	(0.006)	(0.001)				
Sweden	3.471	2.035	0.988	0.071	0.073	0.317	0.195
	(0.014)	(0.015)	(0.001)				
Ukraine	2.849	1.952	0.986	0.037	0.040	0.340	0.204
	(0.020)	(0.028)	(0.001)				
Average	3.065	1.648	0.955	0.052	0.062	0.276	0.182

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in the case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 5: ESTIMATION RESULTS FOR DIFFERENT COUNTRIES

Following the same methodology as in Section 3, I use data on countries where more than 100,000 initial observations are available. The results are reported in Table 5. Taking results for France as a benchmark, I see that using total sales instead of domestic sales leads to a lower estimate of the shape parameter and, perhaps more importantly, to lower values of ρ, which now implies a slightly

larger Pareto tail. This is not surprising as including export revenues results in a fatter right tail which the estimator interprets as a higher share of observations following Pareto.

Results in Table 5 suggest several other important insights. First, the Two-piece distribution performs strictly better than Log-normal and (un-)bounded Pareto for all countries where the estimated ρ is sufficiently far from unity. However, even in cases when the cut-off point is close to unity, e.g., Norway and Italy, it still performs at least as good as the next best option. Values of the parameters averaged across all countries are very close to those employed in Sections 3 and 4 such that the main results are robust to using a larger sample of countries and are not specific to France.

5.3 How important are data points at the extremes?

One may wonder if the main results of the analysis are driven by relatively rare data points located at the extremes. To address this possible concern, I repeat the estimation while sequentially removing 1000 data points from the right and left tails of the original data. First, I remove 1000 firms with the highest measured productivities which constitutes to about 0.1 percent of the original sample. The results are reported in Table 6.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	3.338	1.326	0.959	0.060	0.457	0.217	0.073	0.133
	(0.023)	(0.012)	(0.001)					
Log-normal	0.564	-0.704		0.064	0.425	0.199	0.113	0.192
	(0.001)	(0.001)						
Pareto	1.945	0.296		0.240	1.411	0.845	0.382	0.755
	(0.005)	(0.001)						
Bounded Pareto	0.274	0.209	1.373	0.177	1.086	0.567	0.378	0.704
	(0.025)	(0.001)	(0.012)					

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 6: Truncated Distribution (from the right)

Upon excluding the top 1000 observations from the original data, the Two-piece distribution still dominates the other three alternatives in the overall fit to the data. It yields to the Log-normal only in the bottom 5 percent of the distribution and dominates (un-)bounded Pareto everywhere. The estimated parameters are slightly higher than in Table 1.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	2.983	1.133	0.930	0.044	0.308	0.156	0.028	0.041
	(0.004)	(0.003)	(0.001)					
Log-normal	0.563	-0.698		0.060	0.249	0.122	0.165	0.317
	(0.001)	(0.001)						
Pareto	1.923	0.296		0.224	1.242	0.780	0.339	0.639
	(0.004)	(0.001)						
Bounded Pareto	0.460	0.220	1.479	0.171	0.956	0.530	0.388	0.773
	(0.022)	(0.001)	(0.014)					

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 7: Truncated Distribution (from the left)

Next, I repeat the exercise but now trim the original sample from the left by removing the bottom 1000 observations. The results are presented in Table 7. Again, the results indicate significantly better fit of the Two-piece distribution in comparison to the alternatives. The difference is particularly large for the top 5 percent of available observations. The Two-piece distribution performs slightly worse than the Log-normal in the left tail which is consistent with previous results. Overall, fitting different models on truncated data that excludes extreme observations in the right or the left tails reveals that the general results are not driven by outliers and/or peculiarities of the data at the extremes.

5.4 Alternative measures of productivities

So far, I have analyzed the shape of the productivity distribution using a measure of productivity that is consistent with theories featuring heterogeneous firms and constant markups. Though this particular specification is still the workhorse of quantitative trade theory, there is another important class of models in which markups are no longer constant (for example see Bernard, Eaton, Jensen and Kortum, 2003; Melitz and Ottaviano, 2008; Simonovska, 2015; Edmond, Midrigan and Xu, 2015). A large class of models featuring variables markups and heterogeneous firms is analyzed in Arkolakis, Costinot, Donaldson, Rodríguez-Clare (2015). Mrazova, Neary and Parenti (2015) discuss how conditions on demand and technology shape the distribution of markups and sales. In these models, normalized relative domestic sales would not yield clean measures of productivity but rather the ratio of productivity to firm-specific markups:

$$r_{ii}(\phi) = \left(m(\phi) \frac{w_i}{\phi} \tau_{ii} \right)^{1-\sigma} P_i^{\sigma-1} L_i w_i, \tag{13}$$

where $m(\phi)$ is a firm-specific markup. Unfortunately, the data on firm-level markups are rarely available. However, to check for the robustness of the main results when applied to alternative trade models, I employ a measure of firm-specific markups that may be noisy but could provide some insights on the robustness of the proposed approach. I measure $m(\phi)$ as a ratio between firm's sales and the sum of its cost of employees and materials. As these data are not available for the whole sample, the procedure leaves me with a sample of 633,640 observations. Given the estimates of $m(\phi)$, I calculate the implied productivity parameters, ϕ, as before and use them in the estimation procedure.

I report the results in Table 8 which suggests that the Two-piece distribution outperforms the alternatives in terms of the predictive power when the entire support is considered. In comparison to Log-normal, Two-piece performs better overall and in the top 5 percent and worse in the bottom 5 percent which is consistent with the previous results. It also outperforms (un-)bounded Pareto everywhere except for the top 1 percent. Unbounded Pareto seems to be the best alternative in terms of matching the top 1 percent as the parameter on the right limit of the bounded Pareto distribution cannot be precisely identified. This is also reflected in the estimate of the third parameter of the Two-piece distribution that points to a large (over 22 percent) Pareto tail.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	1.997	1.224	0.773	0.162	1.491	0.673	0.166	0.362
	(0.009)	(0.014)	(0.005)					
Log-normal	0.621	-0.177		0.210	1.319	0.621	0.558	1.154
	(0.002)	(0.001)						
Pareto	1.653	0.457		0.252	2.208	1.056	0.197	0.266
	(0.003)	(0.001)						
Bounded Pareto	1.653	0.457	(2.E+09)	0.252	2.208	1.056	0.197	0.266
	(0.003)	(0.001)	(4.E+09)					

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in the case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 8: MEASURE OF PRODUCTIVITY UNDER VARIABLE MARKUPS

Naturally, firms may differ not only in their productivity and markups but also other characteristics such as wages (for example, see Egger and Kreickermeir, 2009; Helpman, Itskhoki and Redding, 2010; Davis and Harrigan, 2011) or quality of their products (see Verhoogen, 2008; Baldwin and Harrigan, 2011) and potentially in many other characteristics. While looking at every particular case is beyond the scope of the paper, it is possible to check the sensitivity of the results by using measures of productivity based on the cost function. Unfortunately, the data on cost-side variables are not widely available. The best country satisfying this requirement is Japan with a total of 219,454 observations available in 2012. I use data on total cost of goods sold and total cost of

employees to calibrate the productivity parameter from the following relationship:

$$c(\phi)q = \frac{w(\phi)q}{\phi},$$

where $c(\phi)q$ and $w(\phi)q$ are the observed total cost of goods sold and observed cost of employees, respectively. I then estimate the distribution of backed out ϕ applying the QQ-estimator for four parametric models. The results are in Table 9.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	1.085	0.197	0.826	0.216	1.594	0.756	0.384	0.810
	(0.004)	(0.002)	(0.002)					
Log-normal	1.228	-2.656		0.314	1.267	0.593	1.053	2.219
	(0.003)	(0.001)						
Pareto	0.853	0.022		0.482	3.366	1.872	0.384	0.264
	(0.001)	(0.001)						
Bounded Pareto	0.853	0.022	3.E+04	0.482	3.366	1.872	0.384	0.254
	(0.075)	(0.002)	(3.E+04)					

Table notes: In the case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in the case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in the case of the Pareto, (i) and (II) refer to the shape and scale parameters; in the case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 9: Cost-based measure of productivity

Again the Two-piece model fits the data better than any of the alternative three models when considering the entire support. It performs slightly worse than Log-normal in the bottom 5 percent and slightly worse than Pareto in the top 1 percent. This is due to the fact that the true distribution of the cost-based measure is generally more skewed than the one based on sales. This is also confirmed by the value of the estimated $\rho = 0.8260$ which gives a significantly larger weight to the Pareto tail. Also note that due to the skewness, the upper bound of the Bounded Pareto could not be estimated precisely and the shape and the lower bound parameters are nearly identical to the unbounded Pareto. Overall, the main results in this section suggest that the Two-piece distribution is preferable to Log-normal and (un-)bounded Pareto due to closer fit over the entire support.

5.5 Truncation from the right

Several papers argue that unbounded support usually assumed in the literature is a strong assumption (see Helpman, Melitz, Rubinstein, 2007; Feenstra, 2014; Melitz and Redding, 2015). Though I have demonstrated that an unbounded Two-piece distribution outperforms bounded Pareto and does not deviate significantly from the empirical distribution, it may be of value to point out that the extension featuring a mixed distribution with a truncation from the right is possible. Such an extension would be straightforward and would require mixing the following two p.d.f.s:

$$f_L(\phi) = \frac{1}{\sqrt{2\pi}s\phi}e^{-\frac{1}{2}\left(\frac{\ln\phi-\mu}{s}\right)^2} \text{ and } f_{BP}(\phi) = \frac{\alpha\theta^\alpha\phi^{-\alpha-1}}{1-\theta^\alpha\phi_h^{-\alpha}},$$

where ϕ_h would serve as an upper bound. Applying continuity and differentiability conditions and following the same derivations as in the case of the unbounded Pareto (described in the Appendix), one could get a version of a Two-piece distribution bounded from the right.

5.6 External Validity: City Size Distribution

The virtues of the Two-piece distribution can be explored using alternative data sets where the choice between Log-normal and Pareto is not obvious. Perhaps, one of the most well-known of such cases in economics is an ongoing debate on the shape of the city size distribution (see Gabaix, 1999; Eeckhout, 2004, 2009; Levy, 2009). Considering this debate through the lens of the proposed approach can turn out to be particularly fruitful as in many ways it mirrors the current debate in international trade about the shape of the productivity distribution. I use data from Eeckhout (2009) to estimate four alternative parametric models and report the results in Table 10.[11]

First, note that the estimated shape parameter of the Log-normal model is exactly the same as in Eeckhout (2004) which points to the robustness of the results of the QQ-estimator to alternative estimation methods. The location parameter is different due to different units of measurement, i.e., I normalize the data such that the mean is unity. Next, according to the values of RMSE, the Two-piece distribution outperforms Log-normal when considering the whole support and/or any intervals on the support. Finally, estimated parameters of the Two-piece distribution suggest that, indeed, the upper-tail with a measure of 2.22 percent is distributed Pareto.

	Parameters			Root Mean Squared Error				
	(I)	(II)	(III)	All	Bottom 1%	Bottom 5%	Top 5%	Top 1%
Two-piece	1.205	6.178	0.978	0.089	0.209	0.222	0.155	0.256
	(0.034)	(0.774)	(0.003)					
Log-normal	1.751	-1.738		0.099	0.257	0.261	0.191	0.322
	(0.001)	(0.001)						
Pareto	0.619	0.035		0.679	2.909	1.900	1.412	2.728
	(0.004)	(0.001)						
Bounded Pareto	0.126	0.013	5.264	0.354	1.932	1.029	0.868	1.745
	(0.006)	(0.001)	(0.136)					

Table notes: In case of the Two-piece distribution, parameter (I) refers to the shape parameter, α, (II) and (III) refer to the scale parameters, θ and ρ, respectively; in case of the Log-normal distribution, (I) and (II) refer to the scale and location parameters; in case of the Pareto, (i) and (II) refer to the shape and scale parameters; in case of the Bounded Pareto distribution, (I) refers to the shape parameter and (II) and (III) to two location parameters. All parameters are estimated using 100,000 quantile data points.

Table 10: Estimation Results for the City Size Distribution

---

[11]The data are available at `https://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1676`

Let me consider exactly this interval on the support in the right tail to flesh out the difference between Two-piece and Log-normal. I plot the c.d.f.s implied by the respective parameter values along with the empirical c.d.f. for the top 2 percent of the distribution in Figure 7. The results are unambiguous – the Two-piece distribution (blue solid line) outperforms the Log-normal (red dashed line) by a sizable margin.
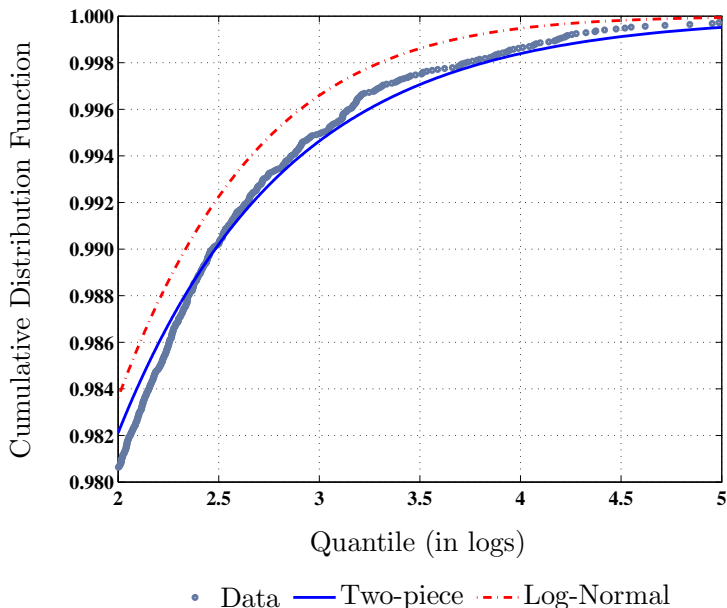

Figure 7: CITY SIZE C.D.F: TWO-PIECE AND LOG-NORMAL VS. DATA

This result lends some evidence in support of the original claim in Gabaix (1999), which is now based on the entire support of the distribution rather than a particular interval thereof and suggests that the upper tail of the city size distribution follows a power law.

6 Conclusion

This work has emphasized that the most often used parametric assumptions about the shape of the productivity distribution fail to come to grips with the data. While the Log-normal distribution fails to capture the shape of the right tail, (un-)bounded Pareto misses on the left one. I use the workhorse quantitative general equilibrium model of trade with heterogeneous firms and show that these deviations from the data can be harmful for the correct calculation of different trade outcomes including the gains from trade.

As a remedy for these problems, I have proposed using a parametric distribution that models the left tail as Log-normal and right tail as Pareto. I use micro data from different countries and estimate the parameters of the proposed model to show that this distribution fits the data better than the alternatives along the entire support and produces negligible errors in counterfactual

exercises. I also provide several possible avenues for further extensions that may prove to be useful in international trade as well as in other areas of economics.

References

1. Arkolakis, C. 2010. "Market Penetration Costs and the New Consumers Margin in International Trade." Journal of Political Economy, 118: 1151–1199.

2. Arkolakis, C. 2015. "A Unified Theory of Firm Selection and Growth." NBER Working Paper No. 17553.

3. Arkolakis, C., Costinot A., and A. Rodriguez-Clare. 2012. "New Trade Models, Same Old Gains?" American Economic Review, 102(1): 94–130.

4. Arkolakis, C., Costinot, A., Donaldson, D. and A. Rodriguez-Clare. 2015. "The elusive procompetitive effects of trade." NBER Working Paper No. 21370

5. Arkolakis, C., Demidova S., Klenow P.J. and A. Rodriguez-Clare. 2008. "Endogenous Variety and the Gains from Trade." American Economic Review, 98(2): 444-50.

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

7. Baldwin R. 2005. "Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model." NBER Working Paper No. 11471.

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

9. Baldwin, R. and J. Harrigan. 2011. "Zeros, Quality, and Space: Trade Theory and Trade Evidence." American Economic Journal: Microeconomics, 3(2): 60–88.

10. Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." American Economic Review, 93: 1268 – 1290.

11. Chaney, T. 2008. "Distorted Gravity: Heterogeneous Firms, Market Structure, and the Geography of International Trade." American Economic Review, 98: 1707–1721

12. Cooray K. and M. Ananda. 2005. "Modeling actuarial data with a composite lognormal-Pareto model." Scandinavian Actuarial Journal, 321–334.

13. Davis, D. and J. Harrigan. 2011. "Good jobs, bad jobs, and trade liberalization." Journal of International Economics 84(1): 26–36.

14. di Giovanni, J., A. A. Levchenko, and R. Ranciere. 2011. "Power laws in firm size and openness to trade: Measurement and implications." Journal of International Economics, 85: 42 – 52.

15. di Giovanni D. and A. A. Levchenko. 2012. "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies." Journal of Political Economy, 120(6): 1083 – 1132.

16. Eaton, J., S. Kortum, and F. Kramarz. 2004. "Dissecting Trade: Firms, Industries, and Export Destinations." American Economic Review, Papers and Proceedings, 94: 150–154

17. Eaton J., Kortum S. and F. Kramarz. 2011. "An Anatomy of International Trade: Evidence From French Firms." Econometrica, 79(5): 1453–1498.

18. Egger, H. and U. Kreickemeier. 2009. "Firm Heterogeneity And The Labor Market Effects Of Trade Liberalization." International Economic Review, 50(1): 187–216.

19. Edmond, C., V. Midrigan and D. Y. Xu. 2015. "Competition, Markups, and the Gains from International Trade." American Economic Review, 105(10): 3183–3221.

20. Eeckhout, J. 2004. Gibrat's Law for (All) Cities. American Economic Review, 94(5): 1429–1451.

21. Eeckhout, J. 2009. "Gibrat's Law for (All) cities: A reply." American Economic Review, 99: 1676–1683.

22. Feenstra, R. C. 2014. "Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity." Mimeo.

23. Fernandes, A., Klenow, P., Meleshchuk, S., Pierola, M. D. and Rodriguez-Clare, U. A. 2015. "The intensive margin puzzle in trade." Mimeo.

24. Freund, C. and M. D. Pierola. 2015. "Export superstars." Review of Economics and Statistics, 97(5): 1023 – 1032.

25. Gabaix, X. 1999. "Zipf's Law for Cities: An Explanation." Quarterly Journal of Economics, 94: 739–767

26. Head, K., Mayer, T. and Thoenig, M. 2014. Welfare and trade without pareto. American Economic Review, 104(5): 310–16.

27. Helpman, E., M. J. Melitz and Y. Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." Quarterly Journal of Economics, 123: 441–487

28. Helpman, E., O. Itskhoki and S. Redding, 2010. "Inequality and Unemployment in a Global Economy," Econometrica, 78(4):1239–1283.

29. Luttmer, E. 2007. "Selection, Growth, and the Size Distribution of Firms." Quarterly Journal of Economics, 122, 1103–1144.

30. Levy, M. 2009. "Gibrat's Law for (all) cities: Comment." American Economic Review, 99: 1672–1675.

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

32. Melitz, M. J. and S. J. Redding. 2015. "New trade models, new welfare implications." American Economic Review, 105(3): 1105–46.

33. Melitz, . J. and G. I. P. Ottaviano. 2008. "Market Size, Trade, and Productivity." The Review of Economic Studies, 75(1):295–316.

34. M. Mrazova, P. Neary and M. Parenti. 2015. "Technology, demand, and the size distribution of firms." University of Oxford Working Papers.

35. Orbis. 2015. Bureau van Dijk. Available at https://orbis.bvdep.com/

36. Ossa R. 2015. "Why trade matters after all." Journal of International Economics, 97(2): 266–277.

37. Scollnik. D. 2007. "On composite lognormal-Pareto models." Scandinavian Actuarial Journal, 20–33.

38. Simon, H. A. 1955. "On a Class of Skew Distribution Functions." Biometrika 42: 425–440.

39. Simon, H. A. and C. P. Bonini. 1958. "The Size Distribution of Business Firms." American Economic Review, 98: 607–617.

40. Simonovska I. 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." Review of Economic Studies, 82(4):1612–1656.

41. Verhoogen, E. 2008. "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector." Quarterly Journal of Economics, 123: 489–530.

42. Yang, M.J. 2014. "Firm heterogeneity and trade in a distorted gravity world." Mimeo.

Appendix A: Derivation of Two-piece distribution

Mixed distributions that combine the Log-normal and Pareto distributions were first developed by the mathematicians Cooray and Ananda (2005) and Scollnik (2007). The distribution proposed in this paper builds on and modifies the version in Scollnik (2007). Let a random variable x be distributed according to the following p.d.f.:

$$f(x) = \begin{cases} \kappa \omega f_L(x) \text{ for } x \in (0, \theta] \\ (1-\omega) f_P(x) \text{ for } x \in [\theta, \infty), \end{cases} \tag{14}$$

where, κ is a normalizing constant derived in what follows and $f_L(x)$ and $f_P(x)$ are the Log-normal and Pareto distributions with the following p.d.f.s:

$$f_L(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \text{ and } f_P(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}.$$

For $f(x)$ to be well-behaved, I impose two conditions: continuity and differentiability. The *continuity* condition ensures that $f(x)$ is continuous at θ and holds if and only if $\kappa \omega f_L(\theta) = (1-\omega) f_P(\theta)$ which entails:

$$\kappa \omega \frac{1}{\sqrt{2\pi}\sigma\theta} e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2} = (1-\omega)\frac{\alpha \theta^\alpha}{\theta^{\alpha+1}} \quad \Rightarrow \quad \kappa \omega = (1-\omega)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2}. \tag{15}$$

The *differentiability* condition requires $f(x)$ to be differentiable at θ and holds whenever $\kappa \omega f_L'(\theta) = (1-\omega) f_P'(\theta)$. First, let me derive the derivatives of the Log-normal component:

$$f_L'(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(-\frac{1}{x^2}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2} - e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2}\left(\frac{\ln x - \mu}{\sigma^2 x^2}\right)\right) = -\frac{1}{\sqrt{2\pi}\sigma x^2}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2}\left(1 + \frac{\ln x - \mu}{\sigma^2}\right). \tag{16}$$

The derivative of the Pareto component is as follows:

$$f_P'(x) = -\alpha(\alpha+1)\frac{\theta^\alpha}{x^{\alpha+2}}. \tag{17}$$

Then, by combining equations (16) and (17), I can specify the differentiability condition as:

$$\kappa \omega \frac{1}{\sqrt{2\pi}\sigma\theta^2}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2}\left(1 + \frac{\ln \theta - \mu}{\sigma^2}\right) = \alpha(\alpha+1)\frac{\theta^\alpha}{\theta^{\alpha+2}},$$

$$\Rightarrow \kappa \omega \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2}\left(1 + \frac{\ln \theta - \mu}{\sigma^2}\right) = (1-\omega)\alpha(\alpha+1). \tag{18}$$

Next, note that the condition in (15) implies:

$$\kappa \omega \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2} = \alpha(1-\omega)$$

$$\Rightarrow (\alpha+1)\kappa \omega \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln \theta - \mu}{\sigma}\right)^2} = (1-\omega)\alpha(\alpha+1),$$

which together with the condition (18) leads to the following:

$$\kappa\omega\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln\theta-\mu}{\sigma}\right)^2}\left(1+\frac{\ln\theta-\mu}{\sigma^2}\right) = (\alpha+1)\kappa\omega\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln\theta-\mu}{\sigma}\right)^2}$$

$$\Rightarrow \left(1+\frac{\ln\theta-\mu}{\sigma^2}\right) = \alpha+1 \quad \text{or} \quad \ln\theta-\mu=\alpha\sigma^2.$$

Plugging this result back into the continuity condition gives the following expression for the relative weight parameter ω which turns out to be a function of the other model's parameters:

$$\omega(\sigma,\alpha,\kappa) = \frac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+\kappa}. \tag{19}$$

Finally, to make sure the the overall p.d.f. is well-behaved, I define the *normalization* condition concerning κ which makes sure the the integral of the p.d.f. along the entire support is unity:

$$\int_0^\infty f(x)d(x) = \kappa\omega\int_0^\theta \frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}dx + \kappa(1-\omega)\int_\theta^\infty \frac{\alpha\theta^\alpha}{x^{\alpha+1}} = 1,$$

$$\Rightarrow \kappa\omega\int_{-\infty}^{\frac{\ln\theta-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt + \kappa(1-\omega)\left(-\frac{\theta^\alpha}{\theta^\alpha}+\frac{\theta^\alpha}{\theta^\alpha}\right) = \omega\kappa\Phi\left(\frac{\ln\theta-\mu}{\sigma}\right)+1-\omega = 1,$$

which puts the following restrictions on κ as a function of θ, μ and σ:

$$\kappa = \left(\Phi\left(\frac{\ln\theta-\mu}{\sigma}\right)\right)^{-1} = \left(\Phi(\sigma\alpha)\right)^{-1}.$$

Now plugging back into the equation (19) I get the following:

$$\omega(\sigma\alpha) = \frac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+[\Phi(\sigma\alpha)]^{-1}} = \frac{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}.$$

Combining this expression with an earlier result $\frac{\mu}{\sigma}=\frac{\ln\theta}{\sigma}-\alpha\sigma$, I can rewrite the original formulation of the probability distribution in the equation 14 as follows:

$$f(x) = \begin{cases} \dfrac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}\dfrac{\alpha}{x}e^{\frac{\ln\theta-\ln x}{\sigma}\left(\alpha\sigma+\frac{\ln\theta-\ln x}{2\sigma}\right)} & \text{for } x\in(0,\theta] \\[3mm] \dfrac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}\dfrac{\alpha}{x}\left(\dfrac{\theta}{x}\right)^\alpha & \text{for } x\in[\theta,\infty), \end{cases} \tag{20}$$

which is a function of one threshold parameter, θ and two shape parameters α and σ. Next, let me derive the c.d.f. of the Two-piece distribution starting with the upper-tail:

$$\begin{aligned} F(x|\theta\leq x) &= \int_\theta^x \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}\frac{\alpha\theta^\alpha}{t^{\alpha+1}}dt = c - \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}\frac{\theta^\alpha}{t^\alpha}\bigg|_\theta^x \\ &= c + \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1} - \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}+1}\frac{\theta^\alpha}{x^\alpha}, \end{aligned}$$

where c is an integrating constant. Using the limiting condition of any c.d.f. yields that:

$$c = 1 - \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}.$$

Then, the Pareto part of the c.d.f. can be expressed as follows:

$$F(x|\theta \leq x) = 1 - \frac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\frac{\theta^\alpha}{x^\alpha}.$$

Next, let me calculate the component of the cumulative distribution function corresponding to the Log-normal part:

$$
\begin{aligned}
F(x|0 \leq x \leq \theta) &= \frac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\int_0^x \frac{1}{\sqrt{2\pi}\sigma t}e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}dt \\
&= \frac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\int_{-\infty}^{\frac{\ln x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\
&= \frac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\Phi\left(\frac{\ln x - \mu}{\sigma}\right),
\end{aligned}
$$

Then, again using the fact that $\frac{\mu}{\sigma} = \frac{\ln\theta}{\sigma} - \alpha\sigma$, the c.d.f. may be expressed as follows:

$$F(x) = \begin{cases} \dfrac{\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\Phi\left(\alpha\sigma + \dfrac{\ln x - \ln\theta}{\sigma}\right) & \text{for } x \in (0, \theta] \\ 1 - \dfrac{1}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}\dfrac{\theta^\alpha}{x^\alpha} & \text{for } x \in [\theta, \infty). \end{cases} \tag{21}$$

Next, I use the fact that at $x = \theta$, the c.d.f. equals to the following:

$$F(\theta) \equiv \rho = \frac{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2}}{\Phi(\sigma\alpha)\sqrt{2\pi}\sigma\alpha e^{\frac{1}{2}(\alpha\sigma)^2} + 1}.$$

This expression provides a clear-cut condition for the permissible parameter space of the second shape parameter σ given α and the two scale parameter θ and ρ, and can be simplified to the following condition.

$$\Phi(\sigma\alpha)\sqrt{2\pi}(\sigma\alpha)e^{\frac{1}{2}(\sigma\alpha)^2} = \frac{\rho}{1 - \rho}.$$

Hence, one can express σ as an implicit function of α and ρ, i.e., $\sigma = s(\alpha, \rho)$. Then, rewriting the expression for c.d.f. in equation (21) gives:

$$F(x) = \begin{cases} \dfrac{\rho}{\Phi(s(\alpha,\rho)\alpha)}\Phi\left(\alpha s(\alpha,\rho) + \dfrac{\ln x - \ln\theta}{s(\alpha,\rho)}\right) & \text{for } x \in (0, \theta] \\ 1 - (1 - \rho)\dfrac{\theta^\alpha}{x^\alpha} & \text{for } x \in [\theta, \infty), \end{cases}$$

Finally, the probability density function in equation 20 can be rewritten as:

$$f(x) = \begin{cases} \dfrac{\rho}{\Phi(s(\alpha,\rho)\alpha)} \dfrac{1}{\sqrt{2\pi}s(\alpha,\rho)x} e^{-\frac{1}{2}\left(\alpha s(\alpha,\rho) - \frac{\ln\theta - \ln x}{s(\alpha,\rho)}\right)^2} & \text{for } x \in (0,\theta] \\[4mm] (1-\rho)\dfrac{\alpha\theta^\alpha}{x^{\alpha+1}} & \text{for } x \in [\theta,\infty), \end{cases}$$

where $F(x)$ and $f(x)$ are exact functions used in the main text.

Appendix B: Estimation and data

In Section 3, I use a QQ-estimator to pin down the parameters of the four distributions. Here, I provide exact functional forms used in the analysis. QQ-estimator minimizes the squared distance between (log) observed quantiles and (log) predicted quantiles for a given value of the c.d.f., F_q. Hence, the estimation involves inverting the c.d.f.s as:

$$
\begin{aligned}
Q_{Log-normal} &= \exp\left(\Phi^{-1}(F_q)s_l + \mu\right), \\
Q_{Pareto} &= x_m(1 - F_q)^{-\frac{1}{s_p}}, \\
Q_{BoundedPareto} &= \left(1 - F_q(1 - x_l^{s_b}x_h^{-s_b})\right)^{-\frac{1}{s_b+1}} x_l^{\frac{s_b}{s_b+1}}.
\end{aligned}
$$

In case of the Two-piece distribution, the c.d.f. must be inverted separately on the intervals $[0,\theta]$ and $[\theta,\infty)$:

$$
Q_{Two-piece} = \begin{cases} \theta\exp\left(s\left[\Phi^{-1}\left(\dfrac{F_q\Phi(s\alpha)}{\rho}\right) - \alpha s\right]\right) & \text{for } x \in (0,\theta] \\[4mm] \theta\left(\dfrac{1-\rho}{1-F_q}\right)^{\frac{1}{\alpha}} & \text{for } x \in [\theta,\infty). \end{cases}
$$

where s is constrained according to the following:

$$\Phi(s\alpha)\sqrt{2\pi}(s\alpha)e^{\frac{1}{2}(s\alpha)^2} = \frac{\rho}{1-\rho}.$$

Note that $Q_{Two-piece}$ distribution is a continuous function. This is easy to check by the identity that if $F_q = \rho$ then $x = \theta$ in both tails.

Before, I explain how the computational procedure works, let me describe the original dataset employed by the study. The original data on total sales and export revenues come from proprietary database of Bureau Van Dijk's Orbis database. The data have been cleaned to keep only unconsolidated accounts, and reflect balance sheet data from the latest available report in any single year for every entity. I also exclude firms classified as operating in financial and insurance sectors. To calculate domestic sales I subtract export revenues from total sales and keep only those observations that are non-negative. This leaves 928,569 observations for France in 2012. They are then normalized described in the main text as to reflect measures of productivity.

For computational purposes, it is inefficient to calculate F_q for all 928,569 because it (i) significantly slows down the optimization routine and (ii) does not increase precision of the estimated parameters. Instead, using original data I produce 100,000 empirical c.d.f. points each associated with an empirical quantile that numerically cover the interval on the support between 0.00001 and 0.99999. I then use these 100,000 data points for estimation. The results are completely robust to using larger (than 100,000 and up to the possible maximum) number of quantiles.

Appendix C: Selection statistics under different distribution models

Here, I derive the two selection statistics that depend on the shape and location of the productivity distribution and are necessary for calculating different trade outcomes. Note that for the *true* values of the two statistics, I use empirical c.d.f. function that computes the value of the c.d.f. for each variable in the data and is available in most statistical packages such as MATLAB where this command is *ecdf*. To calculate the second selection statistics, I take values of productivities to the power $\sigma - 1$ and integrate the resulting distribution from the left, starting with the value of $\phi = \phi_{ij}^*$, by employing trapezoid-based numerical integration. The corresponding command in MATLAB is *trapz*.

In terms of calculating selection statistics for parametric distributions, I start with the Two-piece distribution which gives expression for the first statistics by looking directly at the cumulative distribution function:

$$
1 - F(\phi_{ij}^*) =
\begin{cases}
1 - \dfrac{\rho}{\Phi(s\alpha)}\Phi\left(\alpha s + \dfrac{\ln \phi_{ij}^* - \ln \theta}{s}\right) & \text{for } \phi_{ij}^* \in (0, \theta] \\[3mm]
(1 - \rho)\dfrac{\theta^\alpha}{(\phi_{ij}^*)^\alpha} & \text{for } \phi_{ij}^* \in [\theta, \infty).
\end{cases}
$$

The second statistics is a bit more tricky to get. Let me start calculating the second selection statistics under for the interval where $\phi_{ij}^* \leq \theta$:

$$
\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi = \int_{\phi_{ij}^*}^{\theta} \phi^{\sigma-1} f(\phi) d\phi + \int_{\theta}^{\infty} \phi^{\sigma-1} f(\phi) d\phi. \tag{22}
$$

The first summand in equation (22) is:

$$
\int_{\phi_{ij}^*}^{\theta} \phi^{\sigma-1} f(\phi) d\phi = \int_0^{\theta} \phi^{\sigma-1} f(\phi) d\phi - \int_0^{\phi_{ij}^*} \phi^{\sigma-1} f(\phi) d\phi \tag{23}
$$

For further simplicity, let me first derive an expression for a generic power r and upper limit u:

$$
\int_0^u \frac{1}{\sqrt{2\pi} sx} x^r e^{-\frac{1}{2}\left(\alpha s + \frac{\ln x - \ln \theta}{s}\right)^2} dx
$$

First, I perform change of variables as follows:

$$
y = \alpha s + \frac{\ln x - \ln \theta}{s}, \quad x = \theta e^{s(y - \alpha s)}, \quad dx = \theta s e^{s(y - \alpha s)} dy.
$$

Plugging this back into the equation above gives the following:

$$\int_0^u \frac{1}{\sqrt{2\pi}s}\theta^{r-1}e^{s(y-\alpha s)(r-1)}e^{-\frac{1}{2}\left(\alpha s+\frac{\ln x - \ln\theta}{s}\right)^2}\theta s e^{s(y-\alpha s)}dy = \int_0^{\alpha s + \frac{\ln u - \ln\theta}{s}}\frac{1}{\sqrt{2\pi}}\theta^r e^{s(y-\alpha s)r}e^{-\frac{1}{2}y^2}dy$$

$$= \theta^r \int_0^{\alpha s + \frac{\ln u - \ln\theta}{s}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(y^2 - 2y(sr)+2(sr)(\alpha s)\right)}dy = \theta^r \int_0^{\alpha s + \frac{\ln u - \ln\theta}{s}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(y^2 - 2y(sr)+2(sr)(\alpha s)\right)}dy$$

$$= \theta^r e^{\frac{(sr)^2 - 2(sr)(\alpha s)}{2}}\int_0^{\alpha s + \frac{\ln u - \ln\theta}{s}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-(sr))^2}dy.$$

The second change of variables:

$$y - (sr) = v; \qquad y = v + sr; \qquad dy = dv.$$

Plugging back into the equation gives:

$$\theta^r e^{\frac{(sr)^2 - 2(sr)(\alpha s)}{2}}\int_0^{\alpha s + \frac{\ln u - \ln\theta}{s} - sr}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(v)^2}dv = \theta^r e^{\frac{(sr)^2 - 2(sr)(\alpha s)}{2}}\Phi\left(s(\alpha - r) + \frac{\ln u - \ln\theta}{s}\right)$$

I use this result to calculate first part of equation (23) by setting $u = \theta$ and $r = \sigma - 1$:

$$\int_0^\theta \phi^{\sigma-1}f(\phi)d\phi = \frac{\rho}{\Phi(s\alpha)}\theta^{\sigma-1}e^{\frac{(s(\sigma-1))^2 - 2(s(\sigma-1))(\alpha s)}{2}}\Phi\left(s(\alpha - (\sigma-1))\right),$$

and to calculate the second part of (23), I set $u = \phi_{ij}^*$ and $r = \sigma - 1$:

$$\int_0^{\phi_{ij}^*}\phi^{\sigma-1}f(\phi)d\phi = \frac{\rho}{\Phi(s\alpha)}\theta^{\sigma-1}e^{\frac{(s(\sigma-1))^2 - 2(s(\sigma-1))(\alpha s)}{2}}\Phi\left(s(\alpha - (\sigma-1)) + \frac{\ln\phi_{ij}^* - \ln\theta}{s}\right)$$

Putting two parts together gives the following:

$$\int_{\phi_{ij}^*}^\theta \phi^{\sigma-1}f(\phi)d\phi = \frac{\rho}{\Phi(s\alpha)}\theta^{\sigma-1}\epsilon\left[\Phi\left(s(\alpha - (\sigma-1))\right) - \Phi\left(s(\alpha - (\sigma-1)) + \frac{\ln\phi_{ij}^* - \ln\theta}{s}\right)\right]$$

where an auxiliary constant is defined as:

$$\epsilon = e^{\frac{(s(\sigma-1))^2 - 2(s(\sigma-1))(\alpha s)}{2}}.$$

Next, I calculate the remaining part of equation (23) – integral over the (θ, ∞) support:

$$\int_\theta^\infty \phi^{\sigma-1}f(\phi)d\phi = \int_\theta^\infty (1-\rho)\alpha\theta^\alpha\phi^{\sigma-1-\alpha-1}d\phi = \frac{(1-\rho)\alpha\theta^\alpha}{\alpha - (\sigma-1)}\theta^{\sigma-1},$$

where I make the assumption (supported by data) that $\alpha > \sigma - 1$. This is necessary for the integral to converge. Putting two parts together, I get the second selection statistics in case when $\phi_{ij}^* \le \theta$:

$$\int_{\phi_{ij}^*}^\infty \phi^{\sigma-1}f(\phi)d\phi = \frac{\epsilon\rho\theta^{\sigma-1}}{\Phi(\alpha s)}\left(\Phi\left(s[\alpha - \sigma + 1]\right) - \Phi\left(s[\alpha - \sigma + 1] + \frac{\ln\phi_{ij}^* - \ln\theta}{s}\right)\right) + \frac{\alpha(1-\rho)\theta^{\sigma-1}}{\alpha - (\sigma-1)}.$$

In the opposite case when $\phi_{ij}^* \ge \theta$ the left Log-normal-type tale becomes irrelevant and the expression becomes simply:

$$\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi = \frac{\alpha(1-\rho)\theta^\alpha}{\alpha-(\sigma-1)} (\phi_{ij}^*)^{\sigma-\alpha-1}$$

Selection statistics under Log-normal distribution

Under the Log-normal assumption with shape parameter s_l and scale parameter μ, the first selection statistics is straightforward and can be directly derived from the well-known c.d.f:

$$1 - F(\phi_{ij}^*) = 1 - \Phi\left(\frac{\ln \phi_{ij}^* - \mu}{s_l}\right).$$

The second selection statistics can be derived as follows:

$$\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi = \int_{\phi_{ij}^*}^{\infty} \frac{1}{\sqrt{2\pi}s_l} \phi^{\sigma-2} e^{-\frac{1}{2}\left(\frac{\ln \phi - \mu}{s_l}\right)^2} d\phi$$

First, let me perform change of variables as follows:

$$y = \frac{\ln \phi - \mu}{s_l} \quad \Rightarrow \quad \phi = \exp(ys_l + \mu) \quad \Rightarrow \quad d\phi = s_l \exp(ys_l + \mu) dy.$$

Plug this back into the integral expression to get:

$$
\begin{aligned}
\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi &= \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}s_l} e^{(ys_l+\mu)(\sigma-2)} e^{-\frac{1}{2}y^2} s_l e^{(ys_l+\mu)} dy = \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(ys_l+\mu)(\sigma-1)-\frac{1}{2}y^2} dy \\
&= \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ys_l(\sigma-1)+\mu(\sigma-1)-\frac{1}{2}y^2} dy = \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2-2ys_l(\sigma-1))+\mu(\sigma-1)} dy \\
&= \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-s_l(\sigma-1))^2+\frac{1}{2}[s_l(\sigma-1)]^2+\mu(\sigma-1)} dy \\
&= e^{\frac{1}{2}[s_l(\sigma-1)]^2+\mu(\sigma-1)} \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-s_l(\sigma-1))^2} dy.
\end{aligned}
$$

An additional change of variable is performed as follows:

$$z = y - s_l(\sigma-1) \quad \Rightarrow \quad y = z + s_l(\sigma-1) \quad \Rightarrow \quad dy = dz,$$

which allows reformulating the integral as:

$$
\begin{aligned}
\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi &= e^{\frac{1}{2}[s_l(\sigma-1)]^2+\mu(\sigma-1)} \int_{\frac{\ln \phi_{ij}^* - \mu}{s_l}-s_l(\sigma-1)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= e^{\frac{1}{2}[s_l(\sigma-1)]^2+\mu(\sigma-1)} \left(1 - \Phi\left(\frac{\ln \phi_{ij}^* - \mu}{s_l} - s_l(\sigma-1)\right)\right).
\end{aligned}
$$

Selection statistics under Unbounded Pareto

The selection statistics under unbounded Pareto are well-known. The c.d.f. gives the following:

$$1 - F(\phi_{ij}^*) = \left(\frac{\phi_m}{\phi_{ij}^*}\right)^{s_p}.$$

The second selection statistics is calculated as follows:

$$\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi = \int_{\phi_{ij}^*}^{\infty} s_p \phi_m^{s_p} \phi^{\sigma-1-s_p-1} d\phi = \frac{s_p}{s_p - (\sigma - 1)} \phi_m^{s_p} (\phi_{ij}^*)^{\sigma-1-s_p},$$

as long as $s_p > \sigma - 1$ and $\phi_{ij}^* \geq \phi_m$.

Selection statistics under Bounded Pareto

The first selection statistics under the bounded Pareto can be characterized as:

$$1 - F(\phi_{ij}^*) = 1 - \frac{1 - \phi_l^{s_b}(\phi_{ij}^*)^{-s_b}}{1 - \phi_l^{s_b}(\phi_h)^{-s_b}}.$$

The second selection statistics is derived as follows:

$$\begin{aligned}
\int_{\phi_{ij}^*}^{\infty} \phi^{\sigma-1} f(\phi) d\phi &= \frac{s_p \phi_l^{s_p}}{1 - \phi_l^{s_p}(\phi_h)^{-s_b}} \int_{\phi_{ij}^*}^{\infty} (\phi)^{\sigma-1-s_b-1} \\
&= \frac{s_p \phi_l^{s_p}}{1 - \phi_l^{s_p}(\phi_h)^{-s_b}} \frac{1}{\sigma - 1 - s_b} \max\left((\phi_h^{\sigma-1-s_b} - (\phi_{ij}^*)^{\sigma-1-s_b}, 0\right).
\end{aligned}$$