From Heavy-Tailed Micro to Macro: on the characterization of firm-level heterogeneity and its aggregation properties.

Ruben Dewitte*

16th September 2020

Abstract

This paper emphasizes the importance of two sufficient statistics to characterize firm-level heterogeneity in the workhorse heterogeneous firms trade model: the Cumulative Distribution Function (CDF) and the mean of firm-level sales. Contradicting the strong focus on the CDF, a close fit to average sales proves to be critical for model performance. Moreover, this average varies largely across finite sample draws due to sales being heavy-tailed, providing evidence that individual firms can influence the aggregate economy. As a result, modeled aggregate trade elasticities and Gains From Trade are unlikely to materialize: they are biased in finite samples and underlying characterizations of firm-level heterogeneity are rejected by the data.

Keywords: Average productivity, firm size distribution, heavy-tailed Distributions, granularity, gains from trade

JEL Codes: L11, F11, F12

*Dewitte: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be. I am grateful to Luca David Opromolla, Gonzague Vannoorenberghe, Olivier Thas, Kris Boudt, Jasson Vindas Diaz, Gerdie Everaert, Glenn Rayp, Angelos Theodorakopoulos, Bruno Merlevede and Marijn Verschelde for their helpful comments and suggestions.

1 Introduction

With the seminal work of Melitz (2003), the trade literature entered an era of trade models stressing the importance of firm-level heterogeneity in shaping international trade relations denoted "New New Trade Theory". The ability of these "new new trade models" to capture detailed firm-level trade reality brings them much closer to businesses and policymakers (Cernat, 2014).[1]

Despite firm-level heterogeneity being the core of these models, there is not yet a consensus on its characterization. In effect, a critical part of the trade literature relies on either of two qualitatively very different distributions: the Pareto and the Lognormal distribution. The Pareto distribution is deemed able to approximate the heavy tail[2] of the heterogeneity distribution (Nigai, 2017) and the Melitz-Pareto model, with parsimonious modification, has been argued to match empirical patterns such as entry and the distribution of sales across markets, and the connection between exports and sales at home (Eaton et al., 2011). The Lognormal distribution, on the other hand, provides a better approximation of the bulk of the productivity distribution (Head et al., 2014; Nigai, 2017) and is argued to match, in contrast to the Pareto distribution, empirical evidence of varying trade elasticities across country pairs (Bas et al., 2017). As two qualitatively such different distributions can both be argued to match firm-level trade reality based on a similar trade model, it appears necessary to evaluate which features of firm-level heterogeneity need to be captured according to this model and how these features can be evaluated in the data.

This paper builds on the idea that two statistics are sufficient to characterize firm-level heterogeneity in the workhorse heterogeneous firms trade model: the Cumulative Distribution Function (CDF) and the mean of the *firm-level sales distribution* evaluated at the respective domestic/exporting cutoffs. Whereas current work tries to obtain a close fit to the CDF (and/or Probability Density Function (PDF)) of the sales distribution, a close fit to its bounded mean proves to be critical. Using domestic sales of French firms in 2012, I demonstrate that sales are heavy-tailed and that, as a result, bounded average sales varies largely across finite sample draws. Due to its vital role, the instability of this bounded average begets important consequences for the evaluation of the

[1]See Bernard et al. (2012) for an overview of the empirical evidence on firm heterogeneity and trade.

[2]This paper refers to heavy-tailed distributions as distributions with the property that some of their moments are infinite. Several related definitions of heavy tails exist. It can be referred to as a distribution of which the tail probabilities decay more slowly than those of any exponential distribution. Also, a distribution can be defined as heavy-tailed if observations are i.i.d. according to a distribution in the domain of attraction of a α-stable law with index $\alpha < 2$.

workhorse heterogeneous firms trade model. First, the validity of the continuum of firms assumption in the model appears to depend on the distributional form chosen to characterize firm-level heterogeneity. Direct comparisons between model outcomes differentiated by distribution choice (for instance Lognormal versus Pareto) are therefore not recommended. Second, the continuum of firms assumption does not hold in the data. Third, aggregate trade elasticities are subject to both a finite sample and a truncation bias. A realistic quantification of the Melitz-Pareto model (Chaney, 2008), for instance, features varying rather than constant trade elasticities. Fourth and last, none of the currently considered distributions in the literature provide a significantly good fit to the data. Gains From Trade[3] (GFT) calculated under currently popular distributional assumptions are unlikely to materialize in the real world.

The workhorse heterogeneous firms trade model features Constant Elasticity of Substitution (CES) demand and a continuum of firms heterogeneous in terms of their productivity. It has been shown that two statistics are sufficient to characterize firm-level heterogeneity in these models: the CDF and the (σ-1)th moment of the productivity distribution evaluated at the cutoffs, with σ the elasticity of substitution between varieties (Melitz and Redding, 2014; Nigai, 2017). In line with Mrázová et al. (2015), this paper shows that if the productivity distribution is closed under power-law transformations, these statistics are equivalent to the CDF and bounded average of the more generally available firm-level sales distribution. Whereas the CDF is commonly used to differentiate between the performance of parametric distributions (see for instance Head et al. (2014); Nigai (2017)), bounded average firm-level sales is revealed to be a new aspect of the distribution that requires attention. This is certainly so considering that bounded average firm-level sales carries relatively more weight in the model's outcome compared to the CDF.

The second contribution of this paper consists of a discussion on the accurate (non-) parametric characterization of these sufficient statistics and their evaluation in heavy-tailed data. Evidence that the firm-level sales distribution is heavy-tailed, exhibiting infinite variance, is omnipresent (see for instance Gabaix and Ibragimov (2011); Bee and Schiavo (2018); Head et al. (2014); Nigai (2017)). As a result, a nonparametric estimator will likely be biased in finite[4] samples: nonparametric

[3]Gains From Trade are defined as the changes in welfare, measured as real income, from a change in variable trade costs.

[4]The term 'finite' is rather demeaning. Not even the largest economy would be considered large enough under the common assumptions of sales being Pareto distributed with a shape parameter above but close to one (Gabaix, 2009, 2011).

estimates of the sample mean will be influenced by rare but extremely large data values in the heavy tail (Romano and Wolf (1999); Hall and Yao (2003); Athreya et al. (1987), ...), resulting in a *finite sample bias.* This influence increases as this mean is bounded from below, resulting in a *truncation bias.* A parametric measurement, on the other hand, operates under the infinite number of firms assumption. Due to the heavy-tailed nature of firm-level sales, these parametric 'population' estimates are neither representative of, nor comparable with the nonparametric measures based on 'finite' data. This paper therefore proposes to rely on a parametric bootstrap to correctly evaluate the sufficient statistics obtained from parametrically implied finite samples relative to nonparametric measures from the data.

The applicability of this proposed bootstrap methodology is tested on domestic sales of 500,388 French firms in 2012. After confirming that firm-level sales indeed is heavy tailed, evidence is provided that none of the currently considered distributions in the literature provide a good fit to the data. The proposed evaluation method rejects the premise that the sufficient statistics observed in the data originate from any distribution currently considered. Moreover, parametric estimates of the first distribution moment (sample mean) are biased, with the size of the bias depending on the heavy-tailedness of the underlying distribution. Finite sample estimates obtained from differently-tailed distributions are therefore incomparable.

The conclusions drawn from the empirical exercise spill over to aggregate model outcomes. Whether or not the equilibrium of the CES heterogeneous firms trade model is stable across different finite sample draws depends on the imposed distributional form. For instance, whereas this is generally the case for Lognormal distributions, it is not for Pareto (right-tailed) distributions. This indicates that comparing finite sample modeled outcomes based on different distributional assumptions is not recommended. It also indicates that the continuum of firms assumption in the CES heterogeneous firms trade models (implying a stable equilibrium across finite sample draws) depends on the heavy-tailedness of the imposed distribution. Provided the infinite variance of firm-level sales, it is shown that the equilibrium of the workhorse heterogeneous firms trade model will be unstable across finite sample draws. Thus, the continuum of firms assumption predominant in the CES heterogeneous firms trade models does not hold up in the data

The impact of the presented results is evaluated on two popular aggregate statistics, the aggregate trade elasticity and Gains From Trade (GFT)[5], are evaluated. First, the trade elasticity is a typical variable that needs to be deduced from small data samples, as the number of firms that establish trade between a country pair can be small. These elasticities will therefore, as discussed above, be influenced by sales values from large firms residing in the heavy tail of the sales distribution and be (finite-sample) biased. The influence of these large firms will increase for costly destinations to which small(er) firms do not export, resulting in a truncation bias. A realistic quantification of the Melitz-Pareto model (Chaney, 2008), for instance, features varying rather than constant trade elasticities across country pairs due to this truncation bias. Second, the combination of distributions fits that are not supported by the data and the existence of finite sample biases lead us to conclude that GFT calculated under popular distributional assumptions are unlikely to materialize in the real world, where only a finite number of firms are present.

This work is closely related to the trade literature that studies the importance of the characterization of firm-level heterogeneity. In the trade literature, productivity is commonly assumed to be Pareto distributed, as Pareto is claimed to provide a good fit at least to the right tail of the productivity distribution and allows for straightforward GFT calculations (Arkolakis et al., 2012; Costinot and Rodríguez-Clare, 2014). The popularity of Pareto has been challenged by the Lognormal distribution (Head et al., 2014; Fernandes et al., 2018), the Weibull (Bee and Schiavo, 2018), Lognormal-Pareto (Nigai, 2017) and Double-Pareto Lognormal (Sager and Timoshenko, 2019) distribution, each of which results in different GFT. Current evaluations of distribution fits mostly focus on obtaining a high coefficient of determination (R^2). As this test statistic does not give an indication of the validity of the distributional assumption for the sufficient statistics separately, a test statistic based on the Kolmogorov-Snirnov test is proposed in this paper.

The contributions of this paper can also be linked to the literature that studies the variability of trade elasticities across country pairs depending on the distributional assumption. The constancy of the trade elasticity across countries under a Pareto distribution is often relied upon as a feature for Gains From Trade calculations (Costinot and Rodríguez-Clare, 2014). This feature of the Pareto distribution has been criticized by among others (Melitz and Redding, 2015; Bas et al., 2017). (Bas et al., 2017, p. 2) call upon the "direct evidence that aggregate elasticities are non-constant across

[5]Gains From trade are defined as the changes in real per-capita income due to an exogenous change in variable trade costs.

country pairs" as evidence in favor of the Lognormal relative to the Pareto distribution. The results in this paper uncover that the Pareto-implied trade elasticity can vary across country pairs when evaluated in finite samples, though not as strongly as the Lognormal distribution.

This paper is also related to the granularity literature (Eaton et al., 2012; Gabaix, 2011; di Giovanni and Levchenko, 2012; Carvalho and Grassi, 2019). The granularity literature mostly relies on the Pareto distribution to motivate and/or explain the transmission of firm-level to aggregate volatility (Gabaix, 2011; di Giovanni and Levchenko, 2012; Carvalho and Grassi, 2019). This paper emphasizes, in line with Eaton et al. (2011), the existence of a finite sample bias for average firm-level sales: when sales are heavy tailed, firms do not reduce to a point on a continuum and carry the potential to affect aggregate statistics. The presented results challenge, however, the premise that this granularity originates from a Pareto-tailed characterization of firm-level heterogeneity.

This rest of the paper is structured as follows. I present a CES heterogeneous firms trade model in Section 2 that accommodates the assumption of both a finite and an infinite number of firms. The aim of the model is to single out the relative importance of the sufficient statistics that determine the model with respect to firm heterogeneity. In Section 3, I discuss the empirical applicability of characterizing firm heterogeneity based on these statistics. I suit the action to the word in Section 5, evaluating the fit of popular distributions to the domestic sales of 500,388 French firms in 2012 (the data is discussed in section 4). Section 6 evaluates the implications of the findings for the equilibrium of the presented CES heterogeneous firms trade model and for two resulting aggregate trade statistics: aggregate trade elasticity and Gains From Trade, before concluding in Section 7.

2 CES heterogeneous firms trade model[6]

This section presents a heterogeneous firms model with a symmetric open economy. The model features Constant Elasticity of Substitution (CES)-demand and monopolistic competition between a finite number of firms who ignore their aggregate impact (Dixit and Stiglitz, 1977; Krugman, 1980; di Giovanni and Levchenko, 2012), while remaining agnostic on the parametric specification of firm-level heterogeneity. As the number of firms increases to infinity, the model becomes comparable to the workhorse heterogeneous firms trade model (Melitz, 2003). Accommodating both

[6]See online Appendix B for a complete workout of the model.

the finite and infinite number of firms assumption allows for the model to be evaluated in finite data while maintaining comparability with the current practice in the literature that assumes an infinite number of firms.

Imagine a world with I symmetric countries, each country $i \in I$ populated by L^i identical households. Each household supplies inelastically one unit of labor, earning wage W^i. They make their consumptions choices over a over a finite number of horizontally differentiated varieties ($\varpi \in \Omega^i$) originating from country $i \in I$ and are assumed to take the Constant Elasticity of Substitution (CES) utility (U) form

$$U^j = \left(\sum_{i=1}^{I} \sum_{\varpi \in \Omega^i} q^{ij} \left(\varpi\right)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

with σ the elasticity of substitution between varieties. Utility maximization defines the optimal consumption and expenditure decisions over the individual varieties

$$\frac{q^{ij}(\varpi)}{Q^j} = \left[\frac{p^{ij}(\varpi)}{P^j} \right]^{-\sigma}, \tag{2}$$

where the set of varieties consumed is considered as an aggregate good $Q \equiv U$ and P is the CES aggregate price index.

There is a finite number of businesses ($b \in B$) which choose to supply a distinct horizontally-differentiated variety. They are heterogeneous in terms of their productivity $\omega_b \in [0, \infty]$ drawn from the unconditional Cumulative Distribution Function (CDF) $G(\omega_b)$ after paying a fixed cost f^{ie} in terms of production factor L^i to enter the market.[7] There is zero probability of firm death.[8] Supply of the production factor to the individual firm is perfectly elastic, so that firms are effectively price (W^i) takers on the input markets. Once active, firms from country i have to pay a fixed cost f^{ij} to produce goods destined for country j. The cost function of the firm involves a fixed production cost, iceberg trade costs $\tau^{ij} > 1$ and a constant marginal costs that depends on its productivity: $f^{ij} + \left(\frac{\tau^{ij} q^{ij}}{\omega}\right) W^i$.

[7] As ω_b is the sole heterogeneity component identifying individual firms, the subscript b is dropped in further derivations.

[8] The static specification in which there is zero probability of firm death follows most of the international trade literature.

Profit maximization of the firm results in a price as a constant markup over marginal costs $p^{ij} = \frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^i}{\omega}$. The revenue of a firm from i in j with productivity $\omega \sim G(\omega)$ can then be specified as:

$$x^{ij} = \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^i}{\omega}\right)^{1-\sigma} Q^j \left(P^j\right)^{\sigma} = C^{ij}\omega^{\sigma-1}, \tag{3}$$

with P^j the CES price index in country j, and C^{ij} is a constant absorbing all variables that do not vary at the firm-level.

Only firms that are productive enough can survive $\omega \geq \omega^{ii*}$ and/or export $\omega \geq \omega^{ij*}$ with the cutoffs defined by zero-profit conditions equating the operating profit in each market to the relevant fixed costs:

$$\sigma f^{ij}W^i = \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^i}{\omega^{ij*}}\right)^{1-\sigma} Q^j \left(P^j\right)^{\sigma}. \tag{4}$$

Combining the zero cutoff profit conditions allows us to write the export cutoff as a function of a foreign domestic productivity cutoff, variable and fixed costs and the wages:[9]

$$\omega^{ij*} = \left(\frac{W^i}{W^j}\right)^{\frac{\sigma}{\sigma-1}} \left(\frac{f^{ij}}{f^{jj}}\right)^{\frac{1}{\sigma-1}} \left(\frac{\tau^{ij}}{\tau^{jj}}\right)\omega^{jj*}. \tag{5}$$

The equilibrium value of these cutoffs are uniquely determined by the free entry condition, requiring the probability of successful entry times the expected future value of entry conditional upon successful entry to equal the sunk entry cost:

[9]Here, the model is focused on parameter values such that there is, in line with empirical evidence, selection into exporting ($\omega^{ij*} > \omega^{ii*}$). See online Appendix B for the implications of this assumption.

$$\sum_{j=1}^{J} \mathbb{E}\left[\pi^{ij} | \omega > \omega^{ij*}\right] = f^{ie} W^i$$

$$\sum_{j=1}^{J} f^{ij} \left[\left(\omega^{ij*}\right)^{1-\sigma} \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1} - \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{0}\right] = f^{ie}$$

$$\sum_{j=1}^{J} f^{ij} \left[\left(\omega^{ij*}\right)^{1-\sigma} m^{\sigma-1}_{\omega^{ij*}} - m^{0}_{\omega^{ij*}}\right] = f^{ie}, \tag{6}$$

where I denote by m^r_y the nonparametric y-bounded, r-th sample moment of the productivity distribution:

$$m^r_y = \frac{1}{B} \sum_{n=1}^{B} \mathbb{I}_{\omega_b \geq y} \omega^r_b, \qquad \text{for } r \in \mathbb{R}^+. \tag{7}$$

If one would assume an infinite rather than finite number of firms, one can rely on the law of large numbers to replace the specified sample moments with their population equivalent μ^r_y, the nonparametric y-bounded, r-th population moment of the productivity distribution:

$$\mu^r_y = \int_y^{\infty} \omega^r g(\omega) d\omega, \qquad \text{for } r \in \mathbb{R}^+, \tag{8}$$

where $g(\omega)$ represents the Probability Density Function (PDF) related to $G(\omega)$. Under the continuum of firms assumption, the free-entry equation 6 is equivalent to the continuous free-entry equation as specified by Melitz (2003).

Using the relation between productivity cutoffs (eq. 5), the free entry condition (eq. 6) determines the unique equilibrium values of these cutoffs. The role of the productivity distribution in this free entry equation establishes that a characterization of firm heterogeneity by means of the productivity distribution in the Melitz (2003)-model relies solely on the bounded 0th and $(\sigma-1)$th moment of the productivity distribution, evaluated at the respective productivity/exporting cutoffs (Melitz and Redding, 2015; Nigai, 2017). The importance of the $(\sigma-1)$th moment in the theoretical model cannot be overstated. Its relative weight in the free entry equation (eq. 6) is larger than that of the CDF:

$$\left(\frac{\omega}{\omega^{ij*}}\right)^{\sigma-1} \geq \omega^0 = 1, \quad \text{such that} \quad \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\left(\omega > \omega^{ij*}\right)\left(\frac{\omega}{\omega^{ij*}}\right)^{\sigma-1} > \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\left(\omega > \omega^{ij*}\right)\omega^0. \quad (9)$$

Moreover, conditional upon the probability of survival in the domestic/export market, this $(\sigma - 1)$th moment "also represents aggregate productivity because it completely summarizes the information in the distribution of productivity levels relevant for all aggregate variables" (Melitz, 2003, p. 1700).

3 Measuring firm-level heterogeneity

As discussed in the previous section, two statistics suffice to characterize firm-level heterogeneity in the presented theoretical model: the 0th and $(\sigma - 1)$th moment of the productivity distribution. This section discusses appropriate methods to capture and evaluate this heterogeneity in firm-level data.

I start by linking the productivity statistics to firm-level sales statistics, as sales are more generally available. I then discuss the estimation of the sufficient statistics necessary to measure firm-level heterogeneity. This discussion differentiates both between parametric and nonparametric estimation methods as well as between the behavior of these estimators in finite and infinite sample sizes (see Table 1 for an overview).

Table 1: Overview sufficient statistic estimators in finite and infinite samples

	Finite[a] nr. of firms	**Infinite nr. of firms**
Nonparametric	$m_{\tilde{y}}^{\tilde{r}}(x) = \frac{1}{B}\sum_{n=1}^{B}\mathbb{I}_{x_b \geq \tilde{y}} x_b^r$	$\mu_{\tilde{y}}^{\tilde{r}}(x) = \int_{\tilde{y}}^{\infty} x^r g(x) dx$
Parametric[b]	$m_{\tilde{y}}^{\tilde{r}}(x\|\boldsymbol{\theta}) = \frac{1}{B}\sum_{n=1}^{B}\mathbb{I}_{x_b \geq \tilde{y}} x_b^r$	$\mu_{\tilde{y}}^{\tilde{r}}(x\|\boldsymbol{\theta}) = \int_{\tilde{y}}^{\infty} x^r g(x\|\boldsymbol{\theta}) dx$

Notes: $r = 0, 1, y \in \mathbb{R}$. [a]Finite sample refers to a sample with a fixed number of observations $B \in \mathbb{Z}$. [b]Parametric estimators assume observations originate from a parametric specification: $x_b \sim g(x|\boldsymbol{\theta})$.

3.1 Sales as proxy for productivity

Relying on productivity is burdensome for a researcher, as it is an unknown that needs to be derived from available data. Data requirements are therefore hard. The distributional relation between sales and productivity in the CES heterogeneous firms model, according to equation 3, provides a solution. If productivity follows a distribution that is closed under power-law transformations, sales will follow the same distribution up to a change of distributional parameters $x^{ij}(\omega) \sim C^{ij}\omega^{\sigma-1}$ (see also Mrázová et al. (2015)).[10] To avoid firm-level heterogeneity interference due to trade costs, it is advised to rely on domestic sales ($x^{ii}(\omega)$) as a proxy for productivity (di Giovanni et al., 2011). The distributional moments of productivity can then easily be approximated by sales up to a constant. By a change of variables $x^{ii} = C^{ii}\omega^r, dx^{ii} = C^{ii}\omega^{r-1}d\omega$, we can rewrite the sufficient statistics:[11]

$$m_y^r = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(x^{ii} > \tilde{y}\right) \frac{(x^{ii})^{\tilde{r}}}{C^{ii}} = m_{\tilde{y}}^{\tilde{r}}, \qquad \text{for } \tilde{r} = 0, 1. \tag{10}$$

Thus, we can sufficiently characterize firm heterogeneity in a CES firm heterogeneous model relying only on the sales' CDF and bounded average. Besides the accessibility of firm-level sales compared to productivity, relying on sales has the advantage of absorbing the elasticity of substitution (σ).

3.2 Parametric measurement

Current practices in the literature mainly focus on parametric population estimators to obtain sufficient statistic estimates (see for instance (Bas et al., 2017; Melitz and Redding, 2015; Bee and Schiavo, 2018; Nigai, 2017)). If the distributional parameters are known, these parameters can be plugged into the analytical expression for the sufficient statistics (bottom right cell of Table 1) that are available for each parametric distribution. Whether or not these 'plug-in' estimates are representative of the 'truth' $\left(\mu_{\tilde{y}}^{\tilde{r}}(x|\boldsymbol{\theta}) \stackrel{?}{=} \mu_{\tilde{y}}^{\tilde{r}}(x)\right)$ however, is conditional on the validity of the distributional assumption.

Current evaluations of this validity mostly focus on obtaining a high coefficient of determination

[10] Most common distributions used in the economic literature are closed under power-law transformations (see online Appendix Table A.2 and Dewitte et al. (2019)).

[11] For $Y = aX^b$ and $G(\cdot)$ closed under power-law transformations, we have $G_Y(y) = G_X(ay^b)$ and $g_Y(y) = aby^{b-1}g_X(ay^b)$.

(R^2). Besides being a low-powered distribution test (Clauset et al., 2009), this test statistic does not give an indication of the validity of the distributional assumption for the separate moment conditions. This section therefore proposes to evaluate the validity of the distributional assumption relying on comparison between the nonparametric sample estimates (top left cell of Table 1) and parametrically bootstrapped sample estimates (bottom left cell of Table 1). I first discuss the distribution choice and fitting procedure before expanding on the proposed evaluation methodology.

3.2.1 Distribution choice

As stated in the introduction, the two distributions usually considered to characterize firm-level heterogeneity are the Lognormal (Fernandes et al., 2018; Head et al., 2014) and Pareto distribution (see Arkolakis et al. (2012) for an overview of work relying on the Melitz-Pareto combination). However, the Pareto distribution is mostly considered to be a good fit only to the tail of the efficiency distribution, while the Lognormal distribution provides a better approximation to the bulk of the distribution. Several authors therefore proposed combinations the Lognormal and Pareto distribution. Sager and Timoshenko (2019) introduced the Double-Pareto Lognormal distribution.[12] Reducing the parameter space of the Double Pareto allows us to consider the Left- and Right-Pareto Lognormal distribution respectively (Reed and Jorgensen, 2004). Luckstead and Devadoss (2017) introduced the piecewise composite Inverse Pareto-Lognormal-Pareto distribution. Also this distribution can be reduced to an Inverse Pareto-Lognormal[13] or Lognormal-Pareto distribution only (Ioannides and Skouras, 2013; Nigai, 2017).[14] This paper considers all these possible combinations of the Lognormal and Pareto distribution. In line with the trade literature, I also consider a Pareto distribution fitted to only to the largest 5% and 1% values in the data (Bee and Schiavo, 2018; di Giovanni et al., 2011; Bas et al., 2017) as well as an optimal tail length determined by the minimum of the Kolmogorov-Smirnov Distance between Pareto and the data calculated for every cutoff value in the data (Clauset et al., 2009).

[12] Sager and Timoshenko (2019) exponentiate their productivity term and refer to their distribution as the Double Exponentially Modified Gaussian (EMG) distribution.

[13] This form of the piecewise composite distribution has not been applied yet in the literature. The piecewise composite in is therefore generalized in online Appendix Section C to consist of any two- or three truncated distributions.

[14] See online Appendix Tables 1 and 2 for the specifications of all distributions considered.

3.2.2 Distribution fit

Distribution parameters can be estimated using maximum likelihood, case in which these estimates are asymptotically consistent and efficient. A plug-in estimate of parametric population moments can be obtained given these parameter estimates, and its asymptotic properties can be calculated for every parametric model using the delta method. If the parametric assumption is correct, these plug-in estimates will also be asymptotically consistent and efficient.

3.2.3 Distribution evaluation

I propose to evaluate the parametric distributions summarizing the distance between the nonparametric and parametric rth sample moment of the distribution by the normalized $1-$ and $\infty-$ norm:[15]

$$S^{\tilde{r}} = \sum_{\tilde{y}} \Delta_{\tilde{y}}^{\tilde{r}}, \qquad T^{\tilde{r}} = \sup_{\tilde{y}} \Delta_{\tilde{y}}^{\tilde{r}}, \tag{11}$$

where $\Delta_{\tilde{y}}^{\tilde{r}}$ defines the Normalized Absolute Deviation (NAD) between nonparametric sample and parametric population moment:

$$\Delta_{\tilde{y}}^{\tilde{r}} = \frac{\left| \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}_{x_b \geq \tilde{y}} x_b^{\tilde{r}} - \int_{\tilde{y}}^{\infty} x^{\tilde{r}} g(x|\boldsymbol{\theta}) dx \right|}{\frac{1}{B} \sum_{b=1}^{B} x_b^{\tilde{r}}}. \tag{12}$$

The maximum NAD between the nonparametric sample and parametric population 0th-moment of the distribution, T^0, corresponds with the well-known Kolmogorov-Smirnov (KS) test statistic. This is the sole test statistic specified on which we can rely to provide statistically underpinned claims regarding the accuracy of the distributional assumption with respect to its empirical counterpart.[16] Whereas the ∞-norm contains only information on the largest distance, the 1-norm provides information on the distance between both distributions over the complete distributional space, weighting all distances equally. The normalization factor allows us to interpret the distances on a scale of zero to one for all moments, similar to the interpretation of the standard KS test statistic. As we rely on estimated parameters, asymptotic distributions are not available for the

[15] See online Appendix Table 3 for the analytical expression of these parametric population moments.

[16] I have no knowledge of statistical tests that evaluate distributional fits based on the bounded first moments of the distribution.

test statistics. We therefore rely on the parametric bootstrap:

1. Assume B i.i.d. random variables with distribution $G(\cdot|\boldsymbol{\theta})$ and with nonparametric sample moments $m_{\tilde{y}}^{\tilde{r}}(x)$ for $\tilde{r} = 0, 1$, as specified in the top left cell of table 1;

2. Estimate the parameters $\boldsymbol{\theta}$ of the distribution using MLE and calculate the parametric (plug-in) population moments $\hat{\mu}_{\tilde{y}}^{\tilde{r}}(x|\hat{\boldsymbol{\theta}})$ for $\tilde{r} = 0, 1$, as specified the bottom right cell of table 1;

3. $H_0 : \mu_0^{\tilde{r}}(x) = \hat{\mu}_0^{\tilde{r}}(x|\hat{\boldsymbol{\theta}})$ with test statistic $t \in \left\{S^{\tilde{r}}, T^{\tilde{r}}\right\}$;

4. Draw N bootstrap samples of size B from $G(\cdot|\hat{\boldsymbol{\theta}})$;

5. For each sample of the parametric distribution, calculate the bootstrapped parametric sample moments $m_{\tilde{y}}^{\tilde{r}}(x^*|\hat{\boldsymbol{\theta}})$ and calculate the test statistics $t^* \in \left\{(S^{\tilde{r}})^*, (T^{\tilde{r}})^*\right\}$;[17]

6. The p-value is then defined as

$$\hat{p} = \frac{1}{N+1}\left[\sum_{n=1}^{N} \mathbb{I}\left(t_n^* \geq t\right) + 1\right].$$

The bootstrap exercise should therefore be interpreted as 'the likelihood of observing a deviation between the nonparametric and parametric sample moments as large as t under the null hypothesis'. This should allow us to evaluate whether the distributional assumption provides a good fit to the data overall and whether it provides a good approximation of specific distribution moments.

3.3 Nonparametric measurement

As the nonparametric sample estimates serve as a benchmark for the parametric ones in the previous section, one could wonder why one should not discard these parametric methods in favor of the nonparametric method. After all, the nonparametric sample estimators of the two sufficient statistics, the sales' CDF and bounded average (defined in the left top cell of Table 1) correspond to the empirical CDF and bounded average of firm-level sales and can therefore easily be retrieved

[17]Note that I do not re-fit the parametric distribution to the bootstrap sample. The vastness of the dataset at my availability in the empirical section results both in a large computational burden but also a very precise estimation of the distribution parameters. The influence of not refitting the parametric distribution to the bootstrap sample is therefore negligent.

from the data. The complication of relying on nonparametric estimates, as it appears, lies in the asymptotic behavior of these estimates.[18]

Starting with the asymptotic behavior of the nonparametric sample estimator of the bounded zeroth moment (the empirical CDF), we can apply the Central Limit Theorem (CLT). This CLT states that empirical CDF estimates $\left(\hat{G}_b(\tilde{y})\right)$ converge pointwise at the standard $b^{\frac{1}{2}}$ rate to a normal distribution: $\sqrt{b}\left(\hat{G}_b(\tilde{y}) - G(\tilde{y})\right) \rightarrow^d \mathcal{N}\left(0, G(\tilde{y})\left(1 - G(\tilde{y})\right)\right)$ (Van der Vaart, 1998, p.266). Finite-sample estimates of the bounded zeroth moment are therefore consistent estimators, even for relatively small sample sizes.

As for the asymptotic behavior of the nonparametric sample estimate of the bounded first moment, bounded average of firm-level sales, things are more complicated. This is because firm-level sales is heavy-tailed, exhibiting infinite variance.[19] As a result, (i) the standard CLT does not apply and (ii) there are insufficient firms in any economy to invoke the law of large numbers. Intuitively, the infinite variance of firm-level sales implies that the sample average will be heavily influenced by rare but extremely large data values. This results in a *finite sample bias* even though the estimator is consistent. If there exists an asymptotic distribution that is non-degenerate, this limiting distribution will therefore be leptokurtic and heavy tailed rather than normally distributed. Moreover, as sample averages are bounded from below, these rare but extremely large data values become even more influential, resulting in a *truncation bias*.

First, in the case of a random variable with infinite variance, the Generalized Central Limit Theorem (GCLT) rather than the standard CLT applies. The GCLT states that a sum of independent random variables from the same distribution, when properly centered and scaled, belong to the domain of attraction of a stable distribution. Furthermore, the only distribution that arises

[18] Alternatively, one could rely on methods that try to nonparametrically estimate the population rather than the sample moments. One example of such approach is Nigai (2017) who relies on trapezoidal numerical integration of the empirical histogram to calculate population moments. In the case of heavy-tailed data, however, such nonparametric procedures are rendered less efficient (Clauset et al., 2009).

[19] Evidence on the non-existence of a finite second moment for firm-level sales is omnipresent. A common method to spot infinite higher moments consists of fitting a Pareto distribution to the tails of the data. The shape parameter of this Pareto distribution then represents the cutoff value below which moments become infinite (Pareto moments above the value of its shape parameter are infinite.). It is a well-known feature of firm-level sales data to approach Zipf's law (a Pareto distribution with shape parameter equal to one), at least in the right tail (see Gabaix (2009) for an overview). Recent evidence in the firm size literature is provided by among others (Head et al., 2014; Nigai, 2017; Bee and Schiavo, 2018), who fit the Pareto distribution to various quantiles of firm-level sales distributions. They consistently report a Pareto shape parameter below two, signaling infinite variance of the sales distribution. I have no knowledge of more robust methods to test for the existence of finite moments in the data (Fedotenkov, 2013, 2014; Trapani, 2016) being applied to firm-level sales data.

as the limit from a suitably scaled and centered sum of random variables is a stable distribution. Thus, the asymptotic distribution of bounded average sales belongs to the family of ($\alpha-$ or Levy-) stable distributions, with α referring to the tail index of such a distribution (Durrett, 2010, p.161). If $\alpha = 2$, this stable distribution is the Normal distribution. When $\alpha < 2$, however, the shape of these stable distributions is leptokurtic and heavy-tailed. In contrast with the standard CLT, therefore, the GLCT does not easily apply in practice. The asymptotic distributions are particularly difficult to estimate (Hall and Yao, 2003; Cornea-Madeira and Davidson, 2015; Romano and Wolf, 1999; Davidson, 2012) and bootstrap solutions to determine the asymptotic distribution prove to be precarious (Romano and Wolf, 1999; Hall and Yao, 2003; Cavaliere et al., 2013; Athreya et al., 1987; Knight, 1989).

Second, the nonparametric sample moments are consistent estimates but the convergence rate under the GCLT is too slow to invoke the law of large numbers (in this setting). The rate of convergence under the GCLT is dependent on the underlying heterogeneity distribution. Gabaix (2011), for instance, states that in the case of a symmetrical distribution with zero mean and power-law distributed tails, the rate of convergence equals $b^{\frac{1}{\alpha}}$. In the case of firm-level sales, with a tail parameter α believed to be close to one, this speed of convergence becomes extremely slow (Gabaix, 2011) relative to when $\alpha \leq 2$. When $\alpha \leq 2$ variance is finite and the limiting distribution corresponds to the Normal distribution with standard speed of convergence. Taleb (2019) provides metrics to evaluate the speed of convergence for several distributions in the infinite variance case relative to the Gaussian. He provides the counterintuitive example of a Pareto distribution with a tail index $\alpha \approx 1.14$ requiring $> 10^9$ more observations than the Gaussian to reach an equivalent stability around the sample mean.

Overall, it is hard to provide statements on the efficiency of the nonparametric estimator while consistency is not attained. While a nonparametric sample estimator can establish a useful benchmark, it seems necessary to look beyond nonparametric methods when measuring firm-level heterogeneity.

4 Data

The empirical analysis relies on firm-level sales from the Orbis Europe database. All, both total and exporting, sales data from firms that reported sales in 2012 (excluding companies with no recent financial data and Public authorities/States/Governments) by means of unconsolidated accounts and that have a designated 'corporate' entity type have been downloaded.[20] This search strategy returned a total of 4,557,782 firms. Cleaning the data for duplicates, missing sales values, focusing on positive sales and on countries with at least 100,000 observations, we are left with 3,415,411 firms. Sales are normalized per country by dividing all observations with their respective mean value. The density of the sales distribution for all 12 countries in the dataset is depicted in online Appendix Figure 1. It can be observed that the quality of the data in the lower tail for some countries (mainly Bulgaria and Finland) is not ideal for the empirical exercise in this paper.

The main analysis in our paper focuses solely on the domestic sales distribution of France in 2012. Relying on domestic rater than total sales avoids the potential interference of firm-level heterogeneity in international trade with the firm size distribution di Giovanni and Levchenko (2012). France is one of the only countries with firms consistently reporting firm-level exports in Orbis Europe. Moreover, focusing on France allows straightforward comparison with papers concentrating on the French firm-level sales distribution (Head et al., 2014; Bas et al., 2017; Nigai, 2017; Bee and Schiavo, 2018). The domestic sales density of the 500,388 French firms that report non-missing, positive domestic sales in 2012 is displayed in Figure 1.

5 Firm-level evidence

The empirical analysis commences by providing evidence that, also in this dataset, firm-level sales is heavy tailed. I then demonstrate that none of the currently considered distributions in the literature provide a good fit to the data. The proposed evaluation methods reject the premise that the sufficient statistics observed in the data originate from any distribution considered. Moreover, if parametric estimates of the first moment of the distribution are biased, the size of the bias depends on the heavy-tailedness of the assumed distribution. Finite sample estimates obtained

[20]The data was obtained from the Orbis Europe database update 05/06/2020 (nr. 197002) and software version 197.

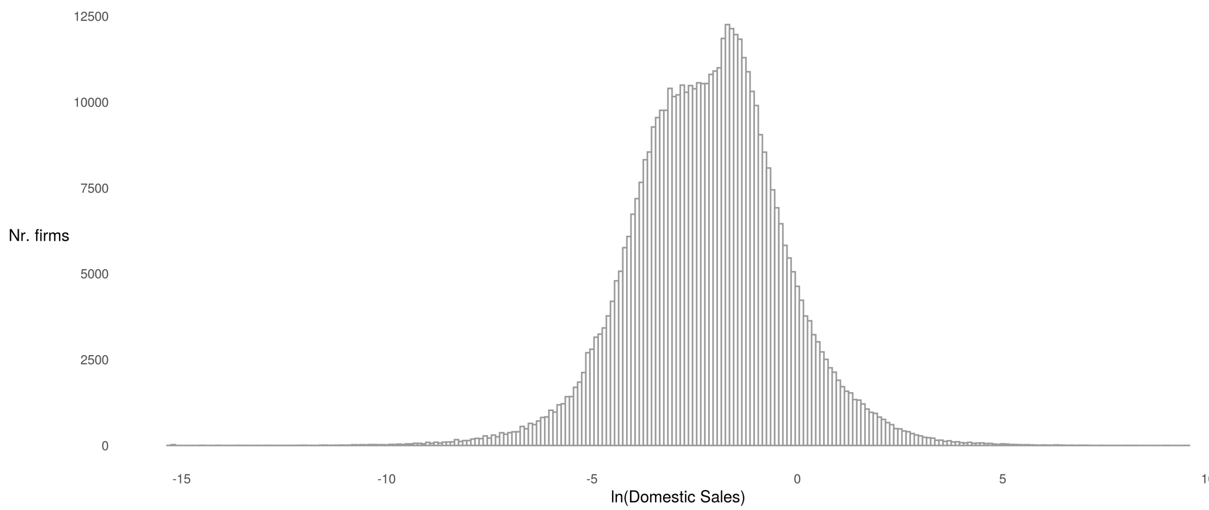


Figure 1: Domestic sales density of 500,388 French firms in 2012
Note: All observations are normalized, dividing the observations with their mean value.

under differently-tailed distributions are therefore incomparable.

5.1 Heavy-tailed sales

Before I employ the proposed methodology, I want to ensure that, also in our data set, firm-level sales is heavy tailed. As stated in footnote 19, a common method to evaluate heavy-tailed data is to fit a Pareto distribution to the (undefined) right tail of the data set and observe whether the shape parameter is smaller than two. The evolution of the Pareto shape parameter in function of its scale parameter is plotted Figure 2. Several conclusions can be drawn from this Figure. First, in line with evidence from Head et al. (2014); Bas et al. (2017); Nigai (2017); Bee and Schiavo (2018), the shape parameter lies below two over a majority of the data range. The shape parameter surpasses the value of two only when the Pareto distribution is fitted to less than 2 observations (the last 0.001% of the data). We can therefore safely conclude that, also in this data set, firm-level sales is heavy tailed. Second, if the Pareto scale parameter is determined such that more than 4.93% of the data is considered Pareto, the shape parameter falls below one. This results in an infinite first moment of the distribution and renders the distribution useless for heterogeneous firms models under the continuum of firms assumption as described in Section 2.[21] Third, Figure 2 demonstrates the motivation in the trade literature to rely on a small percentage of the complete data to fit a

[21] As mentioned by Eaton et al. (2012), distributions with infinite first moment can still be used in heterogeneous firms models if the continuum of firms assumption is abandoned. See also Section 2.

Pareto distribution (see for instance di Giovanni et al. (2011) who consider 13.89% of the data and Bee and Schiavo (2018); Bas et al. (2017) who reduce the data set to the 1% largest firms) in order to obtain a distribution fit in line with model assumptions. Fourth, this figure resonates with the extensive discussion in Perline (2005) on the evidence for Zipf's law[22], or the shape parameter of the Pareto distribution in general, being heavily dependent on sample size. Fifth and last, we can conclude from the continuous increase of the shape parameter for the complete data-range that the Pareto distribution is not a good fit to the (complete) data. If (the tail of) the data would be Pareto distributed, the shape parameter would be relatively constant and independent of sample size. This is neither the case for the complete data set, nor for a right-truncated sample (tail) of the data.

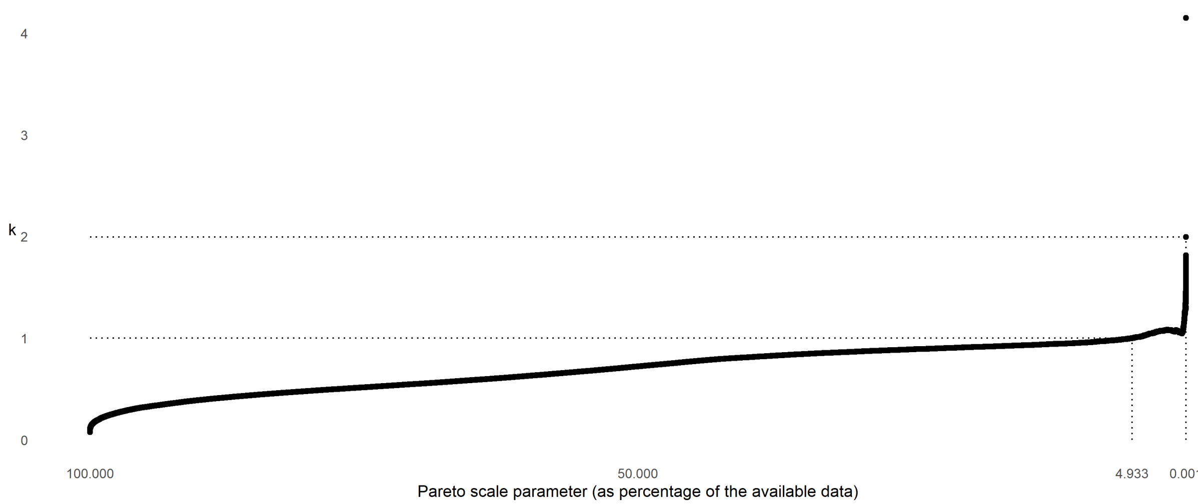


Figure 2: Pareto shape parameter in function of its location parameter

5.2 Distribution evaluation

Having established that firm sales is heavy-tailed, we move on to evaluate the parametric measurement of firm-level heterogeneity in line with the described methodology. Table 2 reports the evaluation statistics for all fitted Pareto-Lognormal combinations of distributions, ordered by the maximum NAD on the zeroth moment T^0, or the Kolmogorov-Smirnov test statistic.[23]

Based on the CDF fit, we observe that none of the currently considered distributions provide a

[22]Zipf's law states that the firm sales distribution follows a Pareto distribution with shape parameter equal (or close to) one (Axtell, 2001; Gabaix, 2009).

[23]All distributions are fitted using Maximum Likelihood.

fit that is not rejected by the data. According to the maximum NAD on the zeroth moment T^0, the largest error of the Pareto CDF fit to 100% of the data is no less than 0.49 out of a scale on 0 to 1. Quite naturally, Pareto distributions fitted to a smaller range in the right tail of the data, 5%, 2.1% (optimal tail length) or 1% respectively, provide an increasingly worse fit to the complete data range. The Lognormal distribution is only ranked just above the Pareto distribution. This demonstrates the need for a combination of distributions to adequately capture heterogeneity in productivity, as distributions with a larger number of parameters provide a better CDF fit. The largest error on the Double-Pareto Lognormal CDF fit, for instance, is only 0.013. Still, this is not sufficient to claim it is a good fit to the data. The large sample size results in a Kolmogorov-Smirnov test with large power, so that even a rather small deviation from the proposed distribution results in a rejection of the null hypothesis. Indeed, if the data would be Double-Pareto Lognormally distributed with parameters as displayed in the second column of Table 2, the errors on the CDF fit would only surpass 0.0019 in 5% of the cases. This result holds for all distributions tested in this paper: the Kolmogorov-Smirnov test rejects the null hypothesis that the sample is drawn from the respective parametric distributions. These results are not due to the large deviation of an outlier in the data but of a consistent bad fit over the complete range of the data, as can be deduced from the total NAD on the zeroth moment S^0.

Table 2: Distribution fits to French domestic sales in 2012.

Distribution	Parms.	T_a^0	S_b^0	T_a^1	S_b^1
Double-Pareto Lognormal	k_1=1.00, μ=-2.15, Var=1.19, k_2=1.03	1.31	2.50	291.14	1448.96
		(0.07;0.19)***	(0.11;0.41)***	(81.88;250.18)***	(214.51;1238.05)***
Inv. Pareto-Lognormal-Pareto	k_1=0.96, μ_2=-2.17, Var_2=1.64, k_3=0.97	1.31	2.18	-	-
		(0.08;0.19)***	(0.11;0.41)***	(-;-)	(-;-)
Lognormal-Pareto	μ_1=-2.21, Var_1=1.78, k_2=1.04	1.48	4.46	221.74	1106.25
		(0.07;0.20)***	(0.11;0.43)***	(68.70;186.21)***	(166.54;922.51)***
Inv. Pareto-Lognormal	k_1=1.00, μ_2=-2.13, Var_2=1.76	1.82	3.25	47.22	220.28
		(0.07;0.20)***	(0.11;0.42)***	(0.39;1.52)***	(0.21;6.28)***
Right-Pareto Lognormal	μ=-2.82, Var=1.71, k_2=1.55	1.89	5.23	30.29	134.76
		(0.07;0.19)***	(0.11;0.40)***	(1.40;6.15)***	(1.08;28.23)***
Left-Pareto Lognormal	k_1=1.30, μ=-1.40, Var=1.66	2.07	3.99	48.58	225.48
		(0.07;0.19)***	(0.11;0.42)***	(0.37;1.33)***	(0.22;5.68)***
Lognormal	μ=-2.17, Var=1.83	2.17	5.05	42.71	194.54
		(0.07;0.20)***	(0.11;0.42)***	(0.43;1.68)***	(0.25;7.11)***
Pareto (100%)$_c$	x_{min}=0.00, k=0.08	49.33	115.05	-	-
		(0.08;0.19)***	(0.11;0.42)***	(-;-)	(-;-)
Pareto (5%)$_c$	x_{min}=2.20, k=1.00	95.00	237.46	-	-
		(0.07;0.19)***	(0.11;0.42)***	(-;-)	(-;-)
Pareto (2.1%)$_c$	x_{min}=5.43, k=1.07	97.90	244.93	7902.78	39352.22
		(0.07;0.20)***	(0.11;0.44)***	(33.23;80.18)***	(63.83;392.95)***
Pareto (1%)$_c$	x_{min}=10.79, k=1.07	99.00	247.69	16793.65	83802.38
		(0.07;0.19)***	(0.11;0.42)***	(39.10;88.62)***	(68.93;435.44)***

Notes: All distributions fitted using Maximum Likelihood.

Values between parentheses report the 5th and 95th quantile of the parametric bootstrapped test statistic with 999 replications. ***, **, * indicate significance of this test at 1%, 5% and 10% respectively.

$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose, $_c$ 100%, 5%, 2.1% (optimal tail length), 1% indicate the range of the right tail of the data over which the Pareto distribution was fitted.

A similar picture emerges from the statistics on the first moment. None of the currently considered distributions appear to be able to provide an estimate of the first moment that is sufficiently close to the empirical values. In contrast to the test statistics on the zeroth moment, I have no statistical basis to draw conclusions regarding the appropriateness of the distribution from this test. Still, these fit statistics contain additional information to discriminate between distribution choices.

The ranking of distributions according to the statistic on the first moment does not accord with the statistics on the zeroth moment, but relates to the heavy-tailedness of the imposed parametric distribution. The Inverse Pareto - Lognormal- Pareto distribution is ranked second according to the KS test statistic but has infinitely large error for the maximum NAD on the first moment. This is because it has a Pareto shape parameter below one and therefore an infinite first moment. Though, also distributions with finite estimates of the first moment can display an unusual high error relative to their CDF fit. The right-Pareto-tailed Lognormal distribution, for instance, with a small Pareto shape parameter of 1.04 displays a maximum error T^1 on its estimate of the first moment that is up to 7 times bigger than the Right-Pareto Lognormal distribution with a Pareto shape parameter of 1.55.[24]

Figure 3 provides a visual perspective on these large errors of the first moment estimates. It displays the bootstrapped distribution of finite sample average sales (the unbounded first moment) drawn from each respective parametric distribution, as described in Section 3. The box plots delineate the 5th, 25th, 50th, 75th and 95th quantile of the limiting distribution of parametric finite sample average sales, while the green circles indicate the average of all these parametric finite sample estimates. Empirical average sales are indicated by the vertical blue line, while the parametric plug-in population estimates, if finite, are indicated by yellow diamonds.

Two conclusions can be drawn from Figure 3. First, it can be observed that the speed of convergence, and therefore the finite sample bias of average sales, depends on the heavy-tailedness of the assumed parametric distribution. We can deduce the finite sample bias from the Figure as the difference between the average parametric bootstrapped sample estimates (green circle) and the parametric plug-in population mean (yellow diamonds). Note that the less heavy-tailed

[24]Even though the Right-Pareto Lognormal is the result of a multiplication of the Lognormal distribution with a Pareto distribution and as such does not have cutoff values, it can be inferred from Figure 2 that a Pareto shape parameter of 1.55 means the Pareto tail covers approximately 0.015% of the data.

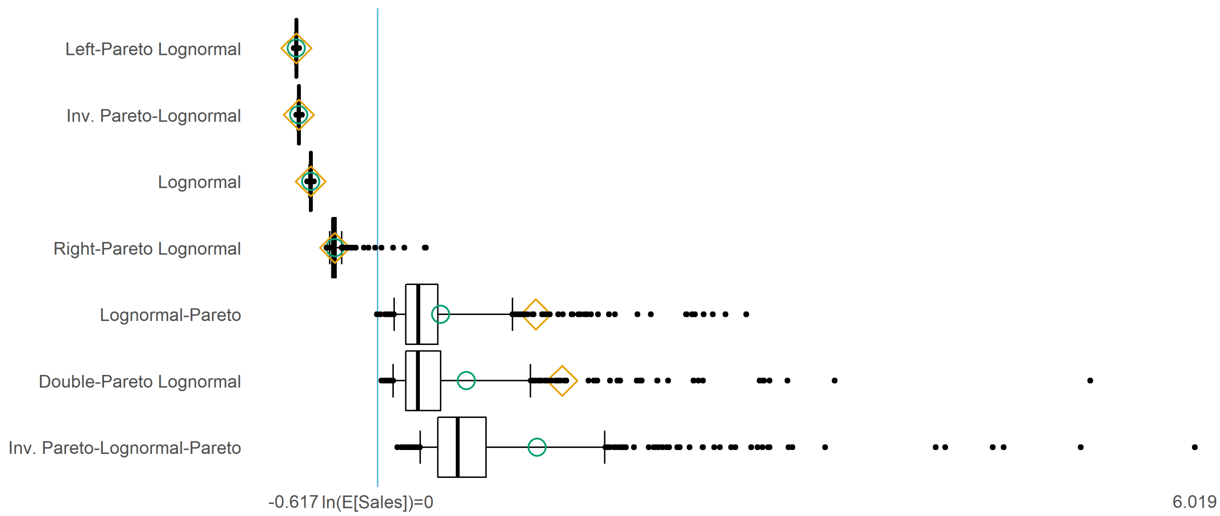


Figure 3: Bootstrapped (1,000 reps) limiting distributions of sales' parametric sample average.
Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the limiting distribution of parametric sample average sales, while the green circles indicate the average of all these parametric sample estimates. Nonparametric sample average sales are indicated by the vertical blue line, while the parametric population estimates, if finite, are indicated by yellow diamonds. All bootstrap sample values obtained from a sample of 500,388 firms. The Pareto distribution is excluded for expositional purpose due to too large deviations: the log of average Pareto (100%) finite sample estimates amounts to 236.98.

Lognormal distribution and Lognormal distributions with Pareto left tails are densely distributed around this mean. This implies a small finite sample bias and indicates consistency is attained. For heavy-tailed distributions, on the other hand, this is not the case. As described in Subsection 3.3, parametric finite sample averages are influenced by rare but extremely large data values. The parametric bootstrap distribution of the sample averages is therefore leptokurtic and heavy-tailed. This results in a large (or even infinite for the Inv. Pareto - Lognormal- Pareto distribution) finite sample bias. Moreover, this finite sample bias increases with the heavy-tailedness of the imposed distributions, an indication of the decreasing speed of convergence with an increasingly heavy tail. Overall, the relation between finite sample and population estimates depends on the heavy-tailedness of a distribution, hampering a direct comparison between finite sample estimates from differently-tailed distributions.

Second, I reject the premise that the empirical sample average can be obtained from any of the proposed distributions. Based on the distribution of parametric bootstrapped finite sample estimates, the following test is constructed (see also Section 3.2): 'how likely is the (positive or negative) deviation of the parametric plug-in population estimate (yellow diamond) from the empirical estimate (vertical blue line) under the parametric assumption.' Under the null hypothesis that the empirical estimate originates from the parametric distribution, the parametric bootstrap distribution should surpass the blue line in at least 5% of the cases. In line with the results summarized in Table 1, we observe that this is the case for none of the currently assumed parametric distribution.

6 Aggregate implications

In the previous section, it was established that none of the currently considered parametric distributions provide a sufficiently good fit to the data and that depending on the distribution, population estimates of this first moment can differ substantially from finite sample estimates. However, as stated in Section 1, this first moment "summarizes the information in the distribution of productivity levels relevant for all aggregate variables" (Melitz, 2003, p.1700). The question is then how important these empirical findings are for quantitative model evaluations in the data. To establish the importance, I center my arguments around the equilibrium in a heterogeneous firms model and two resulting aggregate trade statistics that enjoy widespread attention from both researchers and policymakers: the aggregate trade elasticity and Gains From Trade.

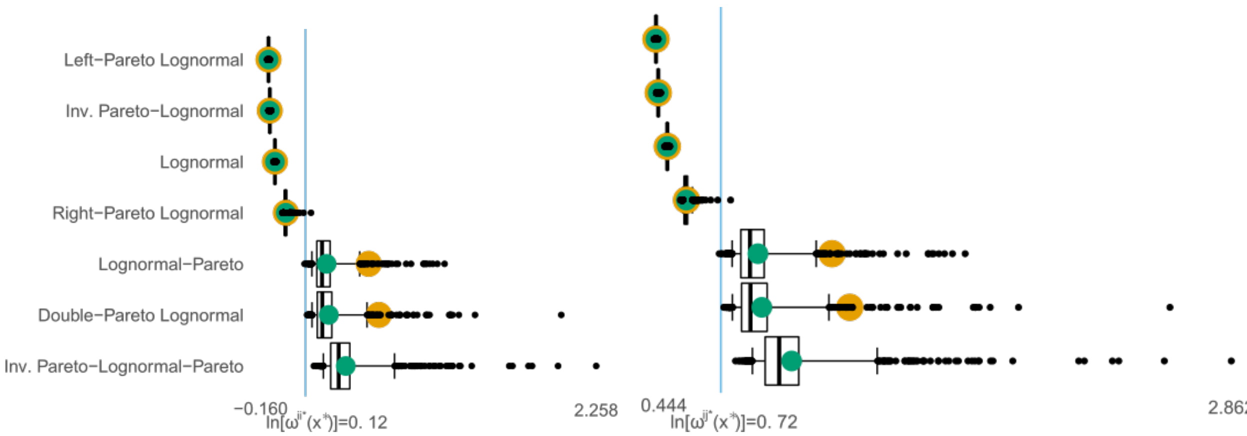
6.1 Melitz (2003) equilibrium

I calibrate the heterogeneous firms model presented in section 1 with both the nonparametric and parametric distribution fits. The parameterization of the model is standard (Melitz and Redding, 2015; Bee and Schiavo, 2018; Head et al., 2014). I work with two symmetric countries i and j and choose labor in one country as the numeraire, so that $W^i = W^j = 1$. The elasticity of substitution σ is set to four. Fixed entry costs are set to 0.545 ($f^e = 0.545$) and fixed costs are set to 1 ($f^{ii} = f^{ij} = 1$). In order to parametrize the model according to the parametric approximation of the heterogeneity distribution, the distributional parameters are recovered from the empirical analysis. The nonparametric estimator is obtained by interpolating the $0th$ and $(\sigma - 1)$th raw sample moments.[25] The variable trade costs are assumed to be 1.83 (Melitz and Redding, 2015).

The results from the calibration exercise are displayed in Figure 4. This figure displays, similarly

[25]See online Appendix B for a discussion on the parametrization of the (non-)parametric model statistics.

to Figure 3, the bootstrapped parametric distributions of both the domestic and exporting cutoffs in the calibrated model. It is immediately observable that the estimates of these cutoffs carry a large resemblance to the estimates of the sample mean in Figure 3.



(a) Domestic Cutoff (b) Exporting Cutoff

Figure 4: Bootstrapped (1,000 reps) limiting distributions of parametric sample cutoffs in the Melitz model.

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the asymptotic distribution of parametric finite sample domestic and exporting cutoffs. Yellow diamonds represent the plug-in (population) estimates. Green circles the average parametric bootstrapped sample estimates, while empirical average sales are indicated by the vertical blue line. All sample values obtained from a sample of 500,388 firms. The Pareto distribution is excluded for expositional reasons due to too large deviations.

First, we observe from this figure that the limiting distributions of the cutoffs are leptokurtic and heavy-tailed when the underlying distribution is heavy-tailed. This observation underwrites the incompatibility of a heavy-tailed sales distribution with a continuum firms assumption (see for instance Eaton et al. (2012); Gabaix (2011); di Giovanni and Levchenko (2012)). When sales are heavy tailed, firms do not reduce to a point on a continuum and carry the potential to affect the equilibrium of heterogeneous firms models. This evidence does not originate from the transmission of firm-level to aggregate volatility (Gabaix, 2011; di Giovanni and Levchenko, 2012) but rather emphasizes the existence of finite sample biases. Eaton et al. (2011) similarly demonstrate the importance for aggregate outcomes of the luck of the technology draw at the level of individual firms by revealing large variation in the manufacturing price level at the country level in finite samples.[26]

[26](di Giovanni and Levchenko, 2012, p.1106) also observed a highly skewed asymptotic distribution of the sample

Second, we can also observe from the calibration results in Figure 4 that the ranking of the parametric relative to the nonparametric estimates correlate strongly with the test statistics on the first moment of the distribution. Corroborating results from the previous subsection, the ranking is vastly different from the log-likelihood and CDF ranking. Thus, whereas the trade literature up till now has been focusing on providing a good fit to the CDF and/or PDF of the sample distribution, the results confirm the theoretical finding that the most important statistic to be matched when quantifying a heterogeneous firms model is the first moment of the sales distribution, or equivalently, the $(\sigma - 1)$th moment of the productivity distribution.

Third, none of the limiting distribution overlaps with the nonparametric estimates. This is in line with the statistical tests reported earlier. I reject the premise that the empirical sample solution of the Melitz (2003)-model originated from one of the considered distributions.

6.2 Aggregate trade elasticity

Aggregate trade elasticities are a central element to evaluate the welfare effects of trade liberalization. They also are a typical example of an aggregate statistic that sometimes has to be deduced from small data samples, as the number of firms that establish trade between two countries can be small. It is therefore of importance to evaluate whether the findings in this paper influences the interpretation of aggregate trade elasticities in finite samples.

The expression for aggregate trade between country i and j can be obtained by summing equation 3 across all active firms:

$$X^{ij} = \left(\frac{\sigma}{\sigma - 1} \tau^{ij} W^i \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma} M^{ie} m^{\sigma-1}_{\omega^{ij*}}, \tag{13}$$

with M^{ie} the number of entering firms in country i. The partial sensitivity of aggregate trade to changes in variable trade costs, the aggregate trade elasticity, is then defined as: (Chaney, 2008; Arkolakis et al., 2012; Melitz and Redding, 2014; Bas et al., 2017):[27]

mean in their simulation exercises. To circumvent resulting aggregate volatility between iterations, they chose to work with the median value of the distribution.

[27]Aggregate trade elasticity is here defined as the direct response of aggregate trade to trade costs, keeping the indirect effect trough the price index via its impact on the domestic cutoff fixed (Melitz and Redding, 2015).

$$\gamma^{ij} = \left.\frac{\partial lnX^{ij}}{\partial ln\tau^{ij}}\right|_{\omega^{ii*}} = 1 - \sigma + \frac{(\omega^{ij*})^{\sigma} g(\omega^{ij*})}{\frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right)\omega^{\sigma-1}}, \tag{14}$$

which can be rewritten in function of the bilateral intensive margin of trade and the weighted extensive margin of trade (Bas et al., 2017):

$$\gamma^{ij} = \underbrace{1-\sigma}_{\text{intensive margin}} - \underbrace{\frac{\min(\omega^{ij})}{E\left[x^{ij}(\omega^{ij})|\omega^{ij} \geq \omega^{ij*}\right]}}_{\text{weights}} \text{ x } \underbrace{\frac{dlnM^{ij}}{dln\tau^{ij}}}_{\text{extensive margin}} \tag{15}$$

where $\frac{dlnM^{ij}}{dln\tau^{ij}} = \frac{\omega^{ij*}g(\omega^{ij*})}{1-G(\omega^{ij*})}$ and $E\left[x^{ij}(\omega^{ij})|\omega^{ij} \geq \omega^{ij*}\right] = \frac{\frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(\omega>\omega^{ij*})\omega^{\sigma-1}}{1-G(\omega^{ij*})}$. Intuitively, the weight on the extensive margin should be decreasing as the market gets easier (that is, as ω^{ij*} declines). This evolution of the aggregate trade elasticity for all fitted distributions is displayed in Figure 5 with an assumed elasticity of substitution $\sigma = 4$. We can observe that for Lognormal and Lognormal-Pareto combinations of distributions, the weight on the extensive margin indeed decreases as the market gets easier, until the intensive margin value ($1 - \sigma = -3$) is reached. Simple Pareto distributions,[28] on the other hand, do not vary over the range of the distribution and maintain a trade elasticity equal to their fitted shape parameter ($k(\sigma - 1)$). This is a result specific to the Pareto distribution for which the intensive margin cancels out in elasticity calculations, as discussed in Chaney (2008).

Given the central role of the first moment of the sales distribution in the calculation of this statistic, however, we should by now not expect these standard results to be validated in the data. Indeed, this is confirmed in Figure 6. In general, trade elasticities of heavy-tailed distributions are biased downwards in finite samples as the weight on the extensive margin is underestimated.[29] As also discussed in section 5, it can be seen that this finite sample bias depends on the heavy-tailedness of the underlying distribution and that direct comparisons of distribution-implied elasticities are therefore not warranted.

[28] Whereas Pareto distributions could not be displayed in graphs in the previous section due to too large deviations, they can be displayed in the graphs in the current and next section. This because the previous section discussed the heterogeneous firms model in levels, while the current section and next section discusses the model in changes, independent of constant deviations.

[29] As it is difficult to obtain a good nonparametric approximation of the density function when the distribution is heavy-tailed, no empirical values are provided.

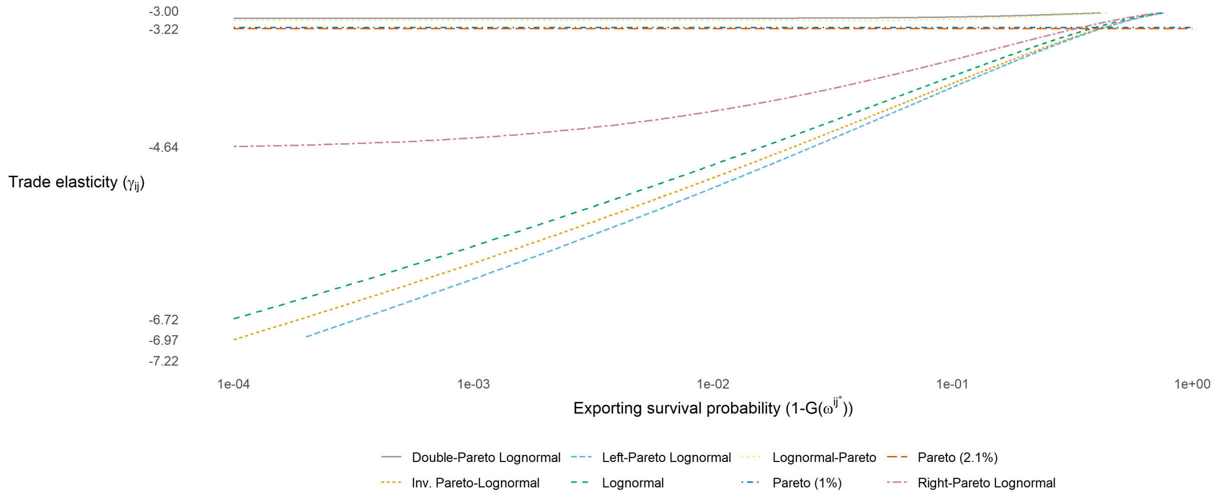


Figure 5: Evolution of trade elasticity estimates along the export survival function of the imposed parametric distribution.

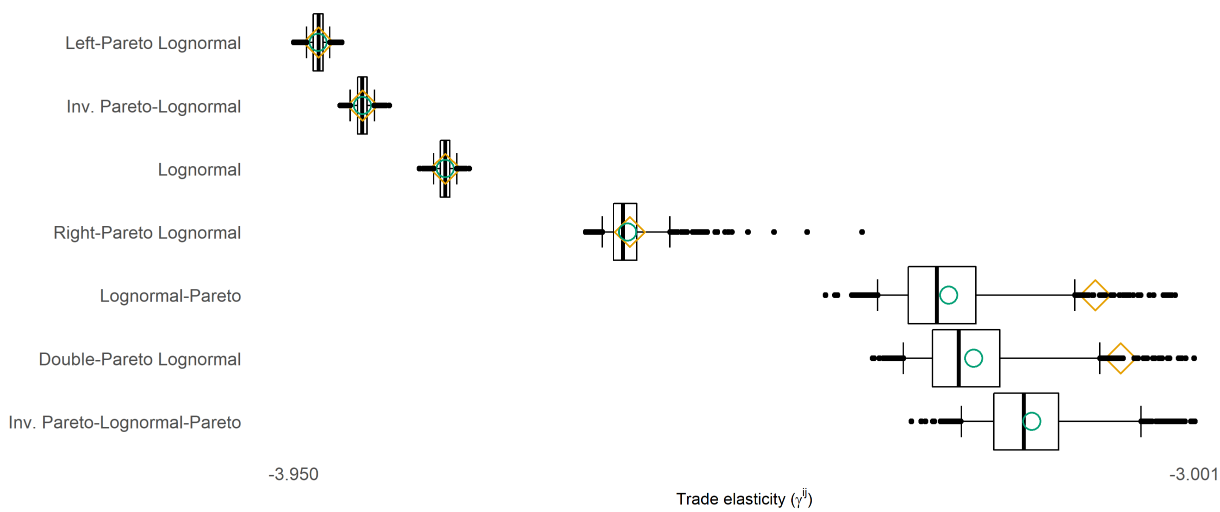


Figure 6: Bootstrapped (1,000 reps) limiting distribution of parametric sample trade elasticities at 10% exporting survival probability $(1 - G(\omega^{ij*}) = 0.9)$

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the asymptotic distribution of parametric finite sample trade elasticities. Yellow diamonds are the plug-in (population) estimates. Green circles the average parametric bootstrapped sample estimates (note that the density estimates in elasticity calculations have not been approximated but represent the parametric density estimates). All sample values obtained from a sample of 500,388 firms.

A finite sample is not the sole source of bias in realized trade elasticities, however. As discussed above, depending on the parametric distributional assumption, large individual firms can affect the aggregate. In the case of trade elasticities, as markets get tougher the individual weight of these larger firms in trade elasticity estimates increases even more, possibly inducing a truncation bias.[30] This can be observed from Figure 7 for two selected parametric distributions, parametric sample estimates tend to decrease as markets get tougher when evaluated in finite samples. The constancy of the trade elasticity under a Pareto distribution is often relied upon as a feature for Gains From Trade calculations across countries (Costinot and Rodríguez-Clare, 2014). It is, however, also considered a weakness of the Pareto distribution (Melitz and Redding, 2015; Bas et al., 2017).[31] The results presented below uncover that, in finite samples, the Pareto distribution can display variance in the trade elasticity across markets as its intensive margin remains active.

6.3 Gains From Trade

From the (partial) trade elasticity the discussion continues on to the full trade elasticity and its related welfare implications. One of the main reasons to quantitatively solve a heterogeneous firms trade model is to quantify changes in welfare due to a trade shock, Gains From Trade (GFT). I evaluate whether these gains can be expected to realize in the real (finite) world and whether they originate from correctly approximating the channels through which trade affects welfare.

These GFT can be written as log changes in real per-capita income due to an exogenous increase in variable trade costs τ_{ij} to τ'_{ij}. This can be further decomposed into the channels through which trade affects welfare: trade costs (τ^{ij}), the number of firms (M^i), the probality of successful entry into the domestic market ($m^0_{\omega^{ii*}}$), the average productivity of firms exporting from i to j ($m^{\sigma-1}_{\omega^{ij*}}$)[32] and the bilateral trade share (λ^{ij}):

[30]Sager and Timoshenko (2019) similarly discuss the existence of a truncation bias in the calculation of trade elasticities due to administrative reporting thresholds for export sales.

[31](Bas et al., 2017, p. 2) call upon the "direct evidence that aggregate elasticities are non-constant across country pairs" as evidence in favor of the Lognormal relative to the Pareto distribution. The authors argue that the variance of the weights on the extensive margin elasticity across markets allows to differentiate between distributions. This argument on the importance of matching the conditional mean as the weight on the extensive margin elasticity is very similar to the argument on the importance of matching the unconditional mean originating from the free-entry condition. When evaluating their implied population estimate with finite sample estimates, however, the authors do not take into account the possible existence of finite sample and truncation biases.

[32]I define average productivity here as average productivity unconditional on successful entry, in contrast to the definition conditional on successful entry in (Melitz, 2003, p.1702).

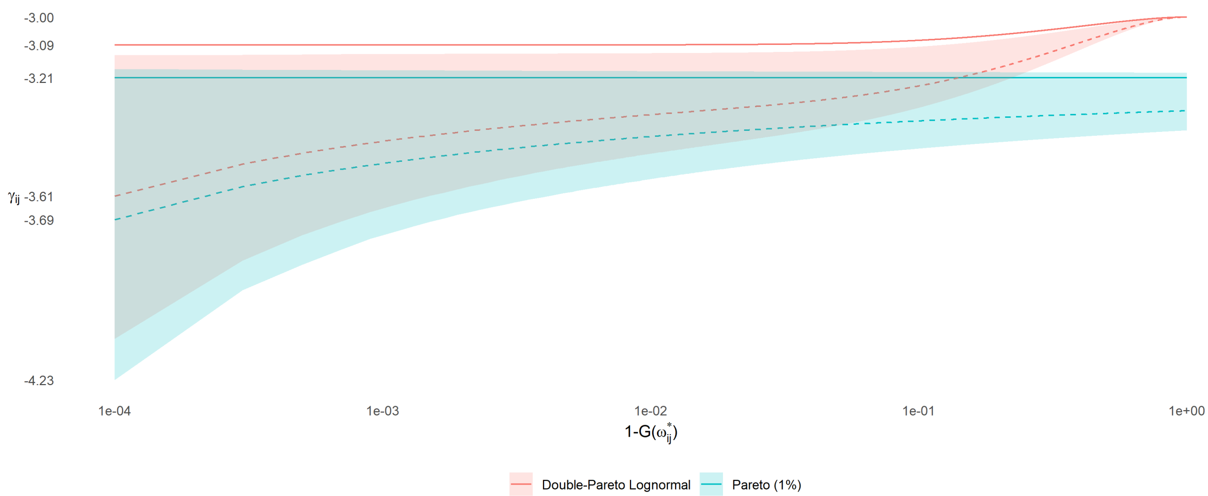


Figure 7: Evolution of trade elasticity estimates along the survival function of the imposed parametric distribution plus bootstrapped (1,000 reps) limiting distribution of parametric sample trade elasticities

Notes: Full lines represent the plug-in (population) elasticity. Dashed lines the average parametric bootstrapped sample estimates (note that the density estimates in elasticity calculations have not been approximated but represent the parametric density estimates). Shaded areas cover the 5th-95th quantile interval of the bootstrapped parametric sample estimates. All sample values obtained from a sample of 500,388 firms.

$$
\begin{aligned}
100 \times ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} &= 100 \times -ln\frac{(P^i)'}{P^i} \qquad (16)\\
&= 100 \times -\left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma-1}\left(ln\frac{(M^i)'}{M^i} - ln\frac{(m^0_{\omega^{ii*}})'}{m^0_{\omega^{ii*}}} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}}\right)\right].
\end{aligned}
$$

The heterogeneous firms model is calibrated equivalently to section 6.1 while reducing the variable trade costs $\tau = 3 \rightarrow \tau' = 1$. The GFT obtained from this exercise are displayed in Figure 8. This far in, the reader will not be surprised anymore by the pattern emerging. The distributions without Pareto tail show no signs of biasedness in finite samples as the parametric plug-in population estimates (yellow diamonds) overlap with the average parametric bootstrapped sample estimates (green circles), but do significantly underestimate Gains From Trade relative to the 'true' Gains From Trade implied by the empirical distribution (vertical blue line). Pareto-tailed distributions, on the other hand, are largely biased in finite samples and tend to overestimate the 'true' gains. The differences in speed of convergence and resulting finite sample biases again demonstrate that direct comparisons between distributions are hard to interpret and can result in erroneous conclusions. It also demonstrates that Gains From Trade predictions based on Pareto-tailed distributions are very uncertain and unlikely to be realized in the real world with only a finite number of firms present.

Two (truncated) distributions positively stand out regarding their performance relative to the 'true' gains: the Pareto distribution respectively covering 2.1% and 1% of the right tail of the data. This acceptable performance is good news from an aggregate perspective, as the Pareto distribution allows for straightforward GFT calculations from aggregate data (Arkolakis et al., 2012; Costinot and Rodríguez-Clare, 2014). It should, however, not be interpreted as evidence in favor of the Pareto distribution to predict firm heterogeneous responses to trade shocks (Melitz and Redding, 2015). As discussed in section 5.2, this truncated Pareto distribution does not provide a good fit to the complete data. One would thus expect that, in line with Melitz and Redding (2015), certain channels through which trade affects welfare can not be approached sufficiently by this distribution. This idea is confirmed in Table 3, which displays the weighted components of welfare gains (see eq. 16) for all considered distributions. It can be seen that the deviation of the parametric results compared to the empirical distribution are relatively small for the changes in number of firms and

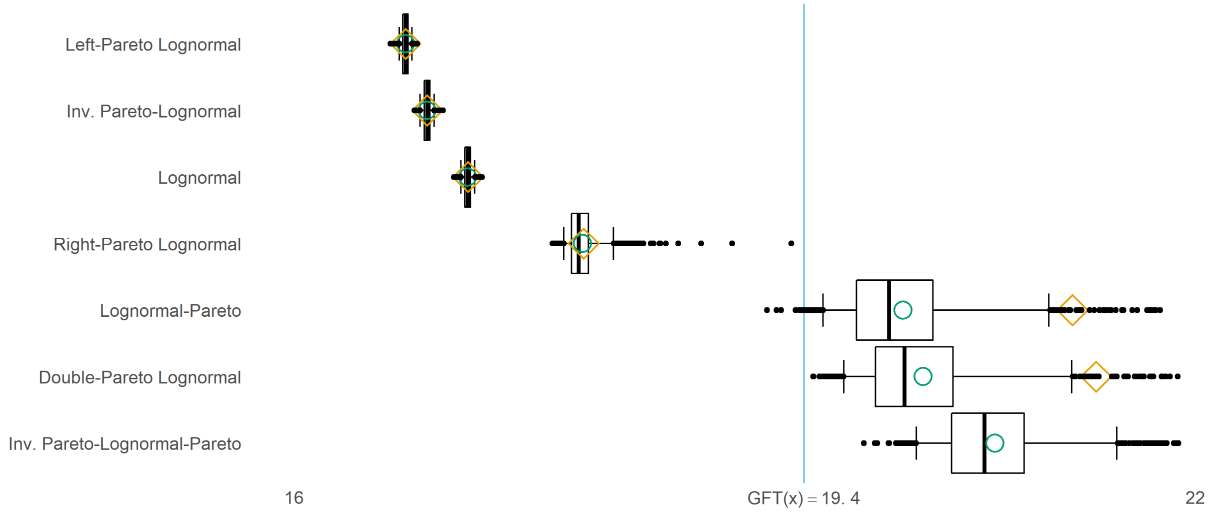


Figure 8: Bootstrapped (1,000 reps) limiting distributions of the parametric sample gains from a reduction in variable trade costs $\tau = 3 \to \tau' = 1$.

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the asymptotic distribution of parametric finite sample GFT. Yellow diamonds represent the plug-in (population) estimates of GFT. Green circles are the average parametric bootstrapped finite sample GFT and the empirical sample GFT are indicated by the vertical blue line. All sample values obtained from a sample of 500,388 firms.

in the probability of successful entry into the domestic market. Large differences originate from the changes in average productivity of exporting firms and from changes in trade shares. Heavy-tailed distributions largely underestimate the positive effect of the increase in average productivity of exporting firms and the negative effect of the increase in the bilateral trade shares compared to the empirical distribution, while the reverse is true for lighter-tailed distributions. These findings once again underscore the importance of the first moment of the sales distribution as a viable predictor for distributional performance in firm heterogeneous trade models.

Table 3: Decomposition of procentual welfare gains from a reduction in variable trade costs $\tau^{ij} = 3 \rightarrow (\tau^{ij})' = 1$.

Distribution	Parms.	$ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i}$	$-\ln\frac{(\tau^{ij})'}{(\tau^{ij})}$	$\frac{1}{\sigma-1}ln\frac{(M^i)'}{M^i}$	$\frac{1}{\sigma-1}ln\frac{(m^0_{\omega^{ii*}})'}{m^0_{\omega^{ii*}}}$	$\frac{1}{\sigma-1}ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}}$	$-\frac{1}{\sigma-1}ln\frac{(\lambda^{ij})'}{\lambda^{ij}}$
Pareto (100%)	2	-	1.10	-	-	-	-
		(-0.00;0.00)***	(1.10;1.10)	(-0.22;-0.22)***	(-0.00;0.00)***	(0.00;0.00)***	(-0.88;-0.88)***
Left-Pareto Lognormal	3	0.17	1.10	-0.17	0.15	0.55	-1.46
		(0.17;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.53;0.56)***	(-1.48;-1.45)***
Inv. Pareto-Lognormal	3	0.17	1.10	-0.17	0.15	0.52	-1.43
		(0.17;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.50;0.54)***	(-1.45;-1.41)***
Lognormal	2	0.17	1.10	-0.17	0.15	0.48	-1.39
		(0.17;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.46;0.49)***	(-1.40;-1.37)***
Right-Pareto Lognormal	3	0.18	1.10	-0.18	0.16	0.33	-1.24
		(0.18;0.18)***	(1.10;1.10)	(-0.18;-0.17)***	(0.16;0.17)***	(0.29;0.36)***	(-1.26;-1.20)***
Empirical	0	0.19	1.10	-0.18	0.18	0.17	-1.07
Pareto (2.1%)	2	0.21	1.10	-0.22	0.22	0.07	-0.96
		(0.19;0.21)	(1.10;1.10)	(-0.22;-0.21)***	(0.21;0.22)***	(0.05;0.16)**	(-1.07;-0.94)**
Pareto (1%)	2	0.21	1.10	-0.22	0.22	0.06	-0.95
		(0.19;0.21)	(1.10;1.10)	(-0.22;-0.21)***	(0.20;0.22)***	(0.06;0.17)**	(-1.07;-0.95)**
Lognormal-Pareto	3	0.21	1.10	-0.22	0.22	0.04	-0.92
		(0.20;0.21)**	(1.10;1.10)	(-0.22;-0.20)***	(0.20;0.22)***	(0.04;0.15)***	(-1.05;-0.93)***
Double-Pareto Lognormal	4	0.21	1.10	-0.22	0.22	0.03	-0.91
		(0.20;0.21)***	(1.10;1.10)	(-0.22;-0.20)***	(0.19;0.22)***	(0.04;0.14)***	(-1.04;-0.92)***
Inv. Pareto-Lognormal-Pareto	4	-	1.10	-	-	-	-
		(0.20;0.21)***	(1.10;1.10)	(-0.21;-0.20)***	(0.19;0.21)***	(0.02;0.11)***	(-1.00;-0.91)***
Pareto (5%)	2	-	1.10	-	-	-	-
		(0.20;0.21)***	(1.10;1.10)	(-0.21;-0.20)***	(0.20;0.21)***	(0.03;0.12)***	(-1.02;-0.92)***

Notes: Values between parentheses report the 5th and 95th quantile of the parametric bootstrapped statistics with 999 replications. ***, **, * indicate the rejection of a signifcant overlap of the parametric bootstrapped statistic with the empirical statistic at 1%, 5% and 10% respectively.

7 Conclusion

This paper emphasizes the importance of two sufficient statistics to characterize firm-level heterogeneity from firm-level sales in a CES heterogeneous firms model. Based on these statistics, and provided the nature of firm-level sales, we generate evidence that individual firms carry the potential to influence the aggregate economy. Aggregate statistics are likely biased in finite samples, affecting current conclusions drawn with respect to the aggregate trade elasticity and GFT calculations. Moreover, we find that none of the currently considered characterizations of firm-level heterogeneity provide a significantly good fit to the data.

The results of this paper point out the importance of a continued research effort on the characterization of firm-level heterogeneity. A good approximation of this heterogeneity is not only desirable from a quantitative perspective, but could deliver information on the underlying firm-level dynamics that result in a specific heterogeneity distribution (Arkolakis, 2016; Luttmer, 2007). This would allow for a closer study of GFT resulting from within-firm efficiency evolution (Costantini and Melitz, 2008; Atkeson and Burstein, 2010) compared to the allocative, between-firm efficiency allocation of the workhorse heterogeneous firms trade model.

The results also carry consequence for the interpretation of GFT statistics. Currently, GFT estimates are usually obtained under the Pareto assumption (Costinot and Rodríguez-Clare, 2014). These statistics are likely an overestimation of reality, however, as the full potential of this heavy-tailed distribution is not attained in finite samples. In this light, the development of models that dispose of the continuum firms assumption provide an important contribution to the field (Eaton et al., 2012; di Giovanni and Levchenko, 2012).

References

Arkolakis, C. (2016). A Unified Theory of Firm Selection and Growth. *The Quarterly Journal of Economics 131*(1), 89–115.

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

Athreya, K. et al. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics 15*(2), 724–731.

Atkeson, A. and A. Burstein (2010). Innovation, firm dynamics, and international trade. *Journal of Political Economy 118*(3), 433–484.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science 293*(5536), 1818–1820.

Bakar, S. A., N. Hamzah, M. Maghsoudi, and S. Nadarajah (2015). Modeling loss data using composite models. *Insurance: Mathematics and Economics 61*, 146 – 154.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1 – 19.

Bee, M. and S. Schiavo (2018). Powerless: gains from trade when firm productivity is not Pareto distributed. *Review of World Economics 154*(1), 15–45.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2012). The empirics of firm heterogeneity and international trade. *Annual Review of Economics 4*(1), 283–313.

Carvalho, V. M. and B. Grassi (2019, April). Large firm dynamics and the business cycle. *American Economic Review 109*(4), 1375–1425.

Cavaliere, G., I. Georgiev, and A. Robert Taylor (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Reviews 32*(2), 204–219.

Cernat, L. (2014). Towards “Trade Policy Analysis 2.0”. From national comparative advantage to firm-level data. Chief Economist Note 4, European Central Bank.

Chaney, T. (2008, September). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-Law Distributions in Empirical Data. *SIAM Review 51*(4), 661–703.

Cornea-Madeira, A. and R. Davidson (2015). A parametric bootstrap for heavy-tailed distributions. *Econometric Theory 31*(3), 449–470.

Costantini, J. and M. Melitz (2008). The dynamics of firm-level adjustment to trade liberalization. *The organization of firms in a global economy 4*, 107–141.

Costinot, A. and A. Rodríguez-Clare (2014). Chapter 4 - Trade Theory with Numbers: Quantifying the Consequences of Globalization. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 197–261. Elsevier.

Davidson, R. (2012). Statistical inference in the presence of heavy tails. *The Econometrics Journal 15*(1), C31–C53.

Dewitte, R., M. Dumont, G. Rayp, and P. Willemé (2019). Unobserved Heterogeneity in the Productivity Distribution and Gains From Trade. MPRA Paper 102711, Ghent University.

di Giovanni, J. and A. A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy 120*(6), 1083–1132.

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 85*(1), 42–52.

Durrett, R. (2010). *Probability: theory and examples*, Volume Fourth Edition. Cambridge university press.

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

Eaton, J., S. S. Kortum, and S. Sotelo (2012). International trade: Linking micro and macro. NBER Working Paper 17864, National bureau of economic research.

Fedotenkov, I. (2013). A bootstrap method to test for the existence of finite moments. *Journal of Nonparametric Statistics 25*(2), 315–322.

Fedotenkov, I. (2014). A note on the bootstrap method for testing the existence of finite moments. *Statistica 74*(4), 447–453.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, M. D. Pierola, and A. Rodríguez-Clare (2018). The Intensive Margin in Trade: Moving Beyond Pareto. Policy Research working paper WPS 8625, World Bank Group.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Gabaix, X. and R. Ibragimov (2011). Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics 29*(1), 24–39.

Hall, P. and Q. Yao (2003). Inference in arch and garch models with heavy–tailed errors. *Econometrica 71*(1), 285–317.

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and Trade without Pareto. *American Economic Review 104*(5), 310–16.

Ioannides, Y. and S. Skouras (2013). Us city size distribution: Robustly pareto, but only in the tail. *Journal of Urban Economics 73*(1), 18 – 29.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *The Annals of Statistics 17*(3), 1168–1175.

Luckstead, J. and S. Devadoss (2017). Pareto tails and lognormal body of us cities size distribution. *Physica A: Statistical Mechanics and its Applications 465*, 573 – 578.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics 122*(3), 1103–1144.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and S. J. Redding (2014). Chapter 1 - Heterogeneous Firms and Trade. In E. H. Gita Gopinath and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 1–54. Elsevier.

Melitz, M. J. and S. J. Redding (2015). New Trade Models, New Welfare Implications. *American Economic Review 105*(3), 1105–46.

Mrázová, M., P. Neary, and M. Parenti (2015). Technology, Demand, And The Size Distribution Of Firms. Economics Series Working Papers 774, University of Oxford, Department of Economics.

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

Perline, R. (2005). Strong, weak and false inverse power laws. *Statistical Science 20*(1), 68–88.

Reed, W. J. and M. Jorgensen (2004). The double pareto-lognormal distribution—a new parametric model for size distributions. *Communications in Statistics - Theory and Methods 33*(8), 1733–1753.

Romano, J. P. and M. Wolf (1999). Subsampling inference for the mean in the heavy-tailed case. *Metrika 50*(1), 55–69.

Sager, E. and O. A. Timoshenko (2019). The double emg distribution and trade elasticities. *Canadian Journal of Economics/Revue canadienne d'économique 52*(4), 1523–1557.

Taleb, N. N. (2019). How much data do you need? an operational, pre-asymptotic metric for fat-tailedness. *International Journal of Forecasting 35*(2), 677–686.

Trapani, L. (2016). Testing for (in) finite moments. *Journal of Econometrics 191*(1), 57–68.

Van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3. Cambridge university press.

Online Appendix to “From Heavy-Tailed Micro to Macro: on the characterization of firm-level heterogeneity and its aggregation properties.”

Ruben Dewitte*

16th September 2020

*Dewitte: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be. I am grateful to Luca David Opromolla, Gonzague Vannoorenberghe, Olivier Thas, Kris Boudt, Jasson Vindas Diaz, Gerdie Everaert, Glenn Rayp, Angelos Theodorakopoulos, Bruno Merlevede and Marijn Verschelde for their helpful comments and suggestions.

Appendix A Additional Figures and Tables

A.1 Figures

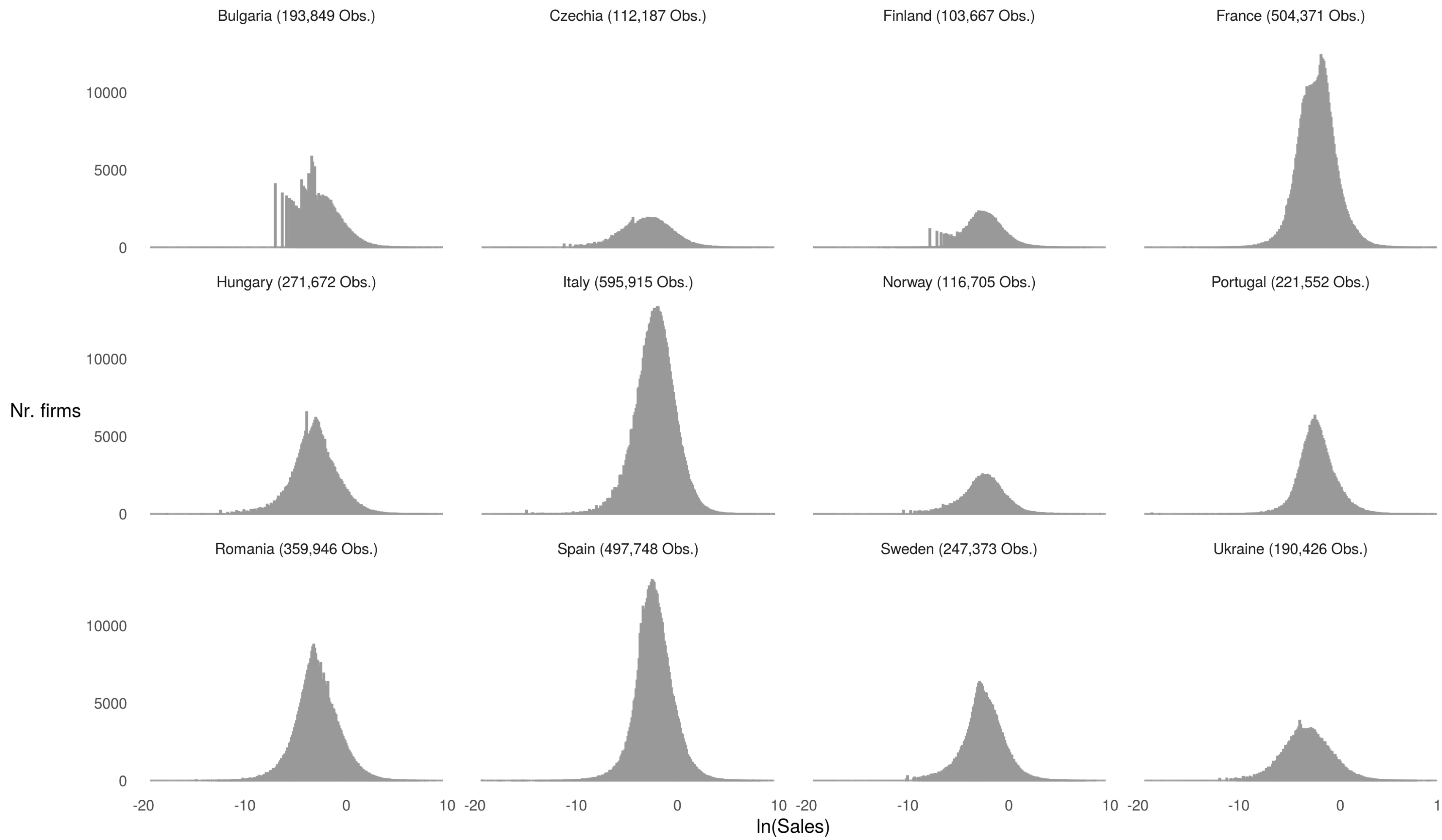


Figure 1: Firm sales density for selected European countries in 2012.
Note: All observations are normalized by country, dividing the observations with their respective mean value.

A.2 Tables

Table 1: Overview of all distributions considered.

Distribution	Abbreviation	Support	Parameters	Change in parameters from power transformation $a\omega^b$
Pareto	P	$[\omega_{min}, \infty[$	k, ω_{min}	$kb, \left(\frac{\omega_{min}}{a}\right)^{\frac{1}{b}}$
Lognormal	LN	$[0, \infty[$	μ, Var	$\frac{\mu - lna}{b}, \frac{Var}{b}$
Piecewise composite	PC	See ind. comp.	$\boldsymbol{\theta}$	See ind. comp.
Double-Pareto Lognormal	DPLN	$[0, \infty[$	k_1, μ, Var, k_2	$\frac{k_1}{b}, b\mu + log(a), Var, \frac{k_2}{b}$
Left-Pareto Lognormal	LPLN	$[0, \infty[$	k_1, μ, Var	$\frac{k_1}{b}, b\mu + log(a), Var$
Right-Pareto Lognormal	RPLN	$[0, \infty[$	μ, Var, k_2	$b\mu + log(a), Var, \frac{k_2}{b}$

Table 2: Overview of the probability and cumulative density functions of distributions considered.

Distribution	**PDF**	**CDF**
P	$\frac{k\omega_{min}^{k}}{\omega^{k+1}}$	$1-\left(\frac{\omega_{min}}{\omega}\right)^{k}$
LN	$\frac{1}{\omega Var\sqrt{2\pi}}e^{-(ln\omega-\mu)^2/2Var^2}$	$\Phi\left(\frac{ln\omega-\mu}{Var}\right)$
PC[a]	$\begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2}m_1^*(x\mid\boldsymbol{\theta}_1) & \text{if} \quad 0<x\leq c_1 \\ \frac{1}{1+\alpha_1+\alpha_2}m_2^*(x\mid\boldsymbol{\theta}_2) & \text{if} \quad c_1<x\leq c_2 \\ \frac{\alpha_2}{1+\alpha_1+\alpha_2}m_3^*(x\mid\boldsymbol{\theta}_3) & \text{if} \quad c_2<x<\infty \end{cases}$	$\begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2}\frac{M_1(x\mid\boldsymbol{\theta}_1)}{M_1(c_1\mid\boldsymbol{\theta}_1)} & \text{if} \quad 0<x\leq c_1 \\ \frac{\alpha_1}{1+\alpha_1+\alpha_2}+\frac{1}{1+\alpha_1+\alpha_2}\frac{M_2(x\mid\boldsymbol{\theta}_2)-M_2(c_1\mid\boldsymbol{\theta}_2)}{M_2(c_2\mid\boldsymbol{\theta}_2)-M_2(c_1\mid\boldsymbol{\theta}_2)} & \text{if} \quad c_1<x\leq c_2 \\ \frac{1+\alpha_1}{1+\alpha_1+\alpha_2}+\frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{M_3(x\mid\boldsymbol{\theta}_3)-M_3(c_2\mid\boldsymbol{\theta}_3)}{1-M_3(c_2\mid\boldsymbol{\theta}_3)} & \text{if} \quad c_2<x<\infty \end{cases}$
DPLN[b]	$\frac{k_2k_1}{k_2+k_1}\left[x^{-k_2-1}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)+ x^{k_1-1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\Phi^c\left(\frac{lnx-\mu+k_1Var^2}{Var}\right)\right]$	$\Phi\left(\frac{lnx-\mu}{Var}\right)-\frac{1}{k_2+k_1}\left[k_1x^{-k_2}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)- k_2x^{k_1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\Phi^c\left(\frac{lnx-\mu+k_1Var^2}{Var}\right)\right]$
LPLN[b]	$k_1x^{k_1-1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\Phi^c\left(\frac{lnx-\mu+k_1Var^2}{Var}\right)$	$\Phi\left(\frac{lnx-\mu}{Var}\right)-x^{k_1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\Phi^c\left(\frac{lnx-\mu+k_1Var^2}{Var}\right)$
RPLN[b]	$k_2x^{-k_2-1}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)$	$\Phi\left(\frac{lnx-\mu}{Var}\right)-x^{-k_2}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)$

Notes: [a] $\forall i\in I: m_i^*(x)=\frac{m_i(x)}{\int_{c_{i-1}}^{c_i}m_i(x)dx}$, [b] Φ and Φ^c stand for the standard normal and complementary standard normal cdfs.

Table 3: Expression of the y-bounded rth moment (μ_y^r) for all distributions considered.

Distribution	μ_y^r	**Additional parameter restrictions**[a]
P	$-(y)^{r-k}\,\frac{k\omega_{min}^k}{r-k}$	$k > r$
LN	$e^{\frac{r\left(rVar^2+2\mu\right)}{2}}\left[1-\Phi\left(\frac{lny-\left(rVar^2+\mu\right)}{Var}\right)\right]$	-
PC	$\begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2}\frac{(\mu_1)_y^r-(\mu_1)_{c_1}^r}{M_1(c_1)}+\frac{1}{1+\alpha_1+\alpha_2}\frac{(\mu_2)_{c_1}^r-(\mu_2)_{c_2}^r}{M_2(c_2)-M_2(c_1)}+\frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_y^r}{1-M_3(c_2)} & \text{if} \quad 0 < y \leq c_2 \\ \frac{1}{1+\alpha_1+\alpha_2}\frac{(\mu_2)_y^r-(\mu_2)_{c_2}^r}{M_2(c_2)-M_2(c_1)}+\frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_{c_2}^r}{1-M_3(c_2)} & \text{if} \quad c_1 < y \leq c_2 \\ \frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_y^r}{1-M_3(c_2)} & \text{if} \quad c_2 < y < \infty \end{cases}$	See ind. comp.
DPLN	$\begin{aligned} &-\frac{k_2k_1}{k_2+k_1}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\frac{y^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{lny-\mu-k_2Var^2}{Var}\right) \\ &-\frac{k_2k_1}{k_2+k_1}\frac{1}{r-k_2}e^{\frac{r^2Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2-\mu}{Var}\right) \\ &-\frac{k_2k_1}{k_2+k_1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\frac{y^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{lny-\mu+k_1Var^2}{Var}\right) \\ &+\frac{k_2k_1}{k_2+k_1}\frac{1}{r+k_1}e^{\frac{r^2Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2-\mu}{Var}\right) \end{aligned}$	$k_2 > r$
LPLN	$\begin{aligned} &-k_1e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\frac{y^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{lny-\mu+k_1Var^2}{Var}\right) \\ &+\frac{k_1}{r+k_1}e^{\frac{r^2Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2+\mu}{Var}\right) \end{aligned}$	-
RPLN	$\begin{aligned} &-k_2e^{k_2\mu+\frac{k_2^2Var^2}{2}}\frac{y^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{lny-\mu-k_2Var^2}{Var}\right) \\ &-\frac{k_2}{r-k_2}e^{\frac{r^2Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2+\mu}{Var}\right) \end{aligned}$	$k_2 > r$

Notes: [a] Additional parameter restrictions represent parameter restrictions needed to keep the statistic finite.

Appendix B Heterogeneous firms model

This appendix provides a detailed description of the heterogeneous firms models relied upon in the paper. Heterogeneous firms trade model with a finite number of firms is presented. The model features Constant Elasticity of Substitution (CES)-demand and monopolistic competition between a finite number of firms who ignore their aggregate impact (Dixit and Stiglitz, 1977; Krugman, 1980; di Giovanni and Levchenko, 2012), while remaining agnostic on the parametric specification of firm-level heterogeneity. For the number of firms going to infinity, the model is comparable to the Melitz (2003)-model.

B.1 Setup

Demand Consumer preferences in country $j \in J$ are defined over a finite number of horizontally differentiated varieties ($\varpi \in \Omega^i$) originating from country $i \in I$ and are assumed to take the Constant Elasticity of Substitution (CES) utility (U) form

$$U^j = \left(\sum_{i=1}^{I} \sum_{\varpi \in \Omega^i} q^{ij} \left(\varpi\right)^{\frac{\sigma-1}{\sigma}} d\varpi \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

with σ the elasticity of substitution between varieties. Utility maximization defines the optimal consumption and expenditure decisions over the individual varieties

$$\frac{q^{ij}(\varpi)}{Q^j} = \left[\frac{p^{ij}(\varpi)}{P^j} \right]^{-\sigma}, \tag{2}$$

where the set of varieties consumed is considered as an aggregate good $Q \equiv U$ and P is the CES aggregate price index.

Supply There is a finite number of businesses ($b \in B$) which choose to supply a distinct horizontally-differentiated variety. They are heterogeneous in terms of their productivity $\omega_b \in [0, \infty]$ drawn from the unconditional Cumulative Distribution Function (CDF) $G(\omega_b)$ after paying a fixed cost f^{ie} in terms of production factor L^i to enter the market.[1] There is zero probability of firm death.[2] Supply of the production factor to the individual firm is perfectly elastic, so that firms are effectively price (W^i) takers on the input markets. Once active, firms from country i have to pay a fixed cost f^{ij} to produce goods destined for country j. The cost function of the firm involves a fixed production cost, iceberg trade costs $\tau^{ij} > 1$ and a constant marginal costs that depends on its productivity: $f^{ij} + \left(\frac{\tau^{ij} q^{ij}}{\omega}\right) W^i$. Profit maximization of the firm, then:

[1] As ω_b is the sole heterogeneity component identifying individual firms, the subscript b is dropped in further derivations.

[2] The static specification in which there is zero probability of firm death follows most of the international trade literature.

$$\max_{q^{ij}} \pi^{ij} = \max_{q^{ij}} \left[p^{ij} q^{ij} - \left(f^{ij} - \frac{\tau^{ij} q^{ij}}{\omega} \right) W^i \right]$$
$$= \max_{q^{ij}} \left[\left(q^{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(Q^j\right)^{\frac{1}{\sigma}} P^j - \left(f^{ij} - \frac{\tau^{ij} q^{ij}}{\omega} \right) W^i \right], \tag{3}$$

results in an optimal quantity produced:

$$\frac{\partial \pi^{ij}}{\partial q^{ij}} = 0$$
$$\Leftrightarrow$$
$$\frac{\sigma - 1}{\sigma} \left(q^{ij}\right)^{-\frac{1}{\sigma}} \left(Q^j\right)^{\frac{1}{\sigma}} P^j = \frac{\tau^{ij} W^i}{\omega}$$
$$\Leftrightarrow$$
$$q^{ij} = \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega} \right)^{-\sigma} Q^j \left(P^j\right)^{\sigma}. \tag{4}$$

and an equilibrium price as a constant markup over marginal costs $p^{ij} = \frac{\sigma}{\sigma-1} \frac{\tau^{ij} W^i}{\omega}$:

$$\left(\frac{q^{ij}}{(Q^j)} \right)^{\frac{-1}{\sigma}} P^j = p^{ij}$$
$$p^{ij} = \frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega}. \tag{5}$$

The realized revenue expression for firms from country i selling in destination j at time t can then be expressed as:

$$x^{ij} = p^{ij} q^{ij} = \left(q^{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(Q^j\right)^{\frac{1}{\sigma}} P^j$$
$$= \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega} \right)^{1-\sigma} Q^j \left(P^j\right)^{\sigma} \tag{6}$$

B.2 Operating decisions

In line with (Dixit and Stiglitz, 1977; Krugman, 1980; di Giovanni and Levchenko, 2012), I assume that the marginal firm ignores the impact of its own production level on the aggregate economy. The necessary productivity levels for serving each market are then determined by the zero cutoff profit conditions.

$$
\begin{aligned}
\pi^{ij} = 0 &= p^{ij} q^{ij} - \left(f^{ij} - \frac{\tau^{ij} q^{ij}}{\omega^{ij*}} \right) W^i, \\
&= \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma} - f^{ij} W^i - \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{-\sigma} Q^j \left(P^j \right)^{\sigma} \frac{\tau^{ij}}{\omega^{ij*}} W^i, \\
&= \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma} - f^{ij} W^i - \left(\frac{\sigma}{\sigma - 1} \right)^{-\sigma} \left(\frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma}, \\
&= \left(1 - \frac{\sigma - 1}{\sigma} \right) \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma} - f^{ij} W^i, \\
&\Leftrightarrow \\
\sigma f^{ij} W^i &= \left(\frac{\sigma}{\sigma - 1} \frac{\tau^{ij} W^i}{\omega^{ij*}} \right)^{1-\sigma} Q^j \left(P^j \right)^{\sigma}. \qquad (7)
\end{aligned}
$$

Combining the zero cutoff profit conditions allows us to write the export cutoff as a function of a foreign domestic productivity cutoff, variable and fixed costs and the wages:

$$
\omega^{ij*} = \left(\frac{W^i}{W^j} \right)^{\frac{\sigma}{\sigma-1}} \left(\frac{f^{ij}}{f^{jj}} \right)^{\frac{1}{\sigma-1}} \left(\frac{\tau^{ij}}{\tau^{jj}} \right) \omega^{jj*}. \qquad (8)
$$

Similarly, the zero cutoff profit conditions from a single origin country can be combined, linking the domestic and export productivity cutoffs:

$$
\omega^{ij*} = \frac{\tau^{ij}}{\tau^{ii}} \left(\frac{P^j}{P^i} \right)^{\frac{\sigma}{1-\sigma}} \left(\frac{Q^i}{Q^j} \frac{f^{ij}}{f^{ii}} \right)^{\frac{1}{\sigma-1}} \omega^{ii*}. \qquad (9)
$$

This paper focuses on parameter values such that there is, in line with empirical evidence, selection into exporting ($\omega^{ij*} > \omega^{ii*}$). This implies

- A large fixed cost of exporting relative to the fixed cost of production. The revenue required to cover the fixed export cost is then large relative to the revenue required to cover the fixed production cost, implying that only firms of high productivity find it profitable to serve both markets.
- A high home price index relative to the foreign price index, and a large home market relative to the foreign market. Only high productivity firms receive enough revenue in the relatively small and competitive foreign market to cover the fixed cost of exporting.
- Variable trade costs increase the exporting productivity cutoff relative to the zero-profit productivity cutoff by increasing prices and reducing revenue in the export market.

The equilibrium value of these cutoffs are uniquely determined by the free entry condition, requiring the probability of successful entry times the expected future value of entry conditional upon successful entry to equal the sunk entry cost:

$$\sum_{j=1}^{J} \mathbb{E}\left[\pi^{ij} | \omega > \omega^{ij*}\right] = f^{ie} W^i$$

$$\sum_{j=1}^{J} \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \pi^{ij} = f^{ie} W^i$$

$$\sum_{j=1}^{J} \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \left[\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij} W^i}{\omega}\right)^{1-\sigma} Q^j \left(P^j\right)^{\sigma} - f^{ij} W^i\right] = f^{ie} W^i$$

$$\sum_{j=1}^{J} f^{ij} W^i \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \left[\left(\frac{\omega}{\omega^{ij*}}\right)^{\sigma-1} - 1\right] = f^{ie} W^i$$

$$\sum_{j=1}^{J} f^{ij} \left[\left(\omega^{ij*}\right)^{1-\sigma} \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1} - \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{0}\right] = f^{ie}$$

$$\sum_{j=1}^{J} f^{ij} \left[\left(\omega^{ij*}\right)^{1-\sigma} m_{\omega^{ij*}}^{\sigma-1} - m_{\omega^{ij*}}^{0}\right] = f^{ie}, \tag{10}$$

where I denote by m_y^r the nonparametric y-bounded, r-th sample moment of the productivity distribution:

$$m_y^r = \frac{1}{B} \sum_{n=1}^{B} \mathbb{I}_{\omega_b \geq y} \omega_b^r, \qquad \text{for } r \in \mathbb{R}^+. \tag{11}$$

If one would assume an infinite rather than finite number of firms, one can rely on the law of large numbers to replace the specified sample moments with their population equivalent μ_y^r, the nonparametric y-bounded, r-th population moment of the productivity distribution:

$$\mu_y^r = \int_y^{\infty} \omega^r g(\omega) d\omega, \qquad \text{for } r \in \mathbb{R}^+, \tag{12}$$

where $g(\omega)$ represents the Probability Density Function (PDF) related to $G(\omega)$. Under the continuum of firms assumption, the free-entry equation 6 is equivalent to the continuous free-entry equation as specified by Melitz (2003).

Using the relation between productivity cutoffs (eq. 8), the free entry condition (eq. 10) determines a unique equilibrium values of these cutoffs.[3] Thus, a parametrization of the Melitz (2003)-model in relation to firm heterogeneity relies solely on the bounded (by the respective productivity cutoffs) 0th and $(\sigma-1)$th moments of the productivity distribution (Nigai, 2017).

[3] Sufficient conditions for this equilibrium to exist are that the term in brackets of equation (10) is (i) finite and (ii) a decreasing function of the cutoffs (Melitz, 2003, p.1704). The second condition corresponds to $\frac{g(x)x}{1-G(x)}$ increasing to infinity on $(0, \infty)$.

B.3 Aggregation

Summing equation 6 across all active firms, we obtain an expression for aggregate trade between country i and j:

$$X^{ij} = \left(\frac{\sigma}{\sigma - 1} \tau^{ij} W^i \right)^{1-\sigma} Q^j \left(P^j\right)^{\sigma} M^{ie} m^{\sigma-1}_{\omega^{ij*}} \tag{13}$$

The partial sensitivity of aggregate trade to changes in variable trade costs, the aggregate trade elasticity, is then defined as (Chaney, 2008; Arkolakis et al., 2012; Melitz and Redding, 2014; Bas et al., 2017):[4]

$$\begin{aligned}
\gamma^{ij} = \left.\frac{\partial ln X^{ij}}{\partial ln \tau^{ij}}\right|_{\omega^{ii*}} &= 1 - \sigma + \frac{\Delta ln \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1}}{\Delta ln \omega^{ij*}} \left.\frac{\partial ln \omega^{ij*}}{\partial ln \tau^{ij}}\right|_{\omega^{ii*}} \\
&= 1 - \sigma + \frac{\Delta \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1}}{\Delta \omega^{ij*}} \frac{\omega^{ij*}}{\frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1}} \left.\frac{\partial ln \omega^{ij*}}{\partial ln \tau^{ij}}\right|_{\omega^{ii*}} \\
&= 1 - \sigma - \frac{(\omega^{ij*})^{\sigma} g(\omega^{ij*})}{\frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\omega > \omega^{ij*}\right) \omega^{\sigma-1}},
\end{aligned} \tag{14}$$

by the fundamental theorem of discrete calculus.

The number of successful entrants $\left[1 - G(\omega^{ii*})\right] M^{ie}$ is specified as the ratio of aggregate over average revenue:

$$M^i = \left[1 - G(\omega^{ii*})\right] M^{ie} = \frac{X^i}{\mathbb{E}\left[x^i\right]}. \tag{15}$$

This number of firms can be rewritten, using the free entry condition, goods and labor market clearing $\left(X^i = W^i L^i\right)$, as a function of exogenous variables:

$$\begin{aligned}
M^i &= \frac{W^i L^i}{\sigma \left(\frac{f^{ie}}{1 - G(\omega^{ii*})} + \sum_{j=1}^{J} \frac{1 - G(\omega^{ij*})}{1 - G(\omega^{ii*})} f^{ij} \right) W^i} \\
&= \frac{L^i}{\sigma \left(\frac{f^{ie}}{1 - G(\omega^{ii*})} + \sum_{j=1}^{J} \frac{1 - G(\omega^{ij*})}{1 - G(\omega^{ii*})} f^{ij} \right)}.
\end{aligned} \tag{16}$$

Assuming a two-country symmetric economy and setting the wage of the composite factor as the numeraire, welfare can be calculated as the inverse of the price index

[4] Aggregate trade elasticity is defined here as the direct response of aggregate trade to trade costs, keeping the indirect effect trough the price index via its impact on the domestic cutoff fixed (Melitz and Redding, 2015).

$$\mathbb{W}^i = (P^i)^{-1}. \tag{17}$$

The price index can be deduced from equation 13:

$$P^j = \left[\left(\frac{\sigma}{\sigma - 1} \tau^{ij} W^i \right)^{1-\sigma} \frac{1}{\lambda_{ij}} \frac{M^i}{1 - G(\omega^{ii*})} m^{\sigma-1}_{\omega^{ij*}} \right]^{\frac{1}{1-\sigma}}, \tag{18}$$

where the bilateral trade share is denoted by $\lambda^{ij} = \frac{X^{ij}}{X^j}$.

The percentage changes in welfare from a change in variable trade costs ($\boldsymbol{\tau} \rightarrow \boldsymbol{\tau}'$) can then written as:

$$\begin{aligned}
100 \times ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} &= 100 \times -ln\frac{(P^i)'}{P^i} \\
&= 100 \times -ln\frac{(P^j)'}{P^j} \\
&= 100 \times - \left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma - 1} \left(ln\frac{(M^i)'}{M^i} - ln\frac{1 - G(\omega^{ii*})'}{1 - G(\omega^{ii*})} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}} \right) \right] \\
&= 100 \times - \left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma - 1} \left(ln\frac{(M^i)'}{M^i} - ln\frac{(m^{0}_{\omega^{ij*}})'}{m^{0}_{\omega^{ij*}}} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}} \right) \right].
\end{aligned} \tag{19}$$

B.4 Parametrization

In order to parametrize the previously described model, two statistics related to the productivity distribution need to be parametrized: the 0th and $(\sigma-1)$the y-bounded moments of the *productivity* distribution (Nigai, 2017). As described in section 3.1, this corresponds to the 0th and 1st y-bounded moments of the *sales* distribution if the parametric distribution is stable under power-law transformations.

Assuming a parametric distribution and under the assumption of an *infinite* number of firms, the necessary analytical expressions can be calculated using the distributional parameters from the empirical analysis to capture heterogeneity. This is the standard approach in the literature. As discussed in the main text, one can also capture heterogeneity directly from the empirical, *finite*, data. To allow comparison between GFT obtained assuming a parametric distribution and GFT obtained from the finite data, a parametric bootstrap is performed. This parametric bootstrap generates a range of finite sample estimates under the hypothesis that the observed data is generated by a certain parametric distribution, equivalent to the bootstrap procedure discussed in the main text.

B.4.1 Continuum of firms

When there are an infinite number of firms, the parametrization of the heterogeneity distribution consists of calculating the y-bounded 0th and 1st population moments of the sales distribution:

$$\mu_y^r = \int_y^\infty x^r g(x) dx. \tag{20}$$

The analytical expressions of these parametric implied population moments are gathered in Table 3 for all distributions considered.

B.4.2 Finite number of firms

Under the assumption of a finite number of firms in the economy, the parametrization of the model consists of calculating the y-bounded 0th and 1st moment of the sales distribution:

$$m_y^r = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(x > y\right) x^r. \tag{21}$$

These moments can easily be retrieved if the data is available. To allow comparison between GFT obtained assuming a parametric distribution and GFT obtained from the finite data, a parametric bootstrap is performed. This parametric bootstrap generates a range of finite sample estimates under the hypothesis that the observed data is generated by a certain parametric distribution:

1. Assume B i.i.d. random variables with distribution $G(\cdot|\boldsymbol{\theta})$, with empirical finite sample moments $m_{\tilde{y}}^{\tilde{r}}(x)$ for $r = 0, 1$, as specified in equation 21 and corresponding $GFT(x)$;

2. Estimate the parameters $\boldsymbol{\theta}$ of the distribution using MLE, calculate the parametric plug-in population moments as specified in equation 20, $\hat{\mu}_{\tilde{y}}^{\tilde{r}}(x|\hat{\boldsymbol{\theta}})$ for $r = 0, 1$, and corresponding $GFT(x|\hat{\boldsymbol{\theta}})$;

3. $H_0 : GFT(x) = GFT(x|\hat{\boldsymbol{\theta}})$;

4. Draw N bootstrap samples of size B from $G(\cdot|\hat{\boldsymbol{\theta}})$;

5. For each sample of the parametric distribution, calculate the bootstrapped sample moments $m_{\tilde{y}}^{\tilde{r}}(x^*)$ and calculate the corresponding $GFT(x^*)$.[5]

6. The p-value for the left-, and right-tailed test is then respectively specified as:

$$\hat{p}_l = \frac{1}{N+1} \left[\sum_{n=1}^{N} \mathbb{I}\left(GFT(x^*) \geq GFT(x)\right) + 1 \right]; \qquad \hat{p}_r = \frac{1}{N+1} \left[\sum_{n=1}^{N} \mathbb{I}\left(GFT(x^*) \leq GFT(x)\right) + 1 \right]. \tag{22}$$

[5] Note that I do not re-fit the parametric distribution to the bootstrap sample. The vastness of the dataset available in the empirical section results both in a large computational burden but also a very precise estimation of the distribution parameters. The influence of not refitting the parametric distribution to the bootstrap sample is therefore negligent.

The bootstrap exercise should therefore be interpreted as 'the likelihood of observing GFT as small or as large as GFT_B under the null hypothesis that the observed data originates from the parametric distribution $G(\cdot|\boldsymbol{\theta})$', allowing us to evaluate whether the distributional assumption provides a good fit to calculate GFT within the proposed model.

When calculating the bounded sample moments, complications can arise related to the lower bound y. This lower bound is ex ante unknown, can take values not observed in the data and/or resides in an unrepresentative part of the finite dataset.[6] Each issue is addressed below separately, arguing that these complications have little influence on the results.

1. y can take values within the boundaries of the data but are not observed. The 'approxfun' interpolation function of the R base distribution has been used to approximate the statistics for such lower bounds.[7] As the calculation of Gains From Trade (GFT) relies on domestic cutoffs residing in the dense part of the productivity distribution, the influence of interpolation is negligible.

2. y can take values below the lowest observed value in the data ($y < x_{min}$):

$$\mu_y^r = \underbrace{\sum \mathbb{I}\left(y < x < x_{min}\right) x^r}_{\text{unobserved}} + \underbrace{\frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(x \geq x_{min}\right) x^r}_{observed} . \tag{23}$$

   The error arising from neglecting the unobserved part of the distribution is likely small as (i) the smallest observation x_{min} in the data set is rather small, (ii) the density in the unobserved part is most likely very low and (iii) the relative weight of the observations in the unobserved part is small (see also Figure 1).

3. As the presented model is a stylized model, it is conceivable firms produce below the model's implied zero-profit productivity cutoff, for instance when there is a positive expectation of future profits (**?**). This can explain very low observed productivity values, but will result in an unrepresentative left tail of the distribution (the lower the actual zero-profit productivity cutoff, the more firms will have a positive expectation of future profits and the denser the left tail of the distribution will be). This issue affects both the nonparametric and parametric estimates, as the parametric distribution is fitted to the observed distribution. Also in this case, however, provided the low density in the left tail of the distribution and the low relative weight of the observations in the left tail, the influence of this issue is likely small.

[6]I am grateful to Gonzague Vannoorenberghe for pointing this out.
[7]All code available on request.

Appendix C Composite distribution

Piecewise or composite distributions are distributions obtained from piecing together weighted truncated distributions at a specified threshold:

$$g(\omega) = \begin{cases} \alpha_1 m_1^*(\omega) & \text{if} \quad c_0 < \omega < c_1 \\ \alpha_2 m_2^*(\omega) & \text{if} \quad c_1 < \omega < c_2 \\ \quad\vdots & \qquad\qquad \vdots \\ \alpha_i m_i^*(\omega) & \text{if} \quad c_{i-1} < \omega < c_i \end{cases} \tag{24}$$

where $\forall i \in I : m_i^*(\omega) = \frac{m_i(\omega)}{\int_{c_{i-1}}^{c_i} m_i(\omega) d\omega}$ is the truncated probability density function (PDF) of $m_i(\omega)$. For this distribution to be well-behaved, additional differentiability and continuity conditions are imposed on the truncation points (Bakar et al., 2015). The flexibility allows for many parametric piecewise contributions (see (Bakar et al., 2015) for an overview). Provided the focus of the firm size literature on the Pareto-tail,[8] we are mainly interested in Pareto-tailed piecewise distributions. I have knowledge of the Inv. Pareto - Lognormal - Pareto distribution and Lognormal-Pareto distributions being applied in the city size literature (Ioannides and Skouras, 2013; Luckstead and Devadoss, 2017), while Nigai (2017) proposes to rely on this Lognormal-Pareto distribution when parameterizing the Melitz (2003) model for GFT calculations.

We generalize the definition of the three-composite distribution to allow for multiple possible parametrizations.[9] The probability density function is specified as:

$$g(\omega) = \begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2} \frac{m_1(\omega)}{M_1(c_1)} & \text{if} \quad 0 < \omega \le c_1 \\ \frac{1}{1+\alpha_1+\alpha_2} \frac{m_2(\omega)}{M_2(c_2)-M_2(c_1)} & \text{if} \quad c_1 < \omega \le c_2 \\ \frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{m_3(\omega)}{1-M_3(c_2)} & \text{if} \quad c_2 < \omega < \infty, \end{cases} \tag{25}$$

with the complementary Cumulative Distribution Function:

$$G(\omega) = \begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2} \frac{M_1(\omega)}{M_1(c_1)} & \text{if} \quad 0 < \omega \le c_1 \\ \frac{\alpha_1}{1+\alpha_1+\alpha_2} + \frac{1}{1+\alpha_1+\alpha_2} \frac{M_2(\omega)-M_2(c_1)}{M_2(c_2)-M_2(c_1)} & \text{if} \quad c_1 < \omega \le c_2 \\ \frac{1+\alpha_1}{1+\alpha_1+\alpha_2} + \frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{M_3(\omega)-M_3(c_2)}{1-M_3(c_2)} & \text{if} \quad c_2 < \omega < \infty. \end{cases} \tag{26}$$

The Quantile function can then be specified as:

[8] Note that a disadvantage of composite distributions is the ambiguity surrounding their generative process. For instance, it is yet unclear which firm dynamics could explain the existence of hard cutoffs that separate the main component distribution from a Pareto tail.

[9] The complementary code is implemented in the R-package 'distributionsrd' under the 'composite' commands.

$$Q(p)=\begin{cases} Q_1\left(\frac{1+\alpha_1+\alpha_2}{\alpha_1}pM_1(c_1)\right) & \text{if} \quad 0<\omega\leq\frac{\alpha_1}{1+\alpha_1+\alpha_2} \\ Q_2\left[\left(\left(p-\frac{\alpha_1}{1+\alpha_1+\alpha_2}\right)(1+\alpha_1+\alpha_2)\left(M_2(c_2)-M_2(c_1)\right)\right)+M_2(c_1)\right] & \text{if} \quad \frac{\alpha_1}{1+\alpha_1+\alpha_2}<\omega\leq\frac{1+\alpha_1}{1+\alpha_1+\alpha_2} \\ Q_3\left[\left(\left(p-\frac{1+\alpha_1}{1+\alpha_1+\alpha_2}\right)\left(\frac{1+\alpha_1+\alpha_2}{\alpha_2}\right)(1-M_3(c_2))\right)+M_3(c_2)\right] & \text{if} \quad \frac{1+\alpha_1}{1+\alpha_1+\alpha_2}<\omega<\infty. \end{cases} \tag{27}$$

Lastly, the y-bounded r-th raw moment of the distribution equals:

$$\begin{aligned} \mu_y^r &= \int_y^\infty \omega^r g(\omega)d\omega \qquad (28) \\ &= \begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2}\frac{(\mu_1)_y^r-(\mu_1)_{c_1}^r}{M_1(c_1)}+\frac{1}{1+\alpha_1+\alpha_2}\frac{(\mu_2)_{c_1}^r-(\mu_2)_{c_2}^r}{M_2(c_2)-M_2(c_1)}+\frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_y^r}{1-M_3(c_2)} & \text{if} \quad 0<y\leq c_1 \\ \frac{1}{1+\alpha_1+\alpha_2}\frac{(\mu_2)_y^r-(\mu_2)_{c_2}^r}{M_2(c_2)-M_2(c_1)}+\frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_{c_2}^r}{1-M_3(c_2)} & \text{if} \quad c_1<y\leq c_2 \\ \frac{\alpha_2}{1+\alpha_1+\alpha_2}\frac{(\mu_3)_y^r}{1-M_3(c_2)} & \text{if} \quad c_2<y<\infty \end{cases} \end{aligned}$$

Imposing continuity and differentiability allows to determine the distributions weights and cutoffs in function of the distribution parameters. *Continuity* implies that the composite distribution is continuous at the thresholds:

$$\lim_{\omega\to c_1^-} g(\omega)=\lim_{\omega\to c_1^+} g(\omega); \qquad \lim_{\omega\to c_2^-} g(\omega)=\lim_{\omega\to c_2^+} g(\omega), \tag{29}$$

such that

$$\begin{aligned} \frac{\alpha_1}{1+\alpha_1+\alpha_2}m_1^*(c_1) &= \frac{1}{1+\alpha_1+\alpha_2}m_2^*(c_1) \qquad (30) \\ \alpha_1 &= \frac{m_2(c_1)M_1(c_1)}{m_1(c_1)\left[M_2(c_2)-M_2(c_1)\right]}, \end{aligned}$$

$$\begin{aligned} \frac{1}{1+\alpha_1+\alpha_2}m_2^*(c_2) &= \frac{\alpha_2}{1+\alpha_1+\alpha_2}m_3^*(c_2) \qquad (31) \\ \alpha_2 &= \frac{m_2(c_2)\left[1-M_3(c_2)\right]}{m_3(c_2)\left[M_2(c_2)-M_2(c_1)\right]}. \end{aligned}$$

Similarly, impose *differentiability* conditions at the thresholds:

$$\lim_{\omega\to c_1^-}\frac{dg(\omega)}{d\omega}=\lim_{\omega\to c_1^+}\frac{dg(\omega)}{d\omega}, \qquad \lim_{\omega\to c_2^-}\frac{dg(\omega)}{d\omega}=\lim_{\omega\to c_2^+}\frac{dg(\omega)}{d\omega}, \tag{32}$$

so that

$$\frac{d}{dc_1} ln \left[\frac{m_1(c_1)}{m_2(c_1)} \right] = 0, \qquad \frac{d}{dc_2} ln \left[\frac{m_2(c_2)}{m_3(c_2)} \right] = 0. \tag{33}$$

These differentiability conditions can be analytically calculated on a case-wise basis or can easily be obtained numerically.

Appendix D Replication Nigai (2017)

In this section, I compare the results obtained in this paper with those reported by Nigai (2017). I start by discussing the nonparametric estimator proposed by Nigai (2017) to capture the 'true' heterogeneity distribution. A narrow replication of Nigai (2017) is performed to ensure comparability before engaging in a wide replication exercise, applying the methodology proposed in this paper to the international trade model of Nigai (2017). Contrary to Nigai (2017), this replication finds non-negligible errors in the quantitative analysis: the analysis rejects the premise that GFT derived from the data can be obtained from any of the proposed distributions.[10]

D.1 Nonparametric approach

Nigai (2017) takes a different approach than this paper to nonparametrically capture firm-level heterogeneity. Rather than aiming to acquire the heterogeneity implied by the finite nr. of firms present in sales data $(m^0_{\tilde{y}}, m^1_{\tilde{y}})$, the author aims to nonparametrically capture heterogeneity in productivity implied by an infinite number of firms $(\mu^0_y, \mu^{\sigma-1}_y)$.

For the zeroth moments, the approach of Nigai (2017) coincides with the approach taken in this paper, recovering the moments from the empirical CDF of the productivity distribution. As discussed in section 3.3, this is a feasible strategy provided the efficiency and consistency of this estimator.

Considering the $(\sigma - 1)$th moment, this is calculated by Nigai (2017) in two steps. First, the author approaches the PDF using a normalized histogram with 1,000 bins. In a second stage, the $(\sigma-1)$th moment is calculated using trapezoidal numerical integration. As the asymptotic behavior of this estimator was not discussed, we analyze this behavior with a parametric bootstrap, similar to the main analysis of the paper.

I calculate the bootstrapped distribution of the unbounded nonparametric $(\sigma - 1)$th moment of productivity calculated on draws from each respective parametric distribution. The box plots delineate the 5th, 25th, 50th, 75th and 95th quantile of the limiting distribution of unbounded parametric sample $(\sigma - 1)$th estimates. Green circles indicate the average of all these parametric finite sample estimates, while the parametric plug-in population estimates, if finite, are indicated

[10]Note that a complete replication of Nigai (2017) is not possible as the dataset differs from the one reported by Nigai (2017). Similar to Nigai (2017), this paper relies on the Orbis database and downloaded all, both total and exporting, sales data from firms that reported sales in 2012 (excluding companies with no recent financial data and Public authorities/States/Governments) by means of unconsolidated accounts and that have a designated 'corporate' entity type. The data was obtained from the Orbis Europe database update 05/06/2020 (nr. 197002) and software version 197. This search strategy resulted in domestic sales data on 500,388 French firms, in comparison to the 928,569 observations reported by Nigai (2017). After conferring with the author, this difference can most likely be attributed to the fact that the author relies on the the last sales data of firms present in the database in 2012, rather than reported sales in 2012. Conflating firm-level sales of different years as such could have introduced additional, undesired, heterogeneity in the sample. Nevertheless, fitted distributions report similar parameter estimates. Nigai (2017) reports a Lognormal fit with mean -0.701 and standard deviation 0.569 relative to the Lognormal fit obtained in this paper with mean -0.724 and standard deviation 0.611. The shape parameter of the Lognormal-Pareto distribution is measured at 3.033 by Nigai (2017) while it amounts to 3.12 in this paper. The analysis therefore relies on parameter estimates as reported by Nigai (2017) for this replication exercise.

by yellow diamonds. Figure 2, similar to Figure 3, provides us with a visual perspective on the asymptotic behavior of the proposed nonparametric estimator for the unbounded $(\sigma-1)$the moment of productivity. An efficient and consistent estimator should have sample estimates that are closely centered around the parametric population estimates. This does not seem to be the case. Just as the nonparametric finite sample estimator proposed in this paper, the estimator appears to be biased. Rather than underestimating the 'true' moment, however, this estimator overestimates it. This phenomenon can most probably be attributed to trapezoidal numerical integration overestimating the area under the convex right tail of a right-skewed distribution.

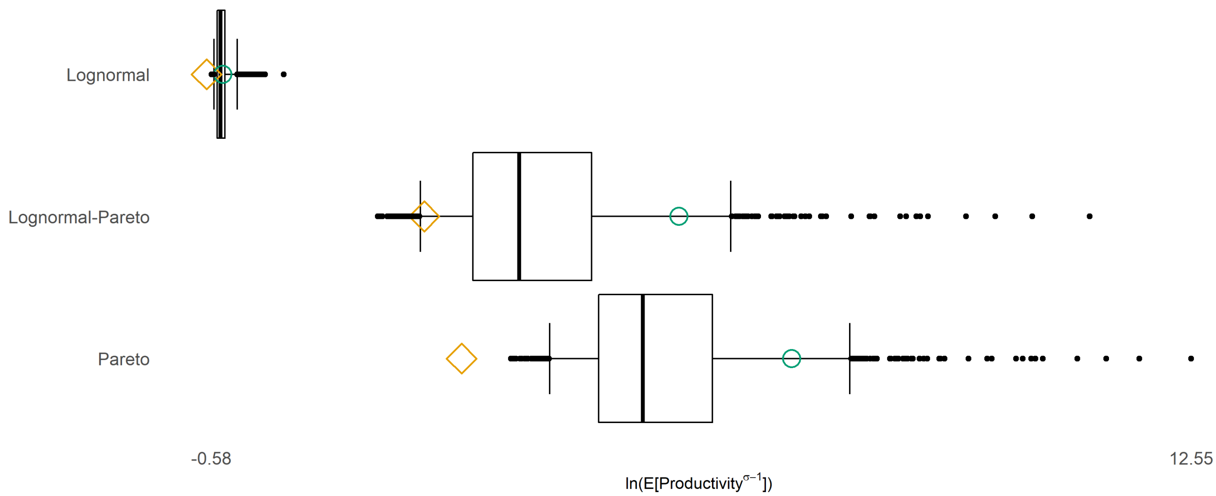


Figure 2: Bootstrapped (1,000 reps) limiting distributions of the unbounded parametric $(\sigma-1)$th moment of productivity.

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the limiting distribution of unboundedm parametric sample $(\sigma-1)$th moments of productivity. Green circles indicate the average of all these parametric sample estimates. The parametric population estimates, if finite, are indicated by yellow diamonds. All bootstrap sample values obtained from a sample of 928,569 firms.

D.2 Narrow replication

The narrow replication exercise focuses on Experiment 1 of Nigai (2017), in which the author calculates welfare gains for from falling variable costs between a large and small economy while fixed export costs are low. The analysis starts by calculating the parametric population GFT for the Lognormal, Pareto and Lognormal-pareto distribution with parameters as reported by Nigai (2017). These estimates are displayed in Figure 3, with the yellow triangles and circles representing GFT for the large and small economy respectively.[11]

[11]Note that the reported GFT for the Pareto distribution differ from those reported in Nigai (2017). This is most likely due to a bug in the code when calculating the moments of the Pareto distribution. The y-bounded rth moment of the distribution equals $\left(\mu_y^r = -(y)^{r-k}\,\frac{k\omega_{min}^k}{r-k}\right)$. If y takes an (according the distribution impossible) value lower than ω_{min}, this statistic becomes values larger than the parametrically implied maximum $\left(max(\mu_y^r) = -(\omega_{min})^{r-k}\,\frac{k\omega_{min}^k}{r-k}\right)$

The proposed nonparametric estimators of GFT is parametrically bootstrapped. The resulting limiting distributions are represented by the box plots in Figure 3. These figures delineate the 5th, 25th, 50th, 75th and 95th quantile of the limiting distribution of nonparametric finite sample GFT. In line with the results found in the previous section regarding the consistency of the proposed nonparametric estimator, the nonparametric GFT are biased relative to the parametrically implied population estimates. Moreover, this bias seems to depend on the underlying parametric distributional form.

These results lead us to conclude that the strategy of the author to rely on the nonparametric estimator as a direct benchmark when evaluating the performance of parametric distributions is erroneous.

Figure 3: Nonparametric bootstrapped (100 reps) limiting distributions of the parametric sample gains from a reduction in variable trade costs $\tau = 3 \rightarrow \tau' = 1.2$.

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the asymptotic distribution of nonparametric GFT. Yellow diamonds represent the plug-in (population) estimates of GFT. Green circles are the average nonparametric bootstrapped finite sample GFT. All sample values obtained from a sample of 928,569 firms.

D.3 Wide replication

As a wide replication exercise, GFT obtained in this paper are evaluated to a heterogeneous firms model specification with asymmetric country size as in Nigai (2017). Also here I focus on Experiment 1, of which the results are displayed in Figure 4. We observe that the results are indeed robust. Contrary to Nigai (2017), this replication finds non-negligible errors in the quantitative

when not correctly restricted. This would also explain why reported GFT by Nigai (2017) for the Pareto distribution are lower than those obtained from the Lognormal distribution, despite Pareto being heavier-tailed according to Figure 3 of Nigai (2017). In this narrow replication, GFT from the Pareto distribution are as expected larger than those obtained from the Lognormal distribution.

analysis: the analysis rejects the premise that GFT derived from the data can be obtained from any of the proposed distributions.

This conclusion contradicts the intuition provided by the author (and commonly relied on in the trade literature), that a good fit to the CDF (Figure 3 in Nigai (2017)) can be directly linked to model performance. As we demonstrate in the main analysis, however, a good fit to the CDF is a necessary but not sufficient condition. A valid parametric approximation also needs to match a higher (the $\sigma - 1$th) moment of the productivity distribution. Only when this parametric approximation provides a *sufficiently* good fit to the CDF (in the Kolmogorov-Smirnov distance sense), it can be claimed there is convergence in distribution and as a result also convergence in higher moments. When this is not the case, higher moments can largely deviate from the data despite a seemingly good CDF fit.

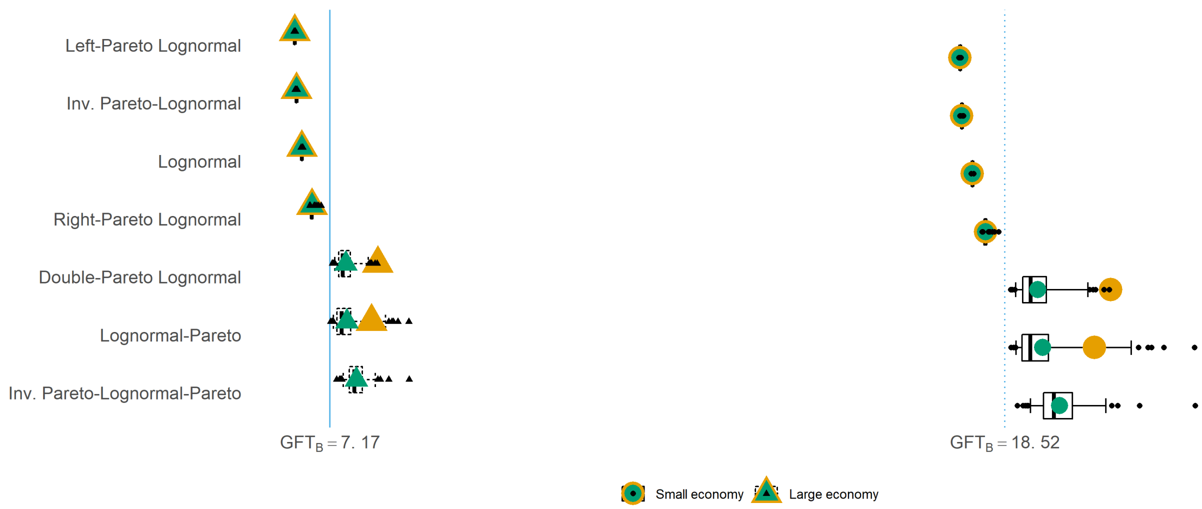


Figure 4: Bootstrapped (100 reps) limiting distributions of the parametric sample gains from a reduction in variable trade costs $\tau = 3 \rightarrow \tau' = 1.2$.

Notes: Box-plots display the 5th, 25th, 50th, 75th and 95th quantile of the asymptotic distribution of parametric finite sample GFT. Yellow diamonds represent the plug-in (population) estimates of GFT. Green circles are the average parametric bootstrapped finite sample GFT and the empirical sample GFT are indicated by the vertical blue line. All sample values obtained from a sample of 500,388 firms.

Appendix References

di Giovanni, J. and A. A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy 120*(6), 1083–1132.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American Economic Review 67*(3), 297–308.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review 70*(5), 950–959.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

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