Unobserved Heterogeneity in the Productivity Distribution and Gains From Trade.

Ruben Dewitte, Michel Dumont, Glenn Rayp, Peter Willemé*

19th December 2022

Abstract

Finding a good parametric approximation to the productivity distribution is a problem of general interest. This paper argues that heterogeneity in productivity is best captured by Finite Mixture Models (FMMs). FMMs build on the existence of unobserved subpopulations in the data. As such, they are generally consistent with models of firm dynamics differing between groups of firms and allow for a very flexible distribution fit. Relative to commonly used parametric alternatives, we find that FMMs are the only distributions able to provide a sufficiently good fit to the data. A Gains From Trade exercise with Portuguese data reveals that only FMMs approximate the 'true' gains reasonably well.

Keywords: Finite Mixture Model, firm size distribution, productivity distribution, Gains From Trade

JEL Codes: L11, F11, F12

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Dumont: Federal Planning Bureau and Ghent University, Rue Belliard - Belliardstraat 14-18 1040 Brussels, dm@plan.be , Rayp: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, glenn.rayp@ugent.be, Willemé: Federal Planning Bureau and Ghent University, Rue Belliard - Belliardstraat 14-18 1040 Brussels, pw@plan.be. We thank Pete Klenow, Sergey Nigai, Luca David Opromolla, Gonzague Vannoorenberghe, Olivier Thas, Gerdie Everaert, Bruno Merlevede, Marijn Verschelde, and participants at the 2018 ETSG conference in Warsaw, the 2019 ETSG conference in Bern, the DEGIT XIV conference in Odense, and the 2019 workshop on "Firm Heterogeneity in Technical Change" in Ghent for their helpful comments and suggestions. Ruben Dewitte acknowledges financial support from Research Foundation Flanders (FWO) grant 12B8822N.

1 Introduction

Parametric approximations of the productivity distribution are of key importance to various research topics in economics. The mechanisms driving firm-level dynamics in aggregate growth models, for instance, are determined by the parametric approximation of the productivity distribution (see, for instance, Luttmer (2007); Arkolakis (2016)). Also, the propagation of firm-level volatility to the aggregate level mainly relies on a Pareto approximation of the right tail of the productivity distribution (Gabaix, 2011; di Giovanni et al., 2011; Carvalho and Grassi, 2019). In the international trade literature, it is recognized that different choices for the productivity distribution significantly affect Gains From Trade (GFT) estimates (Head et al., 2014; Nigai, 2017; Bee and Schiavo, 2018), and alters the channels through which trade affects welfare (Arkolakis et al., 2012; Bas et al., 2017; Melitz and Redding, 2015; Fernandes et al., 2018).

To date, however, there is no consensus on what this parametric approximation should be. Some authors argue a single distributional form such as Pareto (Axtell, 2001), Lognormal (Head et al., 2014) or Weibull (Bee and Schiavo, 2018) suffices to define the productivity distribution. Others build on the idea that a single distribution can not adequately capture the heterogeneity in productivity. This results in combinations of distributions such as the Double-Pareto (Arkolakis, 2016), Double-Pareto Lognormal (Sager and Timoshenko, 2019) or Lognormal-Pareto (Nigai, 2017). Nevertheless, Dewitte (2020) demonstrates that none of the distributions that are currently considered can provide a sufficiently good fit to the data.

This paper argues that heterogeneity in the distribution of firm productivity can be captured most adequately by Finite Mixture Models (FMMs). A FMM is a weighted sum of an a priori unknown number of individual densities. As such, it is a semi-parametric approximation that allows for discrete subpopulations to define the overall distribution. The flexible, semi-parametric nature of FMMs has advantages both from a theoretical and empirical perspective.

From a theoretical point of view, the generative process of a FMM corresponds to a simple combination of the generative processes of the underlying individual densities. A FMM can therefore easily generalize, and is generally consistent with, existing models of firm dynamics. First, FMMs allow to combine a specified generative process of firm dynamics across groups of firms to capture additional, unspecified heterogeneity. Luttmer (2007), for instance, generalizes his single-sector model with a finite mixture specification to a multi-sector model, to capture additional heterogeneity across industries and obtain a satisfactory fit to the data. Similarly, Rossi-Hansberg and Wright (2007) argue the need to account for cross-sectoral differences in their initial single-sector model specification to achieve an accurate description of the cross-sectional size distribution of US firms. Second, a finite mixture specification is generally consistent with the mechanisms considered to differentiate firm dynamics between groups of firms. The differences in growth rates between financially constrained and unconstrained firms by Cabral and Mata (2003), for instance, can be respecified into a finite mixture specification.[1] FMMs provide an empirical tool that can account for dynamics to differ between groups of firms without having, but not excluding the possibility, to specify the mechanisms that drive these differences a priori. These mechanisms can be left 'unobserved'.

We illustrate the superior empirical performance of FMMs using a clear statistical framework to differentiate between a large number (up to 52) of economically relevant parametric distributions. These distributions are fitted to domestic sales of the population of active Portuguese firms in 2006.[2,3] A Kolmogorov-Smirnov test reveals that only FMMs provide a distribution fit that is not rejected by the data. Currently considered distributions such as the Lognormal, Lognormal-Pareto, and Double-Pareto Lognormal are found to *underfit* the data. The Akaike and Bayesian Information Criteria (AIC and BIC) show that the performance of FMMs is not the result of over-fitting.

FMMs outperform other distributions in the ability to capture heterogeneity in the data. We demonstrate how the current focus on improving the fit to the left and/or right tail of the data can worsen the fit to the bulk of the data. The semi-parametric nature of FMMs allows them to approximate the *complete* empirical distribution. FMMs accurately capture the bulk of the data in addition to the left and right tail of the distribution. As the Double-Pareto Lognormal and

[1] Additionally, firm dynamics are argued to differ between groups of firms depending on whether or not they are financially constrained (Cooley and Quadrini, 2001; Cabral and Mata, 2003; Desai et al., 2003; Albuquerque and Hopenhayn, 2004; Clementi and Hopenhayn, 2006; Angelini and Generale, 2008), innovate (Costantini and Melitz, 2008; Atkeson and Burstein, 2010), add or drop products (Klette and Kortum, 2004; Lentz and Mortensen, 2008), add or drop management layers (Caliendo and Rossi-Hansberg, 2012; Caliendo et al., 2020), incur specific market penetration costs (Arkolakis, 2016) ...

[2] As is common in the literature, we capture heterogeneity in productivity from firm-level sales (Head et al., 2014; Nigai, 2017; Bee and Schiavo, 2018). See also section 4.

[3] Having access to a representative dataset on the sales distribution allows us to evaluate the performance of parametric distributions on the complete productivity distribution as well as to focus on both the *left* and right tail. Moreover, it insulates us from erroneous conclusions due to truncated or unrepresentative data in the left tail of the distribution (Perline, 2005).

Lognormal-Pareto distributions can be interpreted as constrained general mixture models, their underfitting of the data indicates that the constraints imposed are not warranted from a statistical point of view.

We demonstrate the economic relevance of our findings for Gains From Trade[4] calculations in heterogeneous firms models à la Melitz (2003). We contribute to the literature providing quantitative expressions necessary to calibrate a heterogeneous firms model for all distributions considered and illustrate the straightforward implementation of FMMs into such models. Our calibration exercise reveals that when reducing variable trade costs by two-thirds, only FMMs can track the 'true' GFT (obtained from the empirical distribution) closely while GFT obtained from commonly used parametric alternatives significantly deviate from these 'true' GFT. We demonstrate that the FMM performance is not a trivial implication of an improved fit to the data. Rather, it demonstrates the ability of FMMs to closely approximate the complete empirical distribution, i.e., to capture heterogeneity in the bulk of the distribution in addition to the heterogeneity in the tails. As a result, FMMs are the only distributional forms able to accurately capture the different channels through which trade affects welfare.

The paper is organized as follows. In the following section, we start by linking the large literature on the parametric approximation of size distributions, spanning the fields of efficiency analysis, physics, regional and actuarial science, to the productivity distribution literature. From this overview, it appears that the literature on productivity distributions lacks a clear statistical framework that differentiates between a sufficiently large number of alternative distributions over a representative data range. We establish a methodology that uniformly fits many distributions to complete and truncated data and present evaluation methods to differentiate between these distributions in section 3. Our database on firm sales is discussed in section 4. We provide our empirical results in section 5 and discuss the implications of these results for GFT calculations in section 6. Section 7 concludes.

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

2 Firm size distributions: a literature overview

This section provides an overview of the literature related to firm size/productivity distributions. We discuss why the Pareto distribution can only match the tail of size distributions while single hump-shaped distributions such as the Lognormal or the Weibull distribution can not accurately match the tail and bulk of the distribution simultaneously. Size distributions are therefore best approximated by a combination of distributions, of which we consider three types: mixture, piecewise composite, and multiplicative distributions. We argue that the flexible, semi-parametric nature of FMMs is appealing both from a theoretical and empirical perspective.

2.1 Single distributions

The *Pareto distribution* has been dominating heterogeneous firms models (Melitz, 2003). Even though the Melitz (2003)-model is not restricted to this distributional choice, its empirical performance (see for instance Axtell (2001); Gabaix (2009); Levy (2009); di Giovanni et al. (2011)) and convenience led to a widespread reliance on the Pareto distribution for social welfare and economic policy analysis.[5] The fit of a Pareto distribution is usually evaluated using its Cumulative Distribution Function (CDF), which follows a straight line on a log-log scale with the shape parameter (k) as slope:

$$G_P(x; x_{min}, k) = 1 - \left(\frac{x_{min}}{x}\right)^k, \qquad x \geq x_{min}. \tag{1}$$

Figure 1 compares a fitted Pareto survival function ($\text{CDF}^c = 1 - \text{CDF}$) with the empirical survival function of Portuguese firm-level sales in 2006 on a log-log scale for the complete dataset (upper panel). It is immediately clear that the Pareto distribution is not a good fit for the complete distribution due to the existence of a hump in the middle.[6]

The popularity of the Pareto distribution, however, rests on its ability to provide a close fit

[5] See Arkolakis et al. (2012) for an overview of work relying on the Melitz-Pareto combination.

[6] See also the Probability Density Function (PDF) in Appendix Figure 2.

to lower-truncated[7] data with predominantly large observations.[8] Just as every curved line looks straight when one zooms in close enough, so too does the distribution of firm sales appear to be straight when truncated sufficiently. Both the left (lower left panel) and right tail (lower right panel) exhibit linearity of the CDF and survival functions respectively on a log-log scale, in line with Pareto behavior in the distribution tails.[9] The apparent straight-line behavior of the tails can therefore just as well be approximated by a surprisingly large class of distributions including, but not restricted to, (finite mixtures of) the Exponential, Lognormal, Gamma and Weibull distributions.[10] Proof of which is the performance of the Lognormal distribution in the lower panels of Figure 1.[11]

These alternative *hump-shaped distributions* are claimed to provide a better fit to complete size distributions (see Bee and Schiavo (2018) for the Weibull and Eeckhout (2004, 2009); Head et al. (2014); Fernandes et al. (2018) for the Lognormal distribution). In the firm size literature, this claim is usually supported by comparing their performance with a limited number of alternative distributions, mostly Pareto, using the low-powered R-squared.[12]

Even though homogeneous hump-shaped distributions such as the Lognormal can adequately fit the tail or the bulk of the empirical distribution, they cannot do both simultaneously. This is easily observable from the upper panel of Figure 1 where the single Lognormal distribution, when fitted to the complete size distribution, does not fit the right tail of the complete productivity distribution while matching the bulk rather satisfactorily.

[7] An *upper-truncated* version of the Pareto distribution has also been used to explain the existence of zero trade flows across country pairs (Helpman et al., 2008; Feenstra, 2018) and to demonstrate the relevance of heterogeneous firms models (Melitz and Redding, 2014). A discussion on the economic relevance of, and an extension of the analysis to, upper-truncated distributions falls outside the scope of this paper. The methodology set out in this paper allows to truncate any kind of distribution both from above and/or below (see Online Appendix B).

[8] Note that the influential paper of Axtell (2001) does not rely on truncated data but unintentionally favors the Pareto distribution due to data binning (Virkar and Clauset, 2014) and methodological choices (Clauset et al., 2009; Bottazzi et al., 2015) characteristic of that time.

[9] The Inverse Pareto distribution is specified as

$$G_{IP}(x; x_{max}, k) = 1 - \left(\frac{x_{max}}{x}\right)^{-k}, \qquad x \leq x_{max}.$$

[10] Perline (2005) defines this class of distributions within the Gumbel domain of attraction.

[11] Even though Pareto and Lognormal distributions exhibit qualitatively different behavior in their upper tails, their apparent quantitative similar behavior in the upper tail for Lognormals with large variance is well-documented (Malevergne et al., 2011).

[12] See Clauset et al. (2009) for an explanation as to why the R-squared has low power in a distributional context.

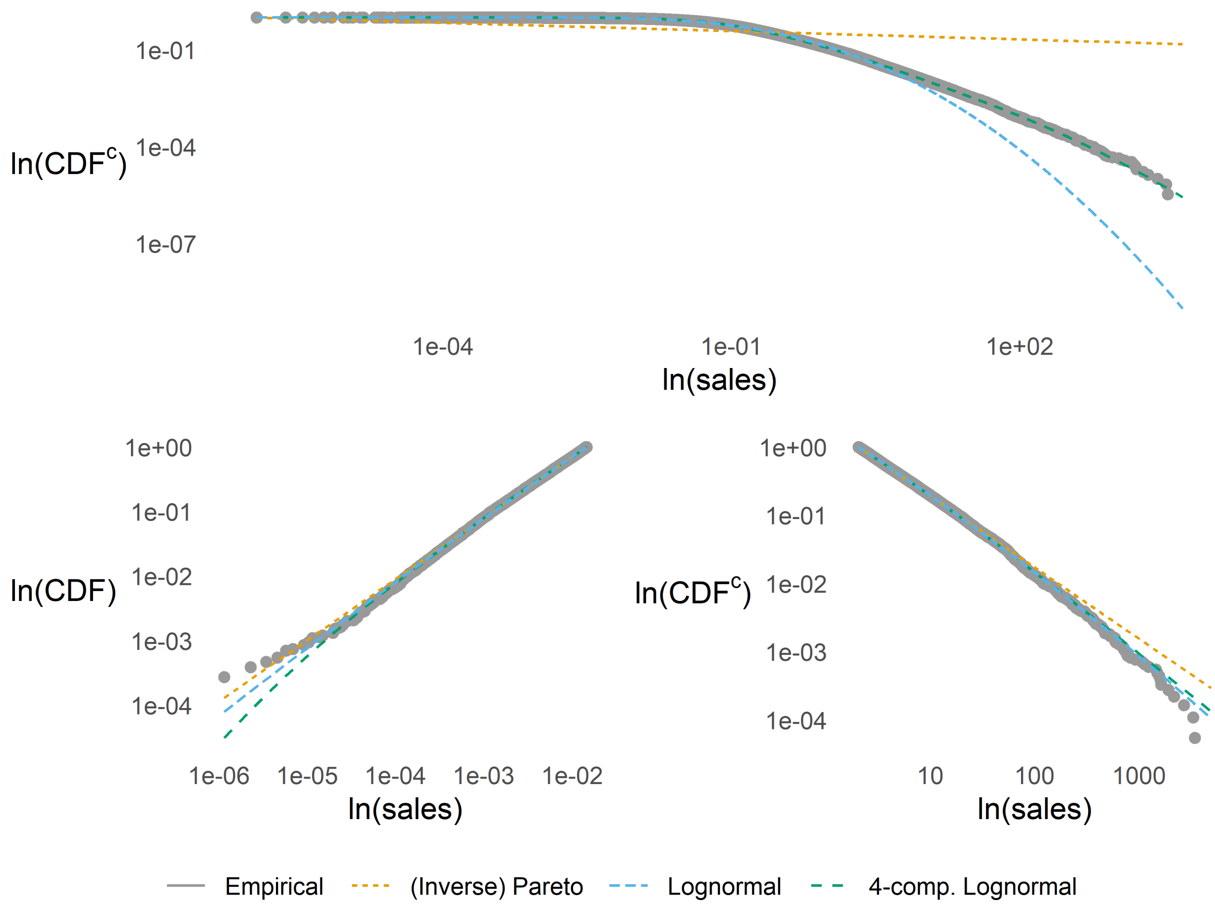

Figure 1: Empirical survival function of Portuguese domestic sales in 2006 (upper panel) on a log-log scale with fitted (Inverse) Pareto and (4-component mixture of) the Lognormal distributions. The lower left and right panels focus on distributions fitted solely to the left and right tail respectively.

Notes: (Truncated) Distributions are fitted using maximum likelihood methods (cf. infra) to the complete and truncated datasets independently. Tail truncation points are determined by the best-fitting (Inverse) Pareto distributions according to the Kolmogorov-Smirnov statistic.

2.2 Combined distributions

As single distributions cannot accurately match both the bulk and the tail(s) of the productivity distribution, recent research focuses on combinations of distributions. We consider three types of combinations: mixture, piecewise composite, and product distributions. To our knowledge, mixture distributions have not been fitted to the productivity distribution. Nevertheless, current applications of both the piecewise composite and product distributions can be interpreted as constraints of the more general mixture specification.

2.2.1 Mixture distributions

Finite Mixture Models (FMMs) are essentially a weighted sum of I individual densities $m_i(\cdot)$:

$$g(x|\boldsymbol{\Psi}) = \sum_{i=1}^{I} \pi_i m_i(x|\boldsymbol{\theta}_i), \qquad \pi_i \geq 0, \quad \sum_{i=1}^{I} \pi_i = 1 \tag{2}$$

where I represents the number of components or discrete subpopulations, π_i is the probability of belonging to component i, $\boldsymbol{\theta_i}$ the component-specific parameter vector of density $m_i(\cdot)$ and $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{I-1}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$ is the vector of all model parameters (McLachlan and Peel, 2000). They are also referred to as Latent Class Models (LCM) provided that the number of components, and thus also the mixing parameter itself, does not have to be specified a priori but is determined by the data. As such, a finite mixture model provides a semi-parametric approach ideal to fully capture the heterogeneity of size distributions.[13]

The aptitude of Finite Mixture models has already been explored in the context of efficiency analysis (see, for instance, Beard et al. (1997); Orea and Kumbhakar (2004); El-Gamal and Inanoglu (2005); Greene (2005)), city sizes (Kwong and Nadarajah, 2019) and actuarial losses (Miljkovic and Grün, 2016). It has, to our knowledge, not been applied to productivity distributions before.

As argued in the introduction, the generative process of a FMM corresponds to a simple combination of the generative processes of the underlying individual densities and can therefore easily

[13] A semi-parametric approach is to be favored over a nonparametric approach in the case of heavy-tailed distributions such as firm size. This is because the heavy tails renders nonparametric procedures less efficient (Clauset et al., 2009; Dewitte, 2020). If the distribution is heavy-tailed, the common nonparametric PDF estimates such as kernel, projection and spline estimates provide misleading peaks in the 'tail' domain or oversmoothe the 'body' of the PDF (Markovich, 2008).

generalize, and is generally consistent with, existing models of firm dynamics.[14]

2.2.2 Piecewise composite distributions

Piecewise composite distributions have a probability density specified as:

$$g(x|\boldsymbol{\theta}) = \begin{cases} \alpha_1 m_1^*(x|\boldsymbol{\theta}_1) & \text{if} \quad c_0 < x \leq c_1 \\ \alpha_2 m_2^*(x|\boldsymbol{\theta}_2) & \text{if} \quad c_1 < x \leq c_2 \\ \quad \vdots & \qquad \vdots \\ \alpha_I m_I^*(x|\boldsymbol{\theta}_I) & \text{if} \quad c_{I-1} < x \leq c_I \end{cases} \tag{3}$$

where $\forall i \in I : m_i^*(x|\boldsymbol{\theta}_i) = \frac{m_i(x|\boldsymbol{\theta}_i)}{\int_{c_{i-1}}^{c_i} m_i(x|\boldsymbol{\theta}_i)dx}$ is the probability density function (PDF) of $m_i(x|\boldsymbol{\theta}_i)$ truncated at the cutoffs c_{i-1}, c_i. For this distribution to be well-behaved, additional differentiability and continuity conditions are imposed that determine the value of both component cutoffs (c_i) and probabilities (α_i) (Bakar et al., 2015), so that the vector of all model parameters reduces to the combination of the component-specific parameter vectors: $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$.

While these composite distributions can be constructed based on many individual parametric distributions, applications mostly focus on Lognormal distributions with Pareto tails. The 'Inverse Pareto-Lognormal-Pareto' distribution has been applied in the city size literature, (Ioannides and Skouras, 2013; Luckstead and Devadoss, 2017), while the 'Lognormal-Pareto' version was applied by Nigai (2017) to the firm size literature (Kondo et al., 2018). Dewitte (2020) generalizes the implementation of the piecewise composite distributions to allow for any underlying density in three- and two- piecewise composite distributions, mainly focusing on Pareto-tailed piecewise composites.

From the distribution specification in equation 3, it can be observed that piecewise composite distributions can be interpreted as mixtures of truncated densities with component probabilities restricted to ensure continuity and differentiability (Scollnik, 2007).[15] This contrasts with the

[14] Note that while this paper conceptualizes the generality of FMMs from a generative perspective, it is not able to provide evidence in favor of any specific generative process. See the methodology section (section 3), Appendix C, and the conclusion (Section 7) for a more elaborate evaluation of current limitations regarding this paper's discussion of (the generative processes of) FMMs.

[15] This becomes even more clear when we rewrite the specification of the piecewise composite distribution (eq. 3) as the weighted sum of truncated densities: $g(x|\boldsymbol{\theta}) = \alpha_1 \mathbb{I}(c_0 < x \leq c_1) m_1^*(x|\boldsymbol{\theta}_1) + \alpha_2 \mathbb{I}(c_1 < x \leq c_2) m_2^*(x|\boldsymbol{\theta}_2) + \ldots + \alpha_I \mathbb{I}(c_{I-1} < x \leq c_I) m_I^*(x|\boldsymbol{\theta}_I)$.

general mixture specification (eq. 2), where component probabilities can be interpreted as the probability that an individual observation belongs to a certain group of observations. Moreover, the generative process of piecewise distributions is rather ambiguous. It is, for instance, not clear yet which firm dynamics could explain the existence of hard cutoffs that separate the Lognormal from the Pareto distribution.

2.2.3 Product distributions

Alternatively, distributions can be combined into a product distribution: a probability distribution constructed as the distribution of the product of random variables with separate distributions. The product distribution mainly used in the literature, the Double-Pareto Lognormal distribution, results from the product of a Lognormal with a (Double-)Pareto distributed random variable (Reed and Jorgensen, 2004). This distribution is found to approximate city size distributions well, (Reed, 2002; Giesen et al., 2010), while Sager and Timoshenko (2019) fitted the distribution to Brazilian export data.

A generative process for this Double-Pareto Lognormal distribution exists (Reed and Hughes, 2002; Reed, 2002; Reed and Jorgensen, 2004) and applies to heterogeneous firms models (Arkolakis, 2016). Interestingly, the Double-Pareto Lognormal distribution can be seen as a structured infinite mixture of Lognormal distributions (Reed, 2002, p.13).[16] The Double-Pareto Lognormal distribution can therefore be absorbed by the more flexible mixture distributions as specified in equation 2. Whereas the Double-Pareto Lognormal may suffer from misspecification and/or oversimplification by imposing a structure on the mixture distribution, a FMM allows the data to determine the mixture structure needed to capture the heterogeneity present in the data.

3 Methodology

The literature review reveals the myriad of empirical evidence in favor of qualitatively very different distributions fits to productivity. This paper adds to the literature by proposing a clear statistical

[16] In the context of firm size this could mean that each age group of firms, with age referring to the time since entry in the market, is distributed Lognormally at a certain point in time. The reason the overall firm size distribution is not Lognormal is that these groups of firms have not all been evolving for the same length of time. The overall distribution of size will be a mixture of Lognormal distributions (across age groups) with time since entry as mixing parameter. When this mixing parameter is exponentially distributed, firm size will be Double-Pareto Lognormally distributed.

framework that differentiates between a sufficiently large number of distributions and evaluates their fit over a representative data range. This section establishes a methodology that uniformly fits the large but relevant range of single and combined distributions to the data. We then present statistical tests to evaluate the distributional fit and differentiate between the fitted distributions.

3.1 Distribution fitting

We rely on Maximum Likelihood (ML)[17] over all firms $b \in B$ to fit all considered distributions to the data. We consider the (Inverse) Pareto, hump-shaped distributions (Lognormal, Weibull, Fréchet, Gamma, Exponential, and Burr), and combinations of these distributions in the form of mixtures, piecewise composite or product distributions. We limit piecewise composite and product distributions to available Pareto-tailed extensions of the considered hump-shaped distributions.[18] In the case of FMMs, ML is wrapped in an Expectation-Maximization (EM) algorithm to estimate the component probabilities.

3.1.1 (Inverse) Pareto

The ML estimator for the shape parameter k over all firms $b = 1, \ldots, B$ can be obtained as

$$k_{IP} = \left[\frac{1}{B} \sum_{b=1}^{B} ln \frac{x_{max}}{x_b} \right]^{-1}, \qquad k_P = \left[\frac{1}{B} \sum_{b=1}^{B} ln \frac{x_b}{x_{min}} \right]^{-1}. \tag{4}$$

The ML estimator of the scale parameters equals the maximum and minimum observation: $\hat{x}_{min} = \min(x)$, $\hat{x}_{max} = \max(x)$, as the likelihood function is monotonically increasing (decreasing) in x_{min} (x_{max}).

[17] The choice for Maximum Likelihood contrasts with the productivity distribution literature, where popular fitting techniques rely on the minimization of squared errors between a log-linearization of the theoretical and empirical PDFs/CDFs (Axtell, 2001; di Giovanni and Levchenko, 2013; Head et al., 2014; Freund and Pierola, 2015; Bas et al., 2017; Nigai, 2017; Bee and Schiavo, 2018). Such methods, however, might not be apt to fit distribution functions. For instance, reported parameters in the literature are, to our knowledge, not obtained from a regression procedure restricted to estimate a properly normalized distribution function. Parameters obtained from an estimation procedure must result in a probability density function that integrates to 1 over the range from the lower bound up to the upper bound (due to its normalization properties) (Clauset et al., 2009). While it is possible to incorporate such constraints in the regression analysis, it has never been reported to our knowledge. Moreover, it is unclear to which extent the standard errors obtained from these methods are valid (Clauset et al., 2009; Bottazzi et al., 2015). Maximum likelihood methods do not suffer from such problems.

[18] See Appendix Tables 1, 2 and 3 for an overview of the specifications for all distributions considered. Considered distributions are chosen based on their occurrence in the economic literature.

3.1.2 Hump-shaped, piecewise composite and product distributions

The maximum likelihood of the considered *hump-shaped distributions* (Lognormal, Weibull, Fréchet, Gamma, Exponential, and Burr) is straightforward, and estimation methods are widely available. We also consider *piecewise composite distributions* as Pareto-tailed extensions of these hump-shaped distributions. The ML estimator of these distributions has no closed-form and needs to be approached numerically, see Dewitte (2020). Pareto-tailed extensions in the form of *product distributions*, on the other hand, are less generally available. We consider the Double-Pareto Lognormal distribution (Reed and Jorgensen, 2004). This distribution is the result of multiplying a Double Pareto, used by among others Arkolakis (2016), with a Lognormal distribution. Reducing the parameter space of the Double Pareto allows us to consider the Left- and Right-Pareto Lognormal distribution, respectively. The ML estimator has no closed-form solution and needs to be approached numerically (Reed and Jorgensen, 2004).

3.1.3 FMM

Direct maximum likelihood estimation of a FMM (see eq. 2) is not straightforward since the number of components I is a priori unknown. The log-likelihood function can be written as

$$logL(x|\boldsymbol{\Psi}) = \sum_{b=1}^{B}\sum_{i=1}^{I} z_{bi}\left[log(\pi_i) + log(m_i(x_b|\boldsymbol{\theta}_i))\right], \tag{5}$$

where z_{bi} is an unobserved component indicator equal to one if the observation x_b originates from subpopulation i and zero otherwise. Two steps need to be taken iteratively in order to be able to maximize this equation. The Expectation (E)-step of the s-th iteration consists of determining the conditional expectation of eq. 5 given the observed data and the current parameter estimates from iteration $s-1$:

$$\begin{aligned} Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s-1)}) &= E\left[logL(x|\boldsymbol{\Psi})|x, \boldsymbol{\Psi}^{(s-1)}\right] \\ &= \sum_{b=1}^{B}\sum_{i=1}^{I} \pi_{bi}^{(s)}\left[log(\pi_i) + log(m_i(x_b|\boldsymbol{\theta}_i))\right], \end{aligned} \tag{6}$$

where the missing data z_{ni} is replaced by the posterior probability that x_b belongs to the ith mixture:

$$\pi_{bi}^{(s)} = E\left[z_{bi}|x_b, \mathbf{\Psi}^{(s-1)}\right] = \frac{\pi_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)})}{\sum_{i=1}^{I} \pi_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)})}. \tag{7}$$

The Maximization (M)-step then, consists of maximizing the Q-function over the parameter vector $\mathbf{\Psi}$:

$$\mathbf{\Psi}^{(s)} = \max_{\mathbf{\Psi}} Q(\mathbf{\Psi}|\mathbf{\Psi}^{(s-1)}). \tag{8}$$

Each iteration updates the E- and M-step until the algorithm converges (See Miljkovic and Grün (2016) and McLachlan and Peel (2000) for a more elaborate overview).

The validity of the proposed estimation technique does not depend on its ability to identify the unobserved component indicator z_{bi}. FMMs can be utilized in two ways. First, they can be used as a semi-parametric, flexible approximation of the overall distribution. Second, they are model-based clustering methods when a certain distribution is imposed (Fop et al., 2018; Grün, 2018). While both applications rely on the idea that discrete subpopulations define the overall distribution, the semi-parametric approximation does not claim to correctly identify these subpopulations (z_{bi}). This paper relies on FMMs as a semi-parametric approximation of the productivity distribution. See Online Appendix D for a more elaborate discussion on the difference between both applications and their relevance for the current analysis.

3.2 Distribution evaluation

We use distinct criteria to differentiate between the distributions. First, we consider whether the proposed parametric distribution provides a sufficiently good fit to the data. We then differentiate between distributions using information criteria.

Goodness of fit We evaluate the parametric distributions by summarizing the distance between the empirical and parametric CDF by the 1- and ∞-norm:

$$S^0 = \sum_y \Delta^0(y), \qquad T^0 = \sup_y \Delta^0(y), \tag{9}$$

where $\Delta^0(y)$ is the normalized absolute deviation:

$$\Delta^0(y) = \left| \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(x_b \geq y) - \int_y^\infty g(x|\mathbf{\Psi}) dx \right|. \tag{10}$$

$\mathbb{I}(A)$ is the indicator of event A and $1 - G_y(x|\mathbf{\Psi}) = \int_y^\infty g(x|\mathbf{\Psi})dx$ is the complementary CDF evaluated at y. The test statistic T^0 corresponds with the Kolmogorov-Smirnov (KS) test statistic, quantifying the maximum distance between the empirical and parametric CDF. This Kolmogorov-Smirnov test statistic allows us to provide statistically underpinned claims regarding the accuracy of the distributional assumption with respect to its empirical counterpart. 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.

As we rely on estimated parameters, asymptotic distributions are not available for the test statistic. We therefore rely on a parametric bootstrap:

1. Assume B i.i.d. random variables with distribution $G(\cdot|\mathbf{\Psi})$;

2. Estimate the parameters $\mathbf{\Psi}$ of the distribution using MLE and calculate the complementary CDF, $1 - G_y(x|\mathbf{\Psi})$, and the test statistic $t \in \{S^0, T^0\}$;

3. Draw N bootstrap samples of size B from $G(\cdot|\hat{\mathbf{\Psi}})$;

4. For each sample of the parametric distribution, calculate the bootstrapped test statistics $t^* \in \left\{(S^{\tilde{0}})^*, (T^{\tilde{0}})^*\right\}$;[19]

5. The p-value is then defined as

$$\hat{p} = \frac{1}{N+1}\left[\sum_{n=1}^{N} \mathbb{I}(t_n^* \geq t) + 1\right]. \tag{11}$$

[19]Note that we do not re-fit the parametric distribution to the bootstrap sample. The vastness of the dataset at our availability in the empirical section results both in a large computational burden and a precise estimation of the distribution parameters. The influence of not refitting the parametric distribution to the bootstrap sample is therefore negligent.

Therefore, the bootstrapped p-value should be interpreted as 'the likelihood of observing a deviation between the empirical and parametric CDF as large as t under the null hypothesis', allowing us to evaluate whether observed data originates from the specified distribution. A rejection of the null hypothesis indicates the data is under-fitted.

Information Criteria We differentiate between distributions based on the log-likelihood, the Akaike or Bayesian Information Criteria. When possible, we can differentiate between two distributions based on the ratio of their likelihoods:

$$LR = \sum_{b=1}^{B} ln \frac{g_1(x_b;\cdot)}{g_2(x_b,\cdot)} \tag{12}$$

with $g_{1,2}$ the probability densities of the respective distributions.If these distributions are non-nested, the test statistic amounts to the sample average of this ratio standardized by a consistent estimate of its standard deviation (Vuong, 1989). The null hypothesis states that both distributions are equally far (in the Kullback and Leibler (1951) divergence/relative entropy sense) from the true distribution. Our test statistic will follow (asymptotically) a Gaussian distribution with mean zero if it is true. If the null is false, and $g_1(\cdot)$ is closer to the truth, the test statistic diverges to $+\infty$ with probability one. If $g_2(\cdot)$ fits the data better, it diverges to $-\infty$ (Vuong, 1989).

To avoid overfitting, the Akaike Information criterion penalizes the log-likelihood information for the number of parameters. It is defined as $AIC = 2np - 2ln(L)$ with np the number of parameters and $ln(L)$ the log-likelihood. Moreover, the AIC is asymptotically equivalent to leave-one-out cross-validation (Stone, 1977). Similarly, the Bayesian Information criterion corrects for the number of parameters as $BIC = npln(B) - 2ln(L)$. Differentiation between distributions relies then on the relative distance of the BICs: $\Delta BIC = BIC_1 - BIC_2$. The value of ΔBIC implies strong evidence in favor of distribution 1 if $\Delta BIC > 10$, moderate evidence if $6 < \Delta BIC \leq 10$ and weak evidence if $2 < \Delta BIC \leq 6$ (Kass and Raftery, 1995). BIC statistics are considered consistent for selecting the number of mixture components when the mixture model estimates a density (McLachlan and Peel, 2000; Celeux et al., 2018) and are therefore favored over AIC statistics.

4 Data

We use firm-level data from Portugal to evaluate the empirical performance of FMMs compared to "traditional" distributions such as the Log-normal or Pareto distribution. The main source of information is Sistema de Contas Integradas das Empresas (SCIE, Enterprise Integrated Accounts System) in the year 2006, a dataset covering the universe of active Portuguese firms that has been used already by, among others: (Carreira and Teixeira, 2016; Dias et al., 2016; Fernandes and Ferreira, 2017; Bastos et al., 2018; Fonseca et al., 2018).[20] It contains data both on firm-level sales and number of employees. Moreover, each firm has a unique identification number that allows us to link this dataset with a dataset on international trade.

The firm size distribution of Portugal was earlier the object of study by Cabral and Mata (2003), who relied on a longitudinal matched employer-employee dataset covering all business units with at least one wage earner in the Portuguese economy (Quadros de Pessoal). They provide evidence that the firm size distribution of Portugal is not very different from other countries such as France, the United States, Germany, Japan and the United Kingdom.

As is common in the literature, we capture heterogeneity in productivity from the distribution of firm-level sales (Head et al., 2014; Nigai, 2017; Bee and Schiavo, 2018). We can rely on the distributional relation between positive domestic sales and productivity, under specific model assumptions (Mrázová et al., 2015; Nigai, 2017; Dewitte, 2020), to approximate the productivity distribution. In a heterogeneous firms model with Constant Elasticity of Substitution (CES) demand, sales (r) will follow the same distribution as productivity (x) up to a change in distributional parameters ($r \sim x^{\sigma-1}$) if the productivity distribution is closed under power-law transformations.[21]

Our reliance on domestic rather than total sales corrects for the impact of international trade on the firm size distribution (di Giovanni et al., 2011). We reduce our dataset discarding self-employed companies[22], resulting in a dataset covering the positive domestic sales of 299,935 Portuguese firms in 2006. We note that parameter estimates of common distributons such as the Lognormal or Pareto distribution obtained from this data are similar to earlier reported parameter estimates

[20] A comparison between SCIE and the OECD SBDS database proves the full coverage of firms in our dataset for the Portuguese economy (see Online Appendix Table 6).

[21] Most common distributions used in the economic literature are closed under power-law transformations (see Online Appendix Table 1).

[22] Disregarding individual companies renders our dataset more comparable with earlier datasets used to evaluate productivity distributions such as the ORBIS database used by Nigai (2017).

obtained from different datasets (Head et al., 2014; Nigai, 2017; Sager and Timoshenko, 2019; Bee and Schiavo, 2018).

5 Results

We fit the distributions to Portuguese domestic sales in the year 2006. We initially focus on fitting the Pareto, Lognormal, combinations of Pareto and Lognormal, and up to 5-component mixtures of Lognormals to the complete data.[23] This proves to be sufficient for our main message. We show that our results hold when focusing on the tails of the data, can be extended to other economically relevant distributions, are robust to sample selection and outliers, also hold in an out-of-sample validation and cross-validation test, and can be externally validated on city size data.

5.1 Complete data

Single distributions cannot sufficiently capture the heterogeneity of the productivity distribution. Table 1 displays the selected distribution fits ordered according to their log-likelihood. One immediately observes that single parametric distributions produce the lowest log-likelihood values. This demonstrates the need, as the evolution of the literature indicates (Nigai, 2017; Sager and Timoshenko, 2019), to combine distributions in order to adequately capture heterogeneity in productivity. Still, such combinations continue to *underfit* the data (see also Dewitte (2020)). The Kolmogorov-Smirnov test statistic (T^0) indicates that the deviation between the empirical and parametric distribution is too large for both the Lognormal-Pareto and Double-Pareto Lognormal distribution, rejecting the hypothesis that the observed data could originate from these parametric distributions. The cumulative error of the CDF fit (S^0) indicates that this deviation is consistent over the complete range of the data and unlikely to originate from outliers.

Finite mixture models are the sole semi-parametric specifications that are not rejected by, and therefore do not underfit, the data. Whereas 3-component Lognormal only improves the distribution fit relative to commonly used parametric alternatives, the 4- and 5-component Lognormal distributions also result in a distribution fit that is not rejected by the data.

[23] Corresponding parameter estimates are reported in Online Appendix Table 8.

Table 1: Selected distribution fits to Portuguese domestic sales in 2006.

		Goodness of fit		Information Criteria		
Distribution	Parms.	T_a^0	S_b^0	Loglike	R_{AIC}	R_{BIC}
5-comp. Lognormal	14	0.18	0.11	12,776	1	2^{+++}
		(0.10;0.25)	(0.08;0.32)			
4-comp. Lognormal	11	0.19	0.11	12,770	2	1
		(0.09;0.25)	(0.08;0.32)			
3-comp. Lognormal	8	0.29	0.34	12,723	3	3^{+++}
		(0.10;0.24)**	(0.09;0.32)**			
Double-Pareto Lognormal	4	0.66	0.80	12,429	4	4^{+++}
		(0.09;0.25)***	(0.08;0.33)***			
2-comp. Lognormal	5	0.53	0.71	12,401	5	5^{+++}
		(0.10;0.24)***	(0.09;0.32)***			
Inv. Pareto-Lognormal-Pareto	4	0.81	1.01	12,231	6	6^{+++}
		(0.09;0.26)***	(0.08;0.34)***			
Inv. Pareto-Lognormal	3	3.02	4.26	9,198	7	7^{+++}
		(0.09;0.24)***	(0.08;0.31)***			
Lognormal-Pareto	3	2.56	3.78	8,721	8	8^{+++}
		(0.09;0.25)***	(0.08;0.32)***			
Left-Pareto Lognormal	3	3.23	4.91	8,059	9	9^{+++}
		(0.10;0.25)***	(0.09;0.32)***			
Right-Pareto Lognormal	3	2.82	4.38	8,028	10	10^{+++}
		(0.09;0.25)***	(0.08;0.32)***			
Lognormal	2	2.93	5.03	7,372	11	11^{+++}
		(0.10;0.25)***	(0.08;0.33)***			
Pareto	2	48.34	68.18	-436,227	12	12^{+++}
		(0.09;0.25)***	(0.08;0.33)***			

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.
$^{+++}$, $^{++}$, $^{+}$ indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favor of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.
$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Figure 2[24] provides a visual insight into the numerical results of Table 1. It plots the normalized absolute deviation between the empirical and parametric CDF. The figure reveals the large deviations of a single Lognormal distribution relative to both bulk and tails of the data. Augmenting the Lognormal distribution with a Pareto right tail as Nigai (2017) improves the fit marginally. While it does result in smaller deviations in the distribution's right tail, this comes at the cost of larger deviations in the left tail of the distribution and an almost equally large deviation in the bulk of the distribution as the Lognormal distribution. The best-fitting Pareto-tailed Lognormal, the Double-Pareto Lognormal, improves the distribution fit and exhibits smaller deviations over the complete data range. Still, we observe significant deviations from the data both in the bulk and the right tail of the data relative to the 4-component Lognormal, which displays small deviations over the complete data range.

This tail performance becomes even more apparent when considering the Quantile-Quantile (QQ)-plot in Figure 3. A QQ-plot allows, relative to Figure 2, to focus on the performance in the distribution's right tail. Whereas it is difficult to differentiate the tail performance of the 4-component Lognormal relative to the Lognormal-Pareto in Figure 2, the QQ-plot displayed in Figure 3 shows the divergence of the Lognormal-Pareto for large values of domestic sales. In contrast, the 4-component Lognormal distribution maintains a relatively solid fit.

The superior performance of FMMs results from their focus on fitting the complete distribution. As demonstrated in Figure 2 for the Lognormal-Pareto distribution, a focus on one part of the distribution (the right tail) can improve the fit in that part and worsen the fit in another part of the distribution (the left tail) relative to the Lognormal. FMMs aim to provide a good fit to the complete distribution by allowing for heterogeneity in productivity across components. We display the PDF of these individual components of the 4-component Lognormal in Figure 4.[25] We observe that to capture the heterogeneity observed in the data, the 4-component Lognormal mainly relies on one component (component 1) to capture the heavy tails of the distribution while components 2, and 4 mainly capture heterogeneity in the bulk of the distribution. Component 3 appears to match some extra fatness in the upper tail. As such, the results of FMMs caution against focusing

[24] This representation of the results is essentially a visually more interpretable version of the Probability-Probability plot (see Online Appendix Figure 3).

[25] See Online Appendix Figure 5 for an overview of the individual components going from a 1- up to a 5-component Lognormal.

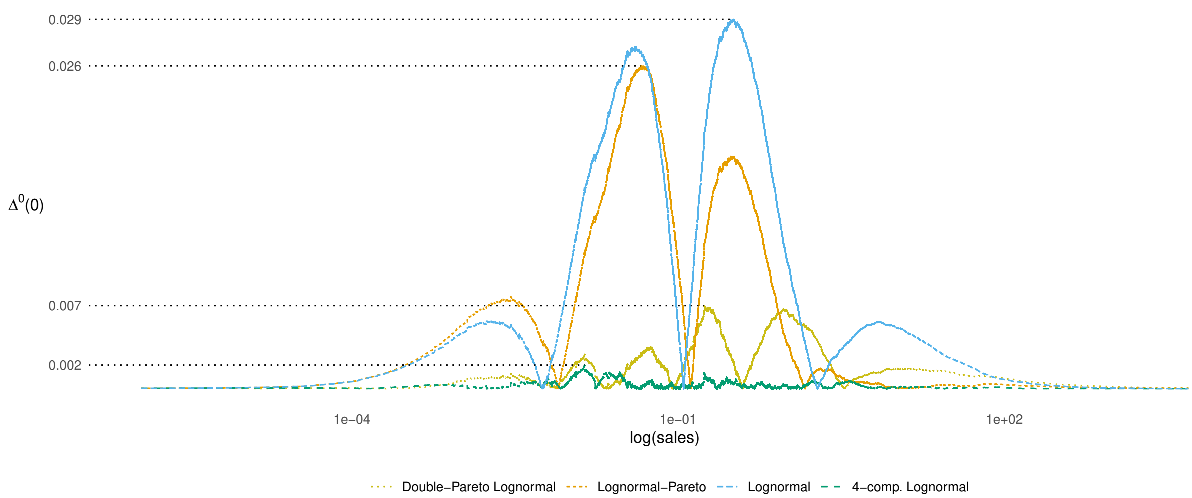

Figure 2: Normalized Absolute Deviation between the empirical and Double-Pareto Lognormal, Lognormal-Pareto, Lognormal and 4-component Lognormal CDFs over the complete range of domestic sales in Portugal, 2006.

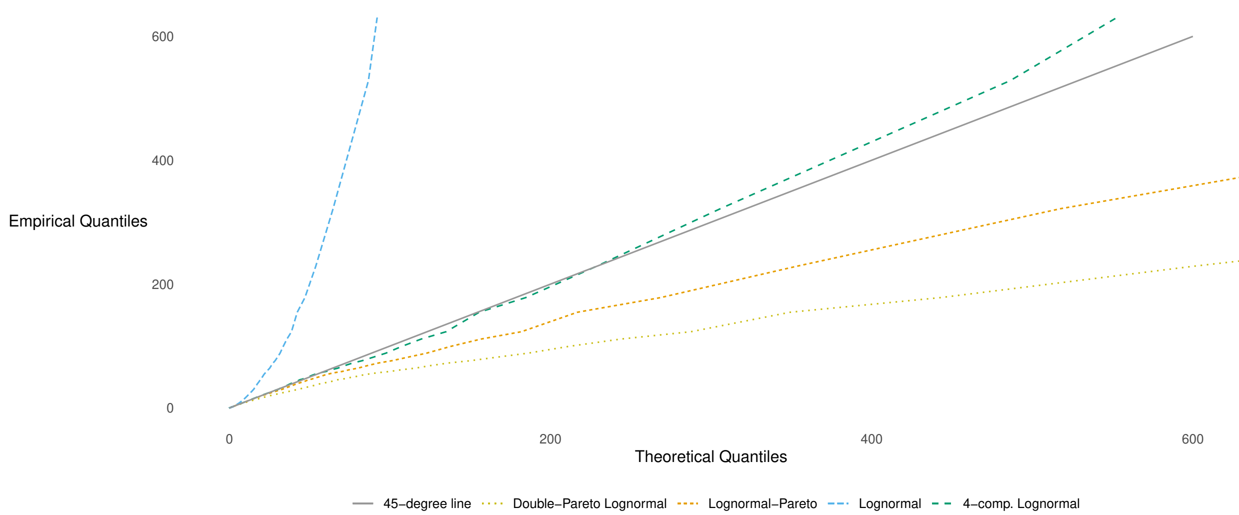

Figure 3: Quantile-Quantile plot for the Double-Pareto Lognormal, Lognormal-Pareto, Lognormal, and 4-component Lognormal over approximately 99.99% of domestic sales in Portugal, 2006.

Note: Quantiles are capped at 600 for expositional purposes, leaving out approximately the upper 0.01% of the data.

on fitting the tails of a distribution and demonstrate the importance of capturing heterogeneity in the bulk of, and consequently the complete, distribution.

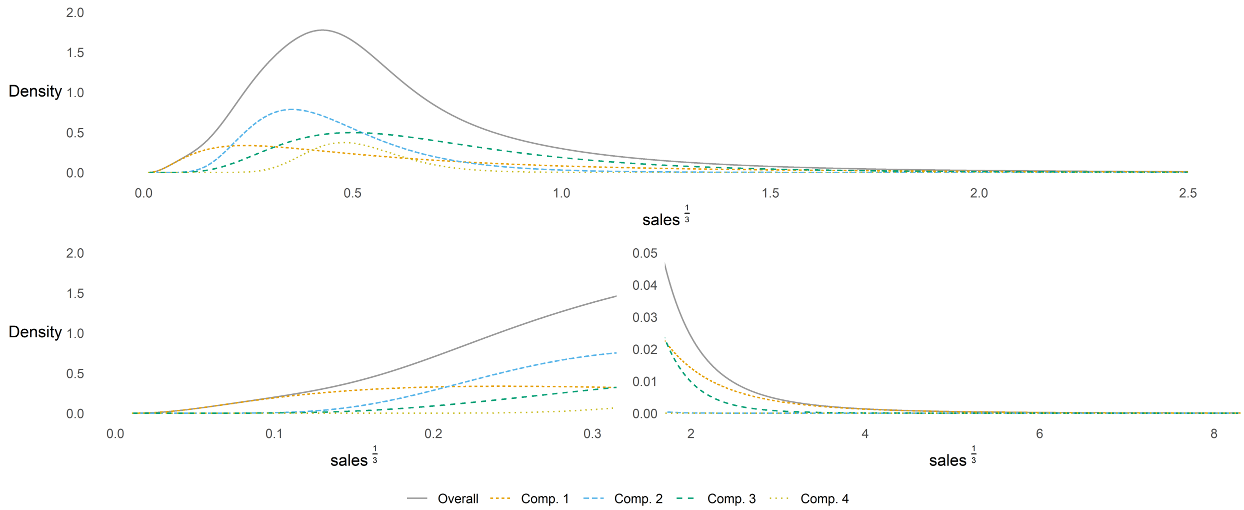

Figure 4: Probability density function of the 4-component Lognormal and its probability-weighted individual components fitted to Portuguese firm productivity in 2006. The lower left and right panels focus in on the left and right tail respectively.

Notes: Productivity is measured as domestic sales (relative to the mean) to the power of $1/(\sigma - 1)$ with σ, the elasticity of substitution between varieties, set to four. Distributions are fitted using maximum likelihood methods (cf. infra) to the complete dataset. For expositional purposes, the upper panel is restricted to productivity values between 0 and 2.5.

The focus on the complete distribution does not result in FMMs over-fitting the data. We evaluate the ability of FMMs to correctly capture the heterogeneity observed in the data using information criteria. The AIC and BIC penalize the log-likelihood for the number of parameters, indicating whether or not an increase in log-likelihood results from over-fitting the data. BIC statistics are considered consistent for selecting the number of mixture components when the mixture model is used to estimate a density (Celeux et al., 2018; McLachlan and Peel, 2000). The BIC values indicate that the 4-component Lognormal provides the best fit to the data. The increase in log-likelihood obtained from an additional, fifth component, is therefore not sufficient to justify the associated larger number of parameters. See also Online Appendix Figure 6 for an overview of the evolution of the distributional fit going from a 1- up to a 5-component Lognormal.

Overall, we demonstrate that the performance of FMMs is not the result of over-fitting but of FMMs being able to capture heterogeneity in productivity of which other distributional forms are not capable. The structure imposed on a general mixture specification to attain specific piecewise composite (in case of the Lognormal-Pareto) or product (in case of the Double-Pareto Lognormal) distributions (see section 2.2) is, therefore, not warranted.

5.2 Truncated data

Clearly, allowing for heterogeneity in distributions provides a better fit when fitting the complete distribution, but what when we only focus on the tails? This is most interesting from the Pareto point of view, which is often claimed to be a good fit to the right tail of the productivity distribution.[26]

Online Appendix Table 9 displays the results of fitting the (Inverse) Pareto to the (left) right tail of the distribution using the methods described in online Appendix B. We recovered the best-fitting truncation point for the (Inverse) Pareto distribution, assigning 8.53% and 6.07% of the data to the left and right tail, respectively. We reduced our dataset according to these truncation parameters and fitted truncated mixtures of Lognormals to both tails of the distribution for comparison. This approach puts the Pareto distribution twice in the advantage. First, it is free from a parametric specification for the bulk of the distribution. Second, the truncation parameter is chosen in function of the best-fitting (Inverse) Pareto distribution. As a result, the (Inverse) Pareto and (mixtures of) the Lognormal provide a good fit to the tails according to the Kolmogorov-Smirnov test.

Nevertheless, despite the advantage for the (Inverse) Pareto distribution, it seems that (mixtures of) the Lognormal distribution provide a significantly better fit to the tails of the data. (Mixtures of) the Lognormal distribution have a higher log-likelihood and lower deviation from the empirical CDF than the (Inverse) Pareto distribution. This results in the likelihood ratio test significantly rejecting Pareto in favor of (mixtures of) the Lognormal distribution, which is in line with earlier results reported in related literature (Clauset et al., 2009). When correcting for the number of parameters, the BIC reveals that the single Lognormal distribution is sufficient to fit the tail only. A mixture of Lognormals insufficiently improves the fit to justify the corresponding increase in the number of parameters.

[26]Note that this argument carries the normative value that obtaining a good fit for larger firms is absolute, regardless of the implications for the fit to smaller firms.

5.3 Robustness and Extensions[27]

We scrutinize the robustness of our results with several additional analyses. First, we examine whether our results are not caused by sample selection. To this end, we restrict our dataset to the manufacturing sector only (see Online Appendix Table 10) and find the performance of FMMs to improve relative to Pareto-tailed distributions. Second, we inspect whether our results are not due to outliers in the tails of the distribution by discarding the 1,000 smallest and largest observations from our dataset. Results in Appendix Table 11 again confirm the superiority of FMMs. Third, the AIC reported in Table 1 is asymptotically equivalent to leave-one-out cross-validation (Stone, 1977). We perform a robustness check on the out-of-sample predictive accuracy of our results using (i) a Monte Carlo Cross-Validation (MCCV), (ii) $k-$fold cross-validation, and (iii) an out-of-sample test for model selection. The results of this exercise (see Online Appendix Table 12) confirm the main results and demonstrate that a mixture of Lognormals improves the model fit without over-fitting the data. Finally, we also provide external validation, in line with Nigai (2017), by fitting the considered distributions to the U.S. Census 2000 city size distribution data.[28] Appendix Table 13 provides the test results, demonstrating that the city size distribution is neither Lognormal, Pareto, nor Pareto-tailed Lognormal. It is best approximated by a 2-component Lognormal distribution (according to the BIC).

The superior performance of FMMs is not limited to the Lognormal distribution. Appendix Table 7 displays the results of fits to the complete data extending to FMMs of distributions often used in the economic literature such as the Exponential, Gamma, Weibull, Burr, and Fréchet distribution. Most of these mixtures are not able to match the performance of the Lognormal. Only the Burr distribution provides an equivalent fit to the PDF and CDF.[29] In comparison, commonly used parametric alternatives as the Double-Pareto Lognormal (Sager and Timoshenko, 2019) and the Lognormal-Pareto (Nigai, 2017) distribution are ranked sixteenth and thirty-first, respectively, according to BIC, out of 52 considered distributions.

The consistent excellent performance of the Lognormal distribution can be motivated from two

[27]See Online Appendix C for a full discussion of the robustness tests and extensions of the results.

[28]This dataset has been subject to an extensive debate in the city size literature, including the discussion between Eeckhout (2004, 2009) and Levy (2009), and is available at https://www.aeaweb.org/aer/data/sept09/20071478_data.zip.

[29]The Burr distribution fails to match higher moments of the data, however. See also section 6.

perspectives. From the perspective of overall fit, a mixture of (log-) normal distributions with sufficient components is assumed to be able to approach all distributions (McLachlan and Peel, 2000). From a generative perspective for individual components, the Lognormal distribution is the realization of applying the Central Limit Theorem (CLT) in the log domain: firm heterogeneity will approximately be Lognormal if it is the multiplicative product of many independent random variables. This corresponds with extensions of heterogeneous firms models à la Melitz (2003) that consider multi-dimensional firm heterogeneity, taking into consideration the product dimension (Bernard et al., 2009) or uncertainty in demand and/or supply (see for instance De Loecker (2011); Bas et al. (2017); Sager and Timoshenko (2019); Gandhi et al. (2020)).

6 Gains From Trade implications

To evaluate the economic relevance of our finding, that FMMs are the only distributions not rejected by the data, we perform a stylized GFT exercise along the lines of Melitz and Redding (2015); Bee and Schiavo (2018).[30] This exercise allows us to demonstrate that (i) only GFT obtained from a FMM are not rejected by the data, (ii) these results are not a trivial implication from the distributional fit, and (iii) FMMs are the only distributional forms that accurately capture the different channels through which trade affects welfare.

Our setup is a two-country symmetric heterogeneous firms model with a finite number of firms.[31] The parameterization of our model is standard (Head et al., 2014; Melitz and Redding, 2015; Bee and Schiavo, 2018). We work with two symmetric countries i and j and choose labor in one country as the numeraire, so that $W^i = W^j = 1$. We choose fixed entry costs $f^e = 0.545$ and set fixed costs equal to one ($f^{ii} = f^{ij} = 1$). The elasticity of substitution is set to four. The productivity distribution is assumed exogenous.[32]

Finally, we need to capture the heterogeneity distribution. Assuming a parametric distribution and under the assumption of an *infinite* number of firms, we can calculate the necessary analytical

[30] Note that we rely on a stylized model that does not represent reality to focus specifically on the performance between parametric distributions.

[31] See Appendix E for a full workout of the model.

[32] If the data generating process of FMMs is an endogenous response to fundamental shocks, this exogeneity assumption results in policy counterfactuals that might not capture some first-order variation of the subsequent response to the shock and therefore to a biased quantification of the gains from trade. See for an example Brooks and Dovis (2020) in the case of credit-constrained firms.

expressions using the distributional parameters from our empirical analysis to capture heterogeneity. Following Nigai (2017), we can also capture heterogeneity directly from the empirical, *finite* data. To compare GFT obtained assuming a parametric distribution and GFT obtained from the finite data, we perform a parametric bootstrap. This parametric bootstrap generates a range of finite sample estimates under the hypothesis that the observed data is generated by a certain parametric distribution, which allows for a comparison with the observed finite data (Dewitte, 2020).

We calculate the changes in welfare due to a trade shock (Gains From Trade), which 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 ($m^{\sigma-1}_{\omega^{ij*}}$)[33] and the bilateral trade share (λ^{ij}):

$$
\begin{aligned}
100 \times ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} &= 100 \times -ln\frac{(P^i)'}{P^i} \qquad (13)\\
&= 100 \times -\left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma-1}\left(ln\frac{(M^i)'}{M^i} - ln\frac{(m^0_{\omega^{ii*}})'}{m^0_{\omega^{ii*}}} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}}\right)\right].
\end{aligned}
$$

Our exercise reduces the variable trade costs from $\tau^{ij} = 3$ to $(\tau^{ij})' = 1$. The obtained GFT are displayed in Figure 5. This figure presents the parametric bootstrapped distribution of GFT using box plots delineating the 5th, 25th, 50th, 75th, and 95th quantile. The vertical blue line indicates empirical GFT. Green circles are the average parametric finite sample GFT, and yellow diamonds show the parametric plug-in population estimates of GFT.

We observe that mixture models are the only distributions able to provide an approximation of GFT that is not rejected by the data. Heavy-(Pareto-) tailed distributions significantly overestimate GFT, while relatively light-tailed distributions underestimate GFT. The distributions in Figure 5 are ordered according to their distance from the empirical GFT. As such, we can interpret the 4-component Lognormal distribution as providing the closest fit to the GFT obtained from the empirical distribution. The empirical values imply an increase in real income per capita of 19.01%

[33] We 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).

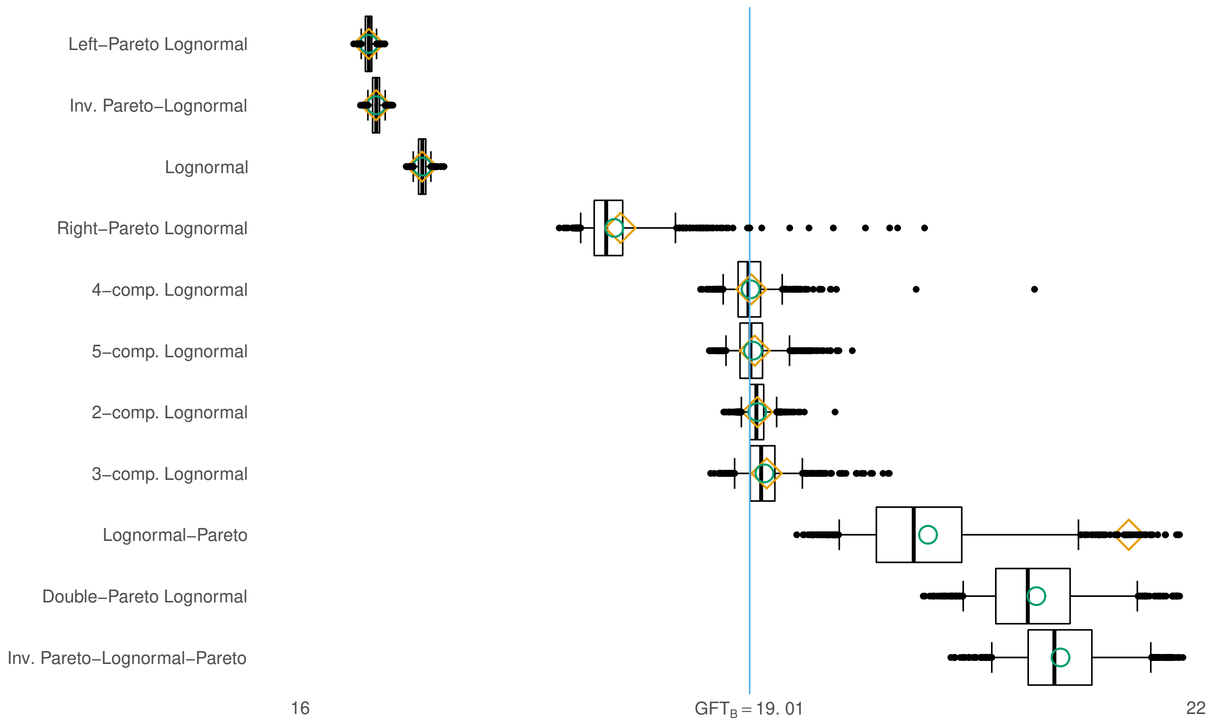

Figure 5: Gains from a reduction in variable trade costs $\tau^{ij} = 3$ to $(\tau^{ij})' = 1$.

Notes: Box-plots display the 5th, 25th, 50th, 75th, and 95th quantile of the asymptotic distribution of parametric finite sample GFT obtained from a bootstrap with 999 replications. Yellow diamonds represent the parametric 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 were obtained from a sample of 299,935 firms.

when reducing variable trade costs from 3 to 1. The 4-component Lognormal distribution closely predicts this to be 19.02%, as can be deduced from the parametric plug-in population estimates (yellow diamonds). Moreover, the close fit results in an excellent approximation of the empirical GFT, as can be deduced from the parametric bootstrapped finite sample GFT being at least as small as the empirical GFT in more than 5% of the cases (the box-plot overlaps with the vertical blue line). This contrasts with the simple Lognormal distribution underestimating the empirical GFT by about 11%, predicting GFT to amount to 16.8%, and the Lognormal-Pareto distribution overestimating the empirical GFT by approximately 13%, predicting a 21.55% increase in welfare.[34] Our results confirm the findings of Dewitte (2020) that GFT obtained from currently considered distributions are unlikely to materialize and emphasize the importance of considering FMMs to capture heterogeneity in productivity.

These results are not a trivial implication of the distributional fits reported in the previous section. A good fit to the CDF does not necessarily imply that higher moments of the distribution are well approximated,[35] while Dewitte (2020) demonstrates that higher distributional moments are essential when evaluating distributional performance regarding GFT predictions.

Therefore, a ranking of the distributions according to GFT performance (see Figure 5) does not closely follow the ranking of the fit to the 0th moment (the CDF) of the distribution (see Table 1). The Double-Pareto Lognormal, for instance, provides a closer fit to the empirical CDF than the Right-Pareto Lognormal, but provides worse GFT approximations. This can be attributed to the relatively heavy tail of the Double-Pareto Lognormal, resulting in a large error when calculating higher moments of the distribution. A ranking of distributions based on the fit to average lower-truncated sales proves to be a better indicator of GFT performance, as can be deduced from the maximal normalized absolute deviation between the empirical and parametric average of lower-truncated sales, T^1, in online Appendix Table 7 (Dewitte, 2020).[36]

[34] A comparison in percentage rather than absolute differences is preferred due to the stylized model this calibration exercise relies on. Absolute differences are likely more sensitive to model specification and parametrization. See Costinot and Rodríguez-Clare (2014) for a discussion on the sensitivity of GFT on model specifications.

[35] Only when a distribution provides a *sufficiently* good fit to the CDF (according to the Kolmogorov-Smirnov distance), one can be ascertained higher moments of the distribution will be well approximated.

[36] We evaluate the parametric average of lower-truncated sales by summarizing the distance between the empirical and parametric average of lower-truncated sales by the 1- and ∞-norm:

$$S^1 = \sum_y \Delta^1(y), \qquad T^r = \sup_y \Delta^1(y),$$

We demonstrate this reasoning more clearly by evaluating the channels through which the differences in GFT between distributions come about. Table 2 reports the weighted components of welfare gains (see eq. 13) for all considered distributional forms. We observe that the deviation of the parametric results compared to the empirical distribution is relatively small for changes in the number of firms and the probability of successful entry into the domestic market. The largest differences can be found for the changes in the average productivity of exporting firms and for the trade shares. Heavy-tailed distributions largely underestimate the positive effect of the increasing average productivity of exporting firms (which relates to average sales (Dewitte, 2020)) and the negative effect of the increasing bilateral trade shares compared to the empirical distribution, while the reverse is true for lighter-tailed distributions. The Lognormal-Pareto distribution, for instance, predicts the weighted average productivity of exporting firms to increase by 2%, an underestimation by ±900%, and the weighted bilateral trade shares by 90%, an underestimation by ±20%. The Lognormal distribution, on the other hand, predicts the weighted average productivity of exporting firms to increase by 53%, an overestimation by ±165%, and the weighted bilateral trade shares by 144%, an overestimation by ±31%.

The distribution-dependent differences in the reaction of the bilateral trade shares to changes in variable trade costs can be traced back to the aggregate trade elasticity.[37] It is a well-known result that the Pareto assumption results in a trade elasticity that is constant across export markets, because the importance of the extensive margin elasticity in the overall trade elasticity is not affected by the difficulty of the market (Chaney, 2008; Bas et al., 2017). Similarly, heavy-tailed distributions such as the Double-Pareto Lognormal (Sager and Timoshenko, 2019) or Lognormal-

where $\Delta^1(y)$ represents the normalized absolute deviation:

$$\Delta^1(y) = \frac{\left| \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}(x_b \geq y) x_b^1 - \int_y^\infty x^1 g(x|\mathbf{\Psi})dx \right|}{\frac{1}{B}\sum_{b=1}^{B} x_b^1}.$$

.

While this normalized absolute deviation provides an indication of the distance between the empirical and parametric lower-truncated average, it is not informative regarding the accuracy of the distributional assumption. The calculated statistics do not generalize to upper-truncated averages. We are grateful to an anonymous referee for pointing this out. A parametric bootstrap is relied upon to provide the asymptotic distribution of the calculated statistics.

[37]The aggregate trade elasticity can be obtained as $\gamma^{ij} = \underbrace{1-\sigma}_{\text{intensive margin}} - \underbrace{\frac{e^{(\sigma-1)\omega^{ij*}}}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)}}_{\text{weights}} \times \underbrace{\frac{d\ln M^{ij}}{dln\tau^{ij}}}_{\text{extensive margin}}$,

where $\frac{dlnM^{ij}}{dln\tau^{ij}} = \frac{e^{\omega^{ij*}} g\left(\omega^{ij*}\right)}{1-G\left(\omega^{ij*}\right)}$

Pareto (See Figure 7 in Online Appendix) distribution predict a quasi-constant trade elasticity. This invariance of the trade elasticity implied by heavy-tailed distributions results in an underestimation of the reaction of the bilateral trade shares to a change in variable trade costs. The light-tailed Lognormal distribution, on the other hand, attaches relatively much importance to the extensive margin elasticity and, as a result, overestimates the change in bilateral trade shares due to trade liberalization. The aggregate trade elasticity predicted by the 4-component FMM model nicely fits in between the predictions of these light- and heavy-tailed distributions, as can be observed in Figure 7. This, in turn, allows FMMs to accurately predict the change in bilateral trade shares for a change in variable trade costs.

Table 2: 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	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	0.16	1.10	-0.17	0.15	0.60	-1.51
		(0.16;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.58;0.62)***	(-1.53;-1.49)***
Inv. Pareto-Lognormal	3	0.17	1.10	-0.17	0.15	0.58	-1.49
		(0.16;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.56;0.60)***	(-1.51;-1.47)***
Lognormal	2	0.17	1.10	-0.17	0.15	0.53	-1.44
		(0.17;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.51;0.55)***	(-1.46;-1.42)***
Right-Pareto Lognormal	3	0.18	1.10	-0.18	0.17	0.28	-1.19
		(0.18;0.19)**	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.23;0.33)**	(-1.24;-1.13)**
Empirical	0	0.19	1.10	-0.18	0.18	0.20	-1.10
4-comp. Lognormal	11	0.19	1.10	-0.18	0.18	0.20	-1.10
		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.18;0.22)	(-1.13;-1.08)
5-comp. Lognormal	14	0.19	1.10	-0.19	0.18	0.20	-1.10
		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.19)	(0.17;0.22)	(-1.12;-1.07)
2-comp. Lognormal	5	0.19	1.10	-0.17	0.17	0.23	-1.13
		(0.19;0.19)	(1.10;1.10)	(-0.18;-0.17)***	(0.16;0.17)***	(0.22;0.25)***	(-1.15;-1.12)***
3-comp. Lognormal	8	0.19	1.10	-0.18	0.18	0.19	-1.09
		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.16;0.22)	(-1.12;-1.06)
Lognormal-Pareto	3	0.22	1.10	-0.22	0.22	0.02	-0.90
		(0.20;0.21)***	(1.10;1.10)	(-0.22;-0.20)***	(0.20;0.22)***	(0.04;0.14)***	(-1.04;-0.93)***
Double-Pareto Lognormal	4	-	1.10	-	-	-	-
		(0.20;0.22)***	(1.10;1.10)	(-0.20;-0.19)***	(0.19;0.20)***	(0.02;0.09)***	(-0.98;-0.90)***
Inv. Pareto-Lognormal-Pareto	4	-	1.10	-	-	-	-
		(0.21;0.22)***	(1.10;1.10)	(-0.20;-0.18)*	(0.18;0.20)***	(0.01;0.08)***	(-0.97;-0.89)***

Notes: $ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i}$ indicates the log changes in real per-capita income due to an exogenous increase in variable trade costs τ_{ij} to τ'_{ij}. This is 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 ($m^{\sigma-1}_{\omega^{ij*}}$) and the bilateral trade share (λ^{ij}).
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.

The reported findings are not the result of a specific parametrization of the model. Figure 4 displays the percentage errors in parametric GFT calculations relative to the empirical benchmark for different parametrization scenarios. Our findings are robust for different values of the elasticity of substitution (left upper panel) and fixed entry costs (left bottom panel), as well as for different starting values for the iceberg trade costs (right upper panel) and a reduction in fixed rather than variable trade costs (right bottom panel).

7 Conclusion

This paper provides evidence that heterogeneity in the firm-level productivity distribution can be captured most adequately by Finite Mixture Models. A clear statistical framework differentiates between the fit of 52 distributions to domestic sales of the population of active Portuguese firms in 2006. The flexible, semi-parametric nature of FMMs results in a substantial empirical performance improvement compared to commonly used parametric alternatives in the firm size literature. FMMs are the only distributions that are not rejected by the data and provide a sufficiently good approximation of Gains From Trade (GFT). Whereas other parametric distributions significantly over- or underestimate the empirical (or 'true') GFT, FMMs can adequately track the 'true' GFT. The superior performance of FMMs follows from their ability to accurately capture heterogeneity in the bulk of the distribution, which is overlooked by commonly used parametric alternatives to FMMs.

FMMs can be relied upon either to capture heterogeneity in productivity or to cluster productivity into discrete sub-populations. Our results provide strong evidence in favor of FMMs from the first perspective. We take no stance on distribution type or the mixing parameter (or mechanism) that defines the underlying discrete subpopulations. It is clear that the two are closely interconnected, and therefore not easily identifiable. Further research is necessary to define which mechanisms result in multiple individual densities defining the overall productivity distribution.

The idea of FMMs also opens many new venues for ongoing research. For instance, the mechanisms driving firm-level dynamics in aggregate growth models are determined by the parametric approximation of the productivity distribution (see, for instance, Luttmer (2007); Arkolakis (2016)). A correct parametric approximation is then essential to motivate the determinants of a firm's productivity growth. Moreover, the estimation of productivity usually relies on an identical first-order Markov process for the complete population. Concurrently, however, it is recognized that productivity dynamics are endogenous to exporting (De Loecker, 2013), importing (Kasahara and Rodrigue, 2008), innovation (Aw et al., 2011), management practices (Bloom and Reenen, 2011; Caliendo et al., 2020), et cetera. Introducing Finite Mixture Modeling into the estimation procedures would allow, semi-parametrically, to control for such discrete subpopulations without the risk of model misspecification. Moreover, the potential identification of these subpopulations provides the opportunity to discriminate between the many different mechanisms (see, for instance, Cabral and Mata (2003); Klette and Kortum (2004); Rossi-Hansberg and Wright (2007); Atkeson and Burstein (2010); Caliendo et al. (2020)) that drive the existence of such subpopulations. Also, the propagation of firm-level volatility to the aggregate level mainly relies on a Pareto specification for the right tail of the productivity distribution (Gabaix, 2011; di Giovanni and Levchenko, 2012; Carvalho and Grassi, 2019). FMMs are sufficiently heavy-tailed to motivate granularity.

References

Albuquerque, R. and H. A. Hopenhayn (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies 71*(2), 285–315.

Angelini, P. and A. Generale (2008). On the evolution of firm size distributions. *The American Economic Review 98*(1), 426–438.

Arkolakis, C. (2016). A Unified Theory of Firm Selection and Growth. *The Quarterly Journal of Economics 131*(1), 89–155.

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

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

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review 101*(4), 1312–44.

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.

Bakar, S. A. A. and S. Nadarajah (2013). CompLognormal: An R Package for Composite Lognormal Distributions. *The R Journal 5*(2), 97–103.

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.

Bastos, P., J. Silva, and E. Verhoogen (2018). Export Destinations and Input Prices. *American Economic Review 108*(2), 353–392.

Beard, T. R., S. B. Caudill, and D. M. Gropper (1997). The diffusion of production processes in the us banking industry: A finite mixture approach. *Journal of Banking & Finance 21*(5), 721–740.

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., S. J. Redding, and P. K. Schott (2009). Products and productivity. *The Scandinavian Journal of Economics 111*(4), 681–709.

Bloom, N. and J. V. Reenen (2011). Chapter 19 - human resource management and productivity. Volume 4 of *Handbook of Labor Economics*, pp. 1697–1767. Elsevier.

Bottazzi, G., D. Pirino, and F. Tamagni (2015). Zipf law and the firm size distribution: a critical discussion of popular estimators. *Journal of evolutionary economics 25*(3), 585–610.

Brooks, W. and A. Dovis (2020). Credit market frictions and trade liberalizations. *Journal of Monetary Economics 111*, 32–47.

Cabral, L. M. B. and J. Mata (2003). On the evolution of the firm size distribution: Facts and theory. *American Economic Review 93*(4), 1075–1090.

Caliendo, L., G. Mion, L. D. Opromolla, and E. Rossi-Hansberg (2020). Productivity and Organization in Portuguese Firms. *Journal of Political Economy forthcoming.*

Caliendo, L. and E. Rossi-Hansberg (2012). The impact of trade on organization and productivity. *The Quarterly Journal of Economics 127*(3), 1393–1467.

Carreira, C. and P. Teixeira (2016). Entry and exit in severe recessions: lessons from the 2008–2013 Portuguese economic crisis. *Small Business Economics 46*(4), 591–617.

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

Celeux, G., S. Frühwirth-Schnatter, and C. P. Robert (2018). *Handbook of mixture analysis*, Chapter 7 - Model selection for mixture models—perspectives and strategies, pp. 121–160. CRC Press.

Chaney, T. (2008). 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.

Clementi, G. L. and H. A. Hopenhayn (2006). A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics 121*(1), 229–265.

Cooley, T. F. and V. Quadrini (2001). Financial markets and firm dynamics. *American Economic Review 91*(5), 1286–1310.

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.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

Desai, M., P. Gompers, and J. Lerner (2003). Institutions, capital constraints and entrepreneurial firm dynamics: Evidence from europe. Working Paper 10165, National Bureau of Economic Research.

Dewitte, R. (2020). From Heavy-tailed Micro to Macro: On the characterization of firm-level heterogeneity and its aggregation properties. MPRA Paper 103170, 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. and A. A. Levchenko (2013). Firm entry, trade, and welfare in zipf's world. *Journal of International Economics 89*(2), 283–296.

di Giovanni, J., A. A. Levchenko, and R. Rancière (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics 85*(1), 42–52.

Dias, D. A., C. R. Marques, and C. Richmond (2016). Misallocation and productivity in the lead up to the Eurozone crisis. *Journal of Macroeconomics 49*, 46–70.

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

Eeckhout, J. (2009). Gibrat's Law for (All) Cities: Reply. *American Economic Review 99*(4), 1676–1683.

El-Gamal, M. A. and H. Inanoglu (2005). Inefficiency and heterogeneity in turkish banking: 1990–2000. *Journal of Applied Econometrics 20*(5), 641–664.

Feenstra, R. C. (2018). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics 110*, 16–27.

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.

Fernandes, A. P. and P. Ferreira (2017). Financing constraints and fixed-term employment: Evidence from the 2008-9 financial crisis. *European Economic Review 92*, 215–238.

Fonseca, T., F. Lima, and S. C. Pereira (2018). Understanding productivity dynamics: A task taxonomy approach. *Research Policy 47*(1), 289–304.

Fop, M., T. B. Murphy, et al. (2018). Variable selection methods for model-based clustering. *Statistics Surveys 12*, 18–65.

Freund, C. and M. D. Pierola (2015). Export Superstars. *The Review of Economics and Statistics 97*(5), 1023–1032.

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.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Giesen, K., A. Zimmermann, and J. Suedekum (2010). The size distribution across all cities – double pareto lognormal strikes. *Journal of Urban Economics 68*(2), 129 – 137.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics 126*(2), 269–303.

Grimm, K. J., G. L. Mazza, and P. Davoudzadeh (2017). Model selection in finite mixture models: A k-fold cross-validation approach. *Structural Equation Modeling: A Multidisciplinary Journal 24*(2), 246–256.

Grün, B. (2018). Model-based clustering. arXiv preprint 1807.01987, arXiv.

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

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics 123*(2), 441–487.

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.

Kasahara, H. and J. Rodrigue (2008). Does the use of imported intermediates increase productivity? plant-level evidence. *Journal of Development Economics 87*(1), 106–118.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Klette, T. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy 112*(5), 986–1018.

Kondo, I., L. T. Lewis, and A. Stella (2018). On the u.s. firm and establishment size distributions. FEDS Working Paper 2018-075.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

Kwong, H. S. and S. Nadarajah (2019). A note on "pareto tails and lognormal body of us cities size distribution". *Physica A: Statistical Mechanics and its Applications 513*, 55 – 62.

Lentz, R. and D. T. Mortensen (2008). An empirical model of growth through product innovation. *Econometrica 76*(6), 1317–1373.

Levy, M. (2009). Gibrat's Law for (All) Cities: Comment. *American Economic Review 99*(4), 1672–1675.

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.

Malevergne, Y., V. Pisarenko, and D. Sornette (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Phys. Rev. E 83*.

Markovich, N. (2008). *Nonparametric analysis of univariate heavy-tailed data: research and practice*, Volume 753. John Wiley & Sons.

McLachlan, G. J. and D. Peel (2000). *Finite mixture models*. New York: Wiley Series in Probability and Statistics.

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.

Miljkovic, T. and B. Grün (2016). Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics 70*, 387 – 396.

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.

Orea, L. and S. C. Kumbhakar (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics 29*(1), 169–183.

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

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science 42*(1), 1–17.

Reed, W. J. and B. D. Hughes (2002). From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E 66*(6), 067103.

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.

Rossi-Hansberg, E. and M. L. J. Wright (2007). Establishment size dynamics in the aggregate economy. *American Economic Review 97*(5), 1639–1666.

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.

Scollnik, D. P. M. (2007). On composite lognormal-pareto models. *Scandinavian Actuarial Journal 2007*(1), 20–33.

Smyth, P. (1996). Clustering using monte carlo cross-validation. In *Kdd*, Volume 1, pp. 26–133.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 44–47.

Virkar, Y. and A. Clauset (2014). Power-law distributions in binned empirical data. *The Annals of Applied Statistics*, 89–119.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica 57*(2), 307–333.

Online Appendix to "Unobserved Heterogeneity in the Productivity Distribution and Gains From Trade"

Ruben Dewitte, Michel Dumont, Glenn Rayp, Peter Willemé*

19th December 2022

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be, Dumont: Federal Planning Bureau and Ghent University, Rue Belliard - Belliardstraat 14-18 1040 Brussels, dm@plan.be , Rayp: Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, glenn.rayp@ugent.be, Willemé: Federal Planning Bureau and Ghent University, Rue Belliard - Belliardstraat 14-18 1040 Brussels, pw@plan.be. We thank Pete Klenow, Sergey Nigai, Luca David Opromolla, Gonzague Vannoorenberghe, Olivier Thas, Gerdie Everaert, Bruno Merlevede, Marijn Verschelde, and participants at the 2018 ETSG conference in Warsaw, the 2019 ETSG conference in Bern, the DEGIT XIV conference in Odense, and the 2019 workshop on "Firm Heterogeneity in Technical Change" in Ghent for their helpful comments and suggestions. Ruben Dewitte acknowledges financial support from Research Foundation Flanders (FWO) grant 12B8822N.

Appendix A Additional Figures and table

A.1 Figures

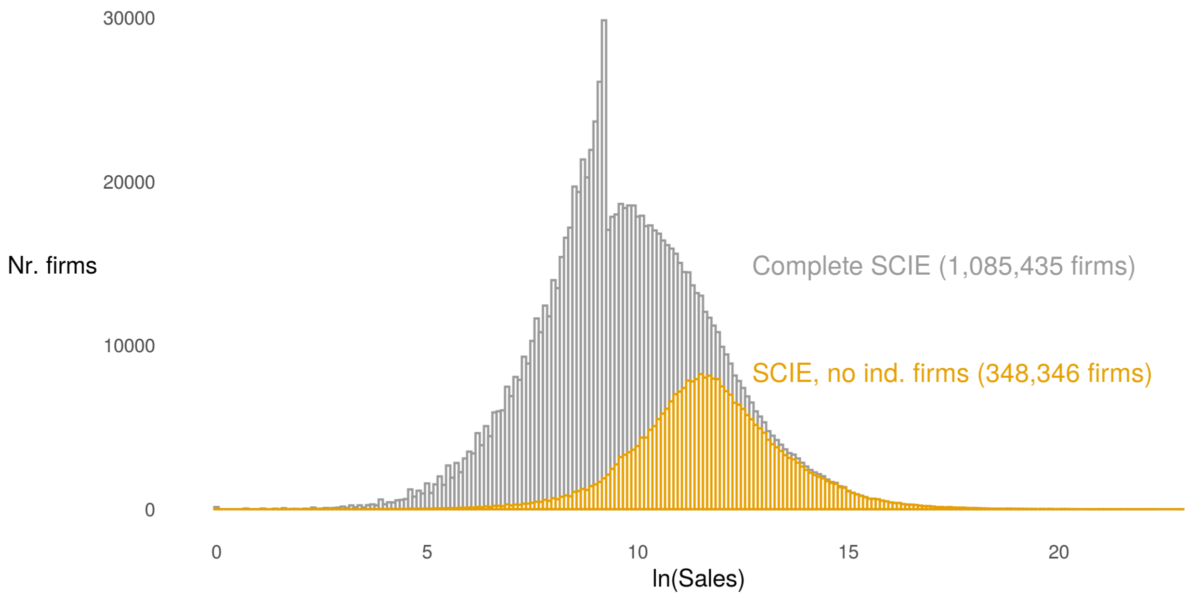

Figure 1: Density comparison of the SCIE dataset with and without individual companies.

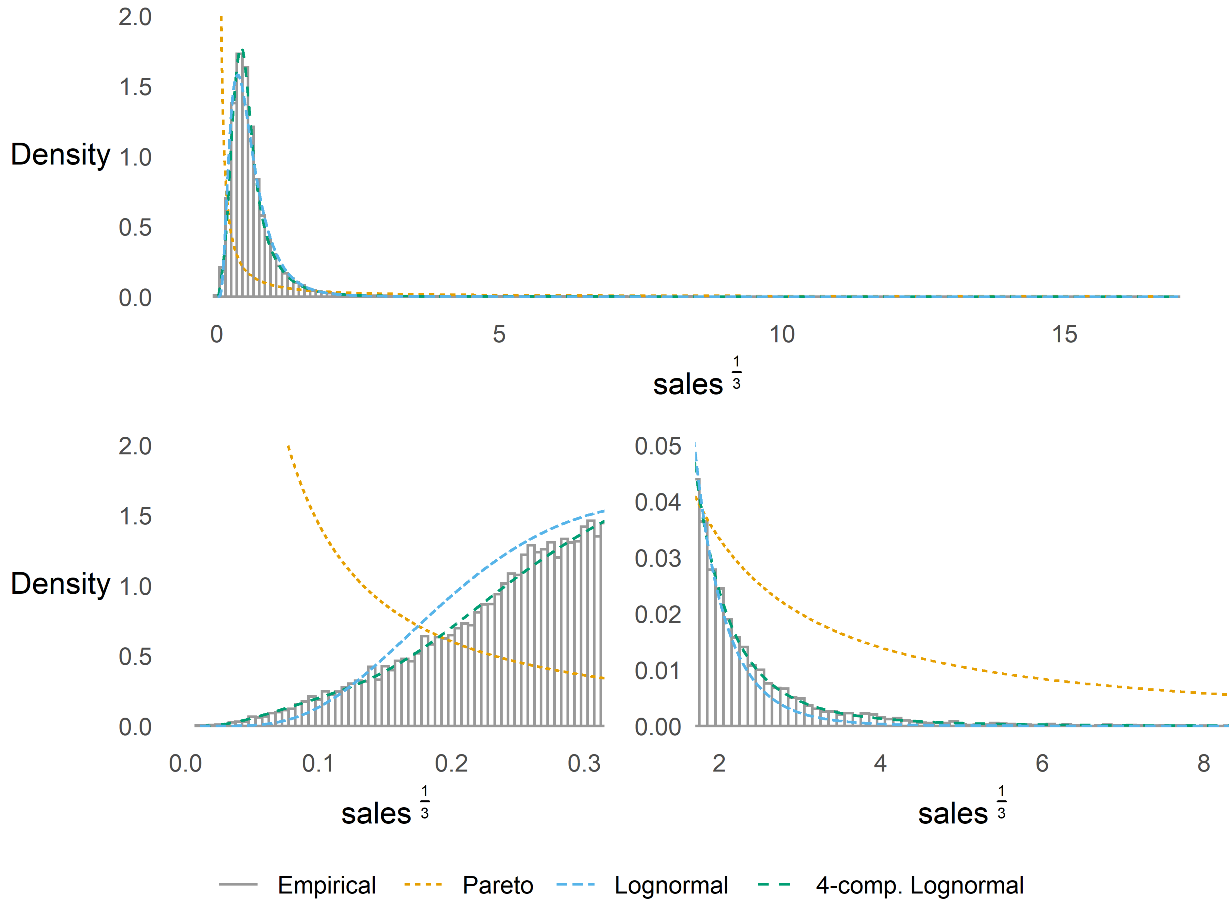

Figure 2: Empirical probability density function of Portuguese firm productivity in 2006 (upper panel) with fitted Pareto and (4-component) Lognormal densities. The lower left and right panels focus in on the left and right tail respectively.

Notes: Productivity is measured as domestic sales (relative to the mean) to the power of $1/(\sigma - 1)$ with σ, the elasticity of substitution between varieties, set to four. Distributions are fitted using maximum likelihood methods (cf. infra) to the complete dataset.

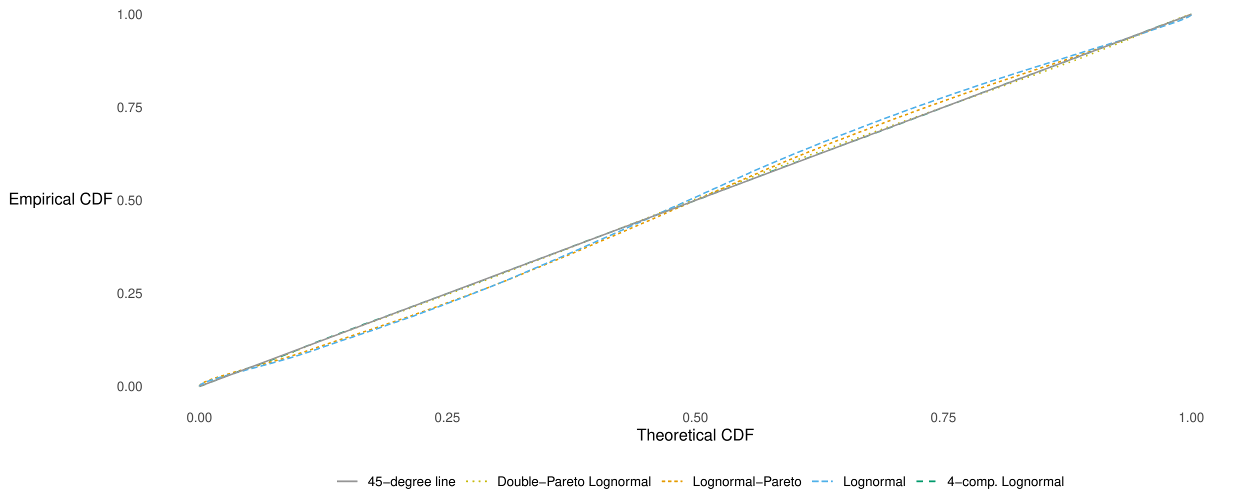

Figure 3: Probability-Probability plot for the Double-Pareto Lognormal, Lognormal-Pareto, Lognormal and 4-component Lognormal over the complete range of domestic sales in Portugal, 2006.

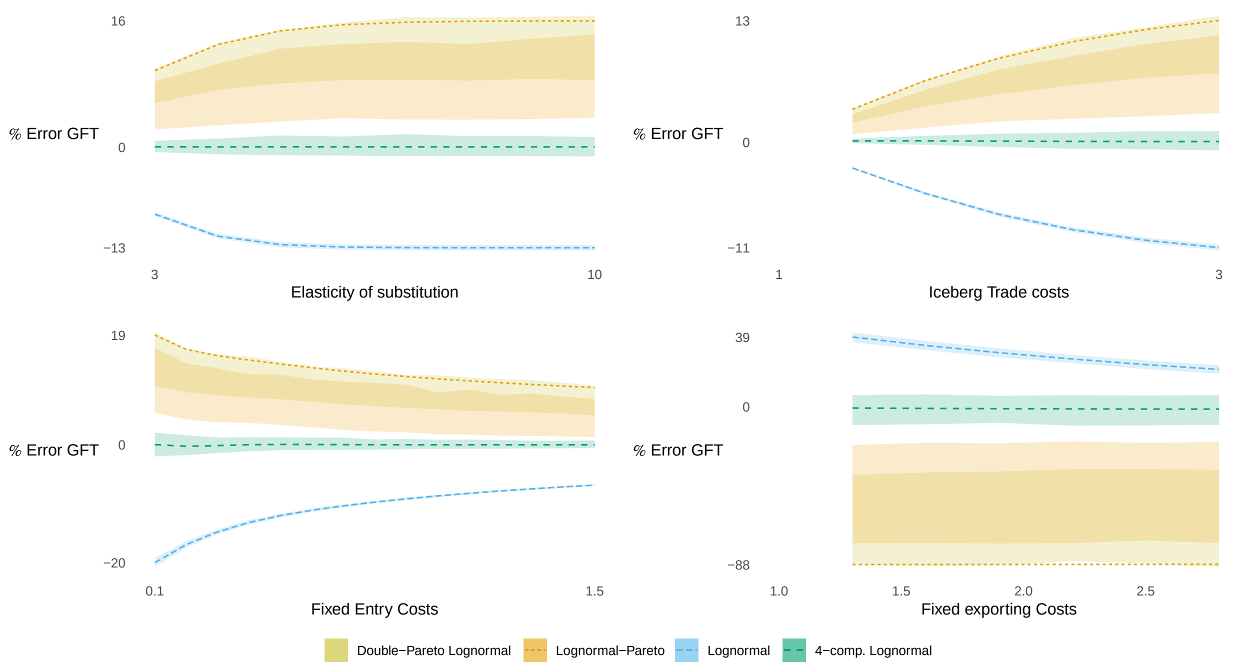

Figure 4: Percentage errors in parametric GFT calculations relative to the empirical benchmark for different values of the elasticity of substitution (left upper panel) and different fixed entry costs (left bottom panel) for a reduction in variable trade costs $\left(\tau^{ij} = 3 \to (\tau^{ij})' = 1\right)$. The right upper panel displays percentage errors in parametric GFT for different starting values of the iceberg trade costs $\left(\tau^{ij} \in [1;3] \to (\tau^{ij})' = 1\right)$. The bottom left panel showcases the error in parametric GFT for a reduction in fixed exporting costs with different starting values $\left(f^{ij} \in [1;3] \to (f^{ij})' = 1\right)$.

Notes: Full lines represent the parametric population GFT, while shaded areas delineate the 5th and 95th quantile of the parametric bootstrapped (999 replications) finite sample GFT. The Double-Pareto Lognormal has no finite population GFT value.

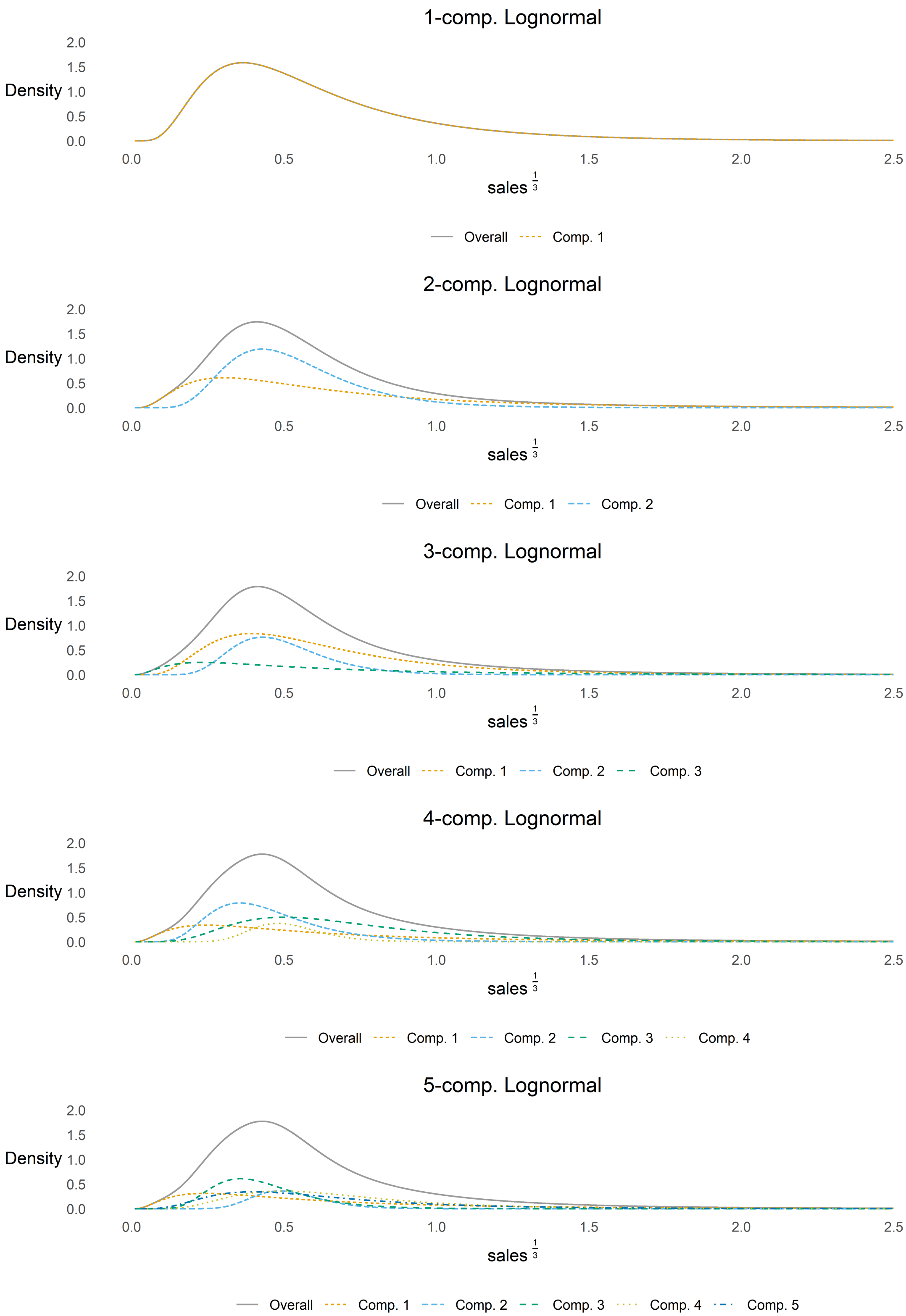

Figure 5: Probability density function of the 1- to 5-component Lognormal and its probability-weighted individual components fitted to Portuguese firm productivity in 2006.

Notes: Productivity is measured as domestic sales (relative to the mean) to the power of $1/(\sigma - 1)$ with σ, the elasticity of substitution between varieties, set to four. Distributions are fitted using maximum likelihood methods (cf. infra) to the complete dataset. For expositional purposes, the panels are restricted to productivity values between 0 and 2.5. Component ranking is not comparable across distributions.

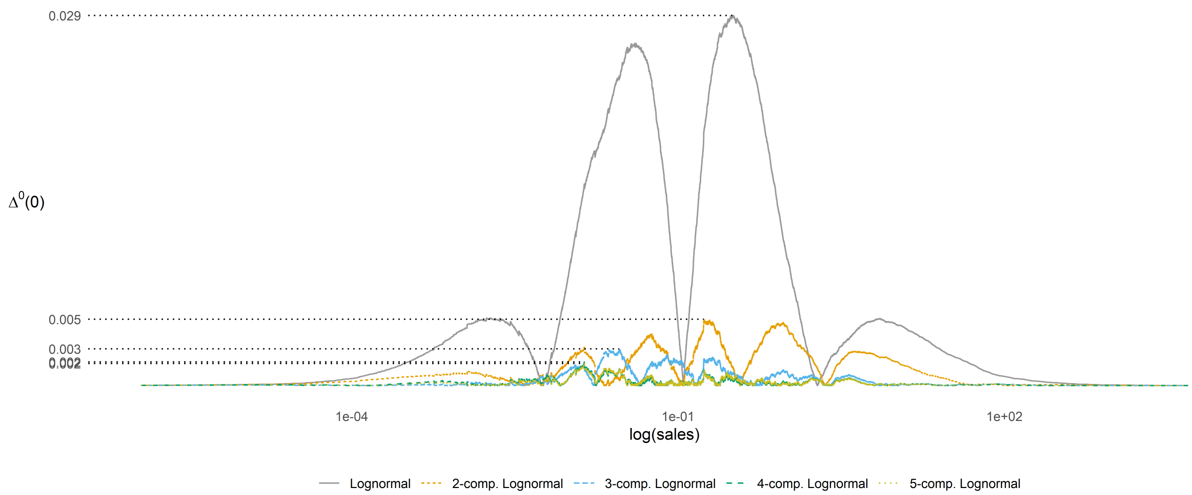

Figure 6: Normalized Absolute Deviation between the empirical and 1- to 5-component Lognormal CDFs over the complete range of domestic sales in Portugal, 2006.

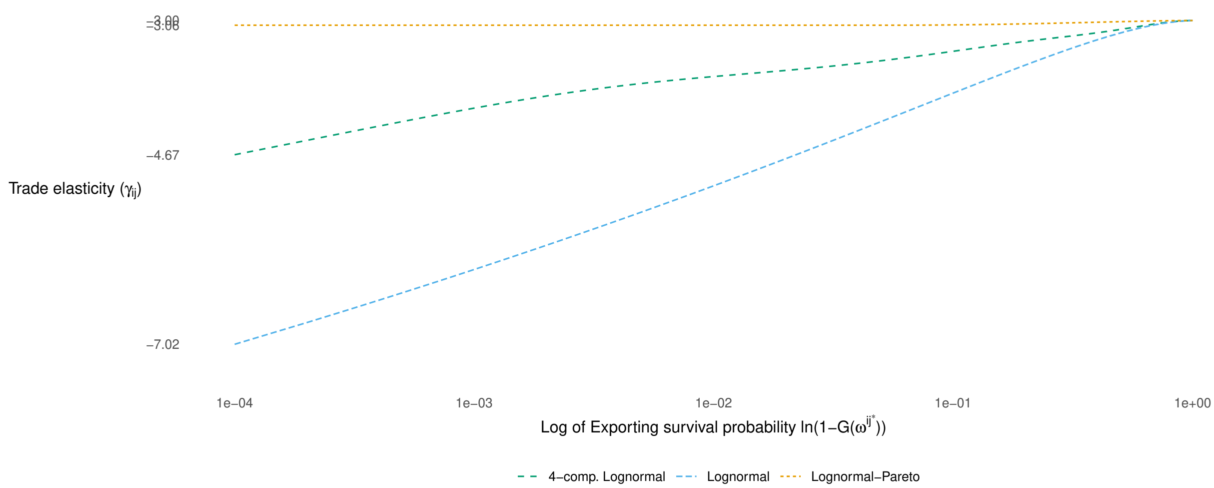

Figure 7: Trade elasticities as a function of the difficulty to reach a market for the Lognormal-pareto, Lognormal and 4-component Lognormal distribution.

Note: Trade elasticities obtained as $\gamma^{ij} = \underbrace{1-\sigma}_{\text{intensive margin}} - \underbrace{\frac{e^{(\sigma-1)\omega^{ij*}}}{\int_{\omega^{ij*}}^{\infty} e^{(\sigma-1)\omega_b} dG(\omega_b)}}_{\text{weights}} \times \underbrace{\frac{d\ln M^{ij}}{dln\tau^{ij}}}_{\text{extensive margin}}$, where $\frac{dlnM^{ij}}{dln\tau^{ij}} = \frac{e^{\omega^{ij*}} g\left(\omega^{ij*}\right)}{1-G\left(\omega^{ij*}\right)}$ with an elasticity of substitution of 4.

A.2 Tables

Table 1: Overview of all distributions considered.

Distribution	**Abbreviation**	**Support**	**Parameters**	**Change in parameters from power transformation** ax^b
Pareto	P	$[x_{min}, \infty[$	k, x_{min}	$kb, \left(\frac{x_{min}}{a}\right)^{\frac{1}{b}}$
Inverse Pareto	IP	$[0, x_{max}]$	k, x_{max}	$kb, \left(\frac{x_{max}}{a}\right)^{\frac{1}{b}}$
Lognormal	LN	$[0, \infty[$	μ, SD	$\frac{\mu - lna}{b}, \frac{SD}{b}$
Weibull	W	$[0, \infty[$	k, s	$bk, \left(\frac{s}{a}\right)^{\frac{1}{b}}$
Exponential	Exp	$[0, \infty[$	s	$W\left(b, \left(\frac{s}{a}\right)^{\frac{1}{b}}\right)$
Burr	B	$[0, \infty[$	k, c, s	$k, bc, \left(\frac{s}{a}\right)^{\frac{1}{b}}$
Fréchet	F	$[0, \infty]$	k, s	$bk, \left(\frac{s}{a}\right)^{\frac{1}{b}}$
Generalized Gamma	GG	$[0, \infty[$	k, c, s	$bk, bc, \left(\frac{s}{a}\right)^{\frac{1}{b}}$
Gamma	G	$[0, \infty[$	k, s	$GG\left(bk, b, \left(\frac{s}{a}\right)^{\frac{1}{b}}\right)$
Finite Mixture Model	FMM	See ind. comp.	$\boldsymbol{\Psi}$	See ind. comp.
Piecewise composite	PC	See ind. comp.	$\boldsymbol{\theta}$	See ind. comp.
Double-Pareto Lognormal	DPLN	$[0, \infty[$	k_1, μ, SD, k_2	$\frac{k_1}{b}, b\mu + log(a), SD, \frac{k_2}{b}$
Left-Pareto Lognormal	LPLN	$[0, \infty[$	k_1, μ, SD	$\frac{k_1}{b}, b\mu + log(a), SD$
Right-Pareto Lognormal	RPLN	$[0, \infty[$	μ, SD, k_2	$b\mu + log(a), SD, \frac{k_2}{b}$

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

Distribution	PDF	CDF
P	$\frac{kx_{min}^{k}}{x^{k+1}}$	$1-\left(\frac{x_{min}}{x}\right)^{k}$
IP	$\frac{kx_{max}^{-k}}{x^{-k+1}}$	$1-\left(\frac{x_{max}}{x}\right)^{-k}$
LN	$\frac{1}{xSD\sqrt{2\pi}}e^{-(lnx-\mu)^2/2SD^2}$	$\Phi\left(\frac{lnx-\mu}{SD}\right)$
W	$\frac{k}{s}\left(\frac{x}{s}\right)^{k-1}e^{-\left(\frac{x}{s}\right)^{k}}$	$1-e^{-\left(\frac{x}{s}\right)^{k}}$
Exp	$\frac{1}{s}e^{-\frac{x}{s}}$	$1-e^{-\frac{x}{s}}$
B	$\frac{\frac{kc}{s}\left(\frac{x}{s}\right)^{c-1}}{\left(1+\left(\frac{x}{s}\right)^{c}\right)^{k+1}}$	$1-\frac{1}{\left(1+\left(\frac{x}{s}\right)^{c}\right)^{k}}$
F	$\frac{k}{s}\left(\frac{x}{s}\right)^{-1-k}e^{-\left(\frac{x}{s}\right)^{-k}}$	$e^{-\left(\frac{x}{s}\right)^{-k}}$
GG^{a}	$\frac{c}{s^{k}\Gamma(\frac{k}{c})}x^{k-1}e^{-\left(\frac{x}{s}\right)^{c}}$	$\frac{1}{\Gamma(\frac{k}{c})}\gamma\left(\frac{k}{c},\left(\frac{x}{s}\right)^{c}\right)$
G^{a}	$\frac{1}{s^{k}\Gamma(k)}x^{k-1}e^{-\frac{x}{s}}$	$\frac{1}{\Gamma(k)}\gamma(k,\frac{x}{s})$

Notes: $^{a}\Gamma(x)$ stands for the Gamma function, while $\gamma(s,x)$ stands for the lower incomplete Gamma function with upper bound x.

Table 3: Overview of the probability and cumulative density functions of combined distributions considered.

Distribution	**PDF**	**CDF**
FMM	$\sum_{i=1}^{I} \pi_i m_i(x\|\boldsymbol{\theta}_i)$	$\sum_{i=1}^{I} \pi_i M(x\|\boldsymbol{\theta}_i)$
PC[a]	$\begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2} m_1^*(x\|\boldsymbol{\theta}_1) & \text{if} \quad 0 < x \leq c_1 \\ \frac{1}{1+\alpha_1+\alpha_2} m_2^*(x\|\boldsymbol{\theta}_2) & \text{if} \quad c_1 < x \leq c_2 \\ \frac{\alpha_2}{1+\alpha_1+\alpha_2} m_3^*(x\|\boldsymbol{\theta}_3) & \text{if} \quad c_2 < x < \infty \end{cases}$	$\begin{cases} \frac{\alpha_1}{1+\alpha_1+\alpha_2} \frac{M_1(x\|\boldsymbol{\theta}_1)}{M_1(c_1\|\boldsymbol{\theta}_1)} & \text{if} \quad 0 < x \leq c_1 \\ \frac{\alpha_1}{1+\alpha_1+\alpha_2} + \frac{1}{1+\alpha_1+\alpha_2} \frac{M_2(x\|\boldsymbol{\theta}_2)-M_2(c_1\|\boldsymbol{\theta}_2)}{M_2(c_2\|\boldsymbol{\theta}_2)-M_2(c_1\|\boldsymbol{\theta}_2)} & \text{if} \quad c_1 < x \leq c_2 \\ \frac{1+\alpha_1}{1+\alpha_1+\alpha_2} + \frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{M_3(x\|\boldsymbol{\theta}_3)-M_3(c_2\|\boldsymbol{\theta}_3)}{1-M_3(c_2\|\boldsymbol{\theta}_3)} & \text{if} \quad c_2 < x < \infty \end{cases}$
DPLN[b]	$\frac{k_2 k_1}{k_2+k_1}\Big[x^{-k_2-1} e^{k_2\mu+\frac{k_2^2 SD^2}{2}} \Phi\left(\frac{lnx-\mu-k_2 SD^2}{SD}\right) + x^{k_1-1} e^{-k_1\mu+\frac{k_1^2 SD^2}{2}} \Phi^c\left(\frac{lnx-\mu+k_1 SD^2}{SD}\right)\Big]$	$\Phi\left(\frac{lnx-\mu}{SD}\right) - \frac{1}{k_2+k_1}\Big[k_1 x^{-k_2} e^{k_2\mu+\frac{k_2^2 SD^2}{2}} \Phi\left(\frac{lnx-\mu-k_2 SD^2}{SD}\right) - k_2 x^{k_1} e^{-k_1\mu+\frac{k_1^2 SD^2}{2}} \Phi^c\left(\frac{lnx-\mu+k_1 SD^2}{SD}\right)\Big]$
LPLN[b]	$k_1 x^{k_1-1} e^{-k_1\mu+\frac{k_1^2 SD^2}{2}} \Phi^c\left(\frac{lnx-\mu+k_1 SD^2}{SD}\right)$	$\Phi\left(\frac{lnx-\mu}{SD}\right) - x^{k_1} e^{-k_1\mu+\frac{k_1^2 SD^2}{2}} \Phi^c\left(\frac{lnx-\mu+k_1 SD^2}{SD}\right)$
RPLN[b]	$k_2 x^{-k_2-1} e^{k_2\mu+\frac{k_2^2 SD^2}{2}} \Phi\left(\frac{lnx-\mu-k_2 SD^2}{SD}\right)$	$\Phi\left(\frac{lnx-\mu}{SD}\right) - x^{-k_2} e^{k_2\mu+\frac{k_2^2 SD^2}{2}} \Phi\left(\frac{lnx-\mu-k_2 SD^2}{SD}\right)$

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 4: Expression of the y-bounded rth moment (μ_y^r) for the single distributions considered.

Distribution	μ_y^r	**Additional parameter restrictions**[a]
P	$-\left(y\right)^{r-k}\frac{k\omega_{min}^{k}}{r-k}$	$k>r$
IP	$k\omega_{max}^{-k}\frac{(\omega_{max})^{r+k}-(y)^{r+k}}{r+k}$	-
LN	$e^{\frac{r\left(rSD^2+2\mu\right)}{2}}\left[1-\Phi\left(\frac{lny-\left(rSD^2+\mu\right)}{SD}\right)\right]$	-
W[c]	$s^{\sigma_s-1}\Gamma\left(\frac{\sigma_s-1}{k}+1,\left(\frac{y}{s}\right)^k\right)$	-
Exp[c]	$s^{\sigma_s-1}\Gamma\left(\sigma_s+1,\frac{y}{s}\right)$	-
B[b]	$s^r k\left[\boldsymbol{B}\left(\frac{r}{c}+1,k-\frac{r}{c}\right)-\boldsymbol{B}\left(\frac{\left(\frac{y}{s}\right)^c}{1+\left(\frac{y}{s}\right)^c};\frac{r}{c}+1,k-\frac{r}{c}\right)\right]$	$c>r,\ kc>r$
F[c]	$s^{\sigma_s-1}\left[1-\Gamma\left(1-\frac{\sigma_s-1}{k},\left(\frac{y}{s}\right)^{-k}\right)\right]$	$k>r$
GG[c]	$\frac{s^{\sigma_s-1}}{\Gamma\left(\frac{k}{c}\right)}\Gamma\left(\frac{\sigma_s-1+k}{c},\left(\frac{y}{s}\right)^c\right)$	-
G[c]	$\frac{s^{\sigma_s-1}}{\Gamma(k)}\Gamma\left(\sigma_s-1+k,\frac{y}{s}\right)$	-

Notes: [a] Additional parameter restrictions represent parameter restrictions needed to keep the statistic finite. [b] $\boldsymbol{B}(a,b)$ stands for the beta function, while $\boldsymbol{B}(x,a,b)$ stands for the lower incomplete beta function with upper bound x. [c] $\Gamma(x)$ stands for the Gamma function, while $\Gamma(s,x)$ stands for the upper incomplete Gamma function with lower bound x.

Table 5: Expression of the y-bounded rth moment (μ_y^r) for the combined considered.

Distribution	μ_y^r	**Additional parameter restrictions**[a]
FMM	$\sum_{i=1}^{I}\pi_i(\mu_i)_y^r$	See ind. comp.
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	$-\frac{k_2k_1}{k_2+k_1}e^{k_2\mu+\frac{k_2^2SD^2}{2}}\frac{y^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{lny-\mu-k_2SD^2}{SD}\right) - \frac{k_2k_1}{k_2+k_1}\frac{1}{r-k_2}e^{\frac{r^2SD^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rSD^2-\mu}{SD}\right) - \frac{k_2k_1}{k_2+k_1}e^{-k_1\mu+\frac{k_1^2SD^2}{2}}\frac{y^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{lny-\mu+k_1SD^2}{SD}\right) + \frac{k_2k_1}{k_2+k_1}\frac{1}{r+k_1}e^{\frac{r^2SD^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rSD^2-\mu}{SD}\right)$	$k_2>r$
LPLN	$-k_1e^{-k_1\mu+\frac{k_1^2SD^2}{2}}\frac{y^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{lny-\mu+k_1SD^2}{SD}\right) + \frac{k_1}{r+k_1}e^{\frac{r^2SD^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rSD^2+\mu}{SD}\right)$	-
RPLN	$-k_2e^{k_2\mu+\frac{k_2^2SD^2}{2}}\frac{y^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{lny-\mu-k_2SD^2}{SD}\right) - \frac{k_2}{r-k_2}e^{\frac{r^2SD^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rSD^2+\mu}{SD}\right)$	$k_2>r$

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

Table 6: Coverage ratio of SCIE vs OECD SDBS database.

	Number of Enterprises						Total Employment						Turnover					
NACE Rev.2	1-9	10-19	20-49	50-249	> 250	Total	1-9	10-19	20-49	50-249	> 250	Total	1-9	10-19	20-49	50-249	> 250	Total
13	100				100	100												
14	100	100	100	100		100		100	100					100	100			
15	100	100	100	100	100	100	100	100	100			100	100	100	100			100
16				100	100	100						100						100
17	100	100	100	100	100	100				100	100	100				100	100	100
18	100	100	100	100	100	100				100	100	100				100	100	100
19	100	100	100	100	100	100		100	100	100				100	100	100		
20	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
21	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
22	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
23					100	100												
24	100	100	100	100	100	100				100	100					100	100	
25	100	100	100	100	100	100				100	100	100				100	100	100
26	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
27	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
28	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
29	100	100	100	100	100	100	100			100	100	100	100			100	100	100
30	100	100	100	100	100	100												
31	100	100	100	100	100	100			100	100	100	100			100	100	100	100
32	100	100	100	100	100	100		100	100	100	100			100	100	100	100	
33	100	100	100	100	100	100	100	100	100				100	100	100			
34	100	100	100	100	100	100			100	100	100				100	100	100	
35	100	100	100	100	100	100			100	100	100				100	100	100	
36	100	100	100	100	100	100		100		100	100	100		100		100	100	100
37	100	100	100	100		100				100		100				100		100
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
41	100	100	100	100	100	100	100	100	100			100	100	100	100			100
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
51	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
52	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
55	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
61	100	100	100	100	100	100		100		100		100		100		100		100
62	100	100	100	100	100	100			100	100		100			100	100		100
63	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
64	100	100	100	100	100	100	100			100	100	100	100			100	100	100
70	100	100	100	100	100	100	100	100	100			100	100	100	100			100
71	100	100	100	100	100	100			100			100			100			100
72	100	100	100	100	100	100	100			100	100	100	100			100	100	100
73	100	100	100	100		100						100						100
74	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Notes: Each cell corresponds to the ratio of our dataset compared to the data from the OECD structural SDBS database for the year 2006. Size classes are based on total employment. Empty cells and absent industries are due to missing information from SBDS, even though the data is available in our SCIE database.

Table 7: Distribution fits to Portuguese domestic sales in 2006.

		Goodness of fit				Information Criteria		
Distribution	Parms.	T_a^0	S_b^0	T_a^1	S_b^1	Loglike	R_{AIC}	R_{BIC}
5-comp. Lognormal	14	0.18	0.11	3.08	2.37	12,776	1	3^{+++}
		(0.10;0.25)	(0.08;0.32)	(2.03;9.73)	(0.82;26.24)			
4-comp. Lognormal	11	0.19	0.11	2.78	0.13	12,770	2	2
		(0.09;0.25)	(0.08;0.32)	(2.07;9.23)	(0.83;24.67)			
5-comp. Burr	19	0.19	0.12	-	-	12,767	3	7^{+++}
		(0.10;0.25)	(0.08;0.32)	(-;-)	(-;-)			
4-comp. Burr	15	0.24	0.14	-	-	12,754	4	6^{+++}
		(0.10;0.25)*	(0.08;0.32)	(-;-)	(-;-)			
3-comp. Burr	11	0.25	0.17	-	-	12,748	6	4^{+++}
		(0.09;0.25)*	(0.08;0.30)	(-;-)	(-;-)			
2-comp. Burr	7	0.20	0.20	-	-	12,745	5	1
		(0.09;0.25)	(0.08;0.32)	(-;-)	(-;-)			
5-comp. Weibull	14	0.25	0.14	6.96	11.95	12,731	7	8^{+++}
		(0.10;0.25)**	(0.08;0.31)	(1.29;5.00)***	(0.59;13.45)*			
3-comp. Lognormal	8	0.29	0.34	4.39	9.91	12,723	8	5^{+++}
		(0.10;0.24)**	(0.09;0.32)**	(2.34;11.34)	(0.93;30.68)			
5-comp. Gamma	14	0.26	0.16	7.27	0.09	12,639	9	9^{+++}
		(0.10;0.26)**	(0.09;0.33)	(1.29;5.11)***	(0.44;14.23)			
Inv. Pareto-Burr	4	0.51	0.61	-	-	12,561	10	10^{+++}
		(0.09;0.24)***	(0.08;0.33)***	(-;-)	(-;-)			
Inv. Pareto-Burr-Pareto	5	0.51	0.61	-	-	12,561	11	11^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			
5-comp. Exponential	9	0.32	0.23	7.96	0.15	12,548	12	12^{+++}

		(0.09;0.26)***	(0.09;0.31)	(1.31;4.78)***	(0.40;12.83)			
4-comp. Weibull	11	0.31	0.25	14.75	27.04	12,543	13	13^{+++}
		(0.09;0.25)***	(0.08;0.31)	(0.87;3.44)***	(0.29;8.78)***			
Burr-Pareto	4	0.73	0.95	-	-	12,451	15	15^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			
Burr	3	0.73	0.95	-	-	12,451	14	14^{+++}
		(0.10;0.24)***	(0.08;0.31)***	(-;-)	(-;-)			
Double-Pareto Lognormal	4	0.66	0.80	-	-	12,429	16	16^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			
2-comp. Lognormal	5	0.53	0.71	8.70	10.15	12,401	17	17^{+++}
		(0.10;0.24)***	(0.09;0.32)***	(1.32;5.87)**	(0.54;16.11)			
Inv. Pareto-Lognormal-Pareto	4	0.81	1.01	-	-	12,231	18	18^{+++}
		(0.09;0.26)***	(0.08;0.34)***	(-;-)	(-;-)			
4-comp. Gamma	11	0.40	0.63	11.95	0.26	12,173	19	19^{+++}
		(0.10;0.25)***	(0.08;0.32)***	(1.00;3.92)***	(0.26;10.38)			
Inv. Pareto-Fréchet-Pareto	4	1.11	1.48	-	-	11,953	20	20^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			
3-comp. Weibull	8	0.69	0.92	20.31	39.45	11,855	21	21^{+++}
		(0.10;0.25)***	(0.09;0.31)***	(0.73;2.60)***	(0.23;6.78)***			
4-comp. Exponential	7	0.57	0.89	13.91	0.36	11,801	22	22^{+++}
		(0.10;0.25)***	(0.09;0.32)***	(0.95;3.61)***	(0.34;9.44)			
Inv. Pareto-Weibull-Pareto	4	1.60	2.00	-	-	11,338	24	24^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
Weibull-Pareto	3	1.60	2.00	-	-	11,338	23	23^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
Inv. Pareto-Gamma-Pareto	4	1.70	2.17	-	-	11,249	26	26^{+++}
		(0.10;0.26)***	(0.08;0.35)***	(-;-)	(-;-)			
Gamma-Pareto	3	1.70	2.17	-	-	11,249	25	25^{+++}

		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
Inv. Pareto-Exponential-Pareto	3	1.97	2.71	-	-	11,044	27	27^{+++}
		(0.10;0.25)***	(0.09;0.33)***	(-;-)	(-;-)			
Exponential-Pareto	2	2.00	2.83	-	-	11,012	28	28^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
3-comp. Gamma	8	1.00	1.56	19.47	0.62	10,288	29	29^{+++}
		(0.10;0.25)***	(0.09;0.32)***	(0.73;2.75)***	(0.30;7.18)			
Inv. Pareto-Lognormal	3	3.02	4.26	45.72	127.97	9,198	30	30^{+++}
		(0.09;0.24)***	(0.08;0.31)***	(0.43;1.67)***	(0.16;4.36)***			
Lognormal-Pareto	3	2.56	3.78	562.07	1683.18	8,721	31	31^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(169.39;451.06)***	(371.67;1342.23)***			
3-comp. Exponential	5	1.64	2.60	22.73	0.87	8,387	32	32^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(0.67;2.36)***	(0.21;6.27)			
Left-Pareto Lognormal	3	3.23	4.91	46.00	127.70	8,059	33	33^{+++}
		(0.10;0.25)***	(0.09;0.32)***	(0.41;1.58)***	(0.18;4.05)***			
Right-Pareto Lognormal	3	2.82	4.38	19.27	49.88	8,028	34	34^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(3.10;11.90)**	(1.23;32.23)**			
Lognormal	2	2.93	5.03	41.38	113.04	7,372	35	35^{+++}
		(0.10;0.25)***	(0.08;0.33)***	(0.47;1.84)***	(0.17;4.76)***			
2-comp. Weibull	5	2.10	3.19	35.20	72.85	6,442	36	36^{+++}
		(0.09;0.24)***	(0.08;0.31)***	(0.42;1.50)***	(0.16;3.80)***			
2-comp. Fréchet	5	6.92	10.64	-	-	-3,041	37	37^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
5-comp. Fréchet	14	6.96	10.63	-	-	-3,045	40	40^{+++}
		(0.10;0.25)***	(0.09;0.31)***	(-;-)	(-;-)			
3-comp. Fréchet	8	6.96	10.63	-	-	-3,046	38	38^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(-;-)	(-;-)			
4-comp. Fréchet	11	6.98	10.63	-	-	-3,047	39	39^{+++}

		(0.10;0.25)***	(0.09;0.32)***	(-;-)	(-;-)			
2-comp. Gamma	5	4.00	5.93	31.79	2.24	-3,381	41	41^{+++}
		(0.09;0.25)***	(0.08;0.32)***	(0.45;1.67)***	(0.14;4.45)			
2-comp. Exponential	3	7.06	11.51	37.63	3.23	-18,112	42	42^{+++}
		(0.10;0.25)***	(0.08;0.33)***	(0.38;1.40)***	(0.14;3.52)*			
Inv. Pareto-Weibull	3	9.18	16.52	54.06	123.38	-29,711	43	44^{+++}
		(0.10;0.25)***	(0.08;0.31)***	(0.26;0.92)***	(0.11;2.22)***			
Weibull	2	9.18	16.51	54.06	123.40	-29,713	44	43^{+++}
		(0.09;0.25)***	(0.09;0.32)***	(0.25;0.90)***	(0.15;2.20)***			
Fréchet	2	8.91	16.72	-	-	-32,908	45	45^{+++}
		(0.10;0.26)***	(0.08;0.33)***	(-;-)	(-;-)			
Fréchet-Pareto	3	8.91	16.72	-	-	-32,908	46	46.5^{+++}
		(0.10;0.25)***	(0.08;0.31)***	(-;-)	(-;-)			
Inv. Pareto-Fréchet	3	8.91	16.72	-	-	-32,908	46	46.5^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			
Inv. Pareto-Gamma	3	20.93	32.98	50.26	9.56	-104,785	48	48^{+++}
		(0.10;0.25)***	(0.08;0.33)***	(0.22;0.76)***	(0.13;1.81)***			
Gamma	2	20.98	33.03	50.29	9.58	-104,878	49	49^{+++}
		(0.10;0.25)***	(0.08;0.32)***	(0.22;0.76)***	(0.11;1.71)***			
Exponential	1	44.64	79.71	60.73	16.76	-299,935	50	50^{+++}
		(0.10;0.25)***	(0.09;0.33)***	(0.15;0.49)***	(0.11;1.08)***			
Inv. Pareto-Exponential	2	44.64	79.71	60.73	16.76	-299,935	51	51^{+++}
		(0.09;0.24)***	(0.08;0.32)***	(0.15;0.51)***	(0.11;1.12)***			
Pareto	2	48.34	68.18	-	-	-436,227	52	52^{+++}
		(0.09;0.25)***	(0.08;0.33)***	(-;-)	(-;-)			

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.

+++, ++, + indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favour of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.

$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Table 8: Coefficients for selected distribution fits to Portuguese domestic sales in 2006.

Distribution	Parms.	Priors	Coefficients
Double-Pareto Lognormal	4	π_1=1.00	k_1=0.99, μ=-2.08, SD=0.94, k_2=0.92
Inv. Pareto-Lognormal	3	π_1=1.00	k_1=1.03, μ_2=-1.95, SD_2=1.68
Inv. Pareto-Lognormal-Pareto	4	π_1=1.00	k_1=0.96, μ_2=-2.04, SD_2=1.44, k_3=0.89
Left-Pareto Lognormal	3	π_1=1.00	k_1=1.44, μ=-1.31, SD=1.60
Lognormal	2	π_1=1.00	μ=-2.00, SD=1.75
2-comp. Lognormal	5	π_1=0.46	μ=-1.91, SD=2.24
		π_2=0.54	μ=-2.08, SD=1.19
3-comp. Lognormal	8	π_1=0.54	μ=-1.84, SD=1.69
		π_2=0.27	μ=-2.23, SD=0.96
		π_3=0.18	μ=-2.14, SD=2.60
4-comp. Lognormal	11	π_1=0.24	μ=-2.14, SD=2.51
		π_2=0.31	μ=-2.62, SD=1.21
		π_3=0.35	μ=-1.35, SD=1.48
		π_4=0.10	μ=-2.06, SD=0.68
5-comp. Lognormal	14	π_1=0.23	μ=-2.14, SD=2.53
		π_2=0.10	μ=-2.02, SD=0.66
		π_3=0.21	μ=-2.68, SD=1.10
		π_4=0.23	μ=-1.45, SD=1.43
		π_5=0.23	μ=-1.79, SD=1.70
Lognormal-Pareto	3	π_1=1.00	μ_1=-2.06, SD_1=1.68, k_2=1.02
Pareto	2	π_1=1.00	x_{min}=0.00, k=0.09
Right-Pareto Lognormal	3	π_1=1.00	μ=-2.72, SD=1.59, k_2=1.39

Notes: All distributions fitted using Maximum Likelihood.

Table 9: Selected distribution fits to the tails of Portuguese domestic sales in 2006.

Distribution	Parms.	Goodness of fit T^0_a	S^0_b	Information Criteria Loglike	R_{AIC}	R_{BIC}
			Left tail (N=25,588, 8.53% of the data)			
5-comp. Trunc. Lognormal	14	0.63	0.04	108,196.19***	5	6^{+++}
		(0.32;0.85)	(0.02;0.10)			
4-comp. Trunc. Lognormal	11	0.61	0.04	108,195.05***	4	5^{+++}
		(0.33;0.85)	(0.02;0.09)			
3-comp. Trunc. Lognormal	8	0.58	0.04	108,194.44***	1	4^{+++}
		(0.33;0.86)	(0.02;0.10)			
2-comp. Trunc. Lognormal	5	0.77	0.06	108,189.93***	3	3^{+++}
		(0.32;0.84)*	(0.02;0.09)			
Trunc. Lognormal	2	1.02	0.10	108,186.99***	2	1
		(0.32;0.85)**	(0.02;0.09)**			
Inv. Pareto	2	0.80	0.10	108,183.90	6	2^{++}
		(0.33;0.84)*	(0.02;0.09)**			
			Right tail (N=18,217, 6.07% of the data)			
5-comp. Trunc. Lognormal	14	0.62	0.03	-47,896.59***	5	6^{+++}
		(0.39;1.00)	(0.02;0.07)			
Trunc. Lognormal	2	0.70	0.04	-47,897.86***	1	1
		(0.38;0.97)	(0.02;0.08)			
2-comp. Trunc. Lognormal	5	0.71	0.04	-47,897.99***	2	3^{+++}
		(0.38;1.01)	(0.02;0.08)			
3-comp. Trunc. Lognormal	8	0.68	0.04	-47,898.60***	3	4^{+++}
		(0.38;0.99)	(0.02;0.08)			
4-comp. Trunc. Lognormal	11	0.68	0.04	-47,898.62***	4	5^{+++}
		(0.39;1.00)	(0.02;0.08)			
Pareto	2	0.86	0.08	-47,910.44	6	2^{+++}
		(0.38;0.99)	(0.02;0.08)*			

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.
Similarly, ***, **, * indicate significance at 1%, 5% and 10% respectively for the likelihood ratio test between (Inverse) Pareto and (mixtures of) the Lognormal distribution.
$^{+++}$, $^{++}$, $^{+}$ indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favor of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.
$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Table 10: Distribution fits to domestic sales of the Portuguese manufacturing sector in 2006.

		Goodness of fit		Information Criteria		
Distribution	Parms.	T_a^0	S_b^0	Loglike	R_{AIC}	R_{BIC}
5-comp. Burr	19	0.24	0.02	-2,095	7	10^{+++}
		(0.25;0.66)	(0.03;0.12)			
5-comp. Lognormal	14	0.28	0.03	-2,095	3	5^{+++}
		(0.25;0.67)	(0.03;0.12)			
4-comp. Burr	15	0.24	0.02	-2,096	5	6^{+++}
		(0.25;0.67)	(0.03;0.12)			
3-comp. Burr	11	0.23	0.03	-2,099	4	3^{+++}
		(0.25;0.67)	(0.03;0.12)			
2-comp. Burr	7	0.22	0.02	-2,099	1	1
		(0.25;0.66)	(0.03;0.12)			
3-comp. Lognormal	8	0.28	0.02	-2,101	2	2^{+++}
		(0.25;0.69)	(0.03;0.12)			
4-comp. Lognormal	11	0.27	0.02	-2,101	6	4^{+++}
		(0.25;0.67)	(0.03;0.12)			
5-comp. Weibull	14	0.34	0.03	-2,104	8	7^{+++}
		(0.26;0.65)	(0.03;0.11)			
5-comp. Gamma	14	0.31	0.04	-2,114	9	8^{+++}
		(0.26;0.66)	(0.03;0.12)			
4-comp. Weibull	11	0.40	0.04	-2,131	10	9^{+++}
		(0.26;0.65)	(0.03;0.11)			
5-comp. Exponential	9	1.29	0.15	-2,171	11	13^{+++}
		(0.25;0.66)***	(0.03;0.12)***			
4-comp. Gamma	11	0.50	0.09	-2,178	12	15^{+++}
		(0.26;0.65)	(0.03;0.12)			
Inv. Pareto-Fréchet-Pareto	4	0.65	0.09	-2,187	13	11^{+++}
		(0.25;0.65)*	(0.03;0.12)			
Inv. Pareto-Burr	4	0.88	0.13	-2,197	14	12^{+++}
		(0.26;0.65)***	(0.03;0.12)**			
Inv. Pareto-Burr-Pareto	5	0.88	0.13	-2,197	15	14^{+++}
		(0.25;0.69)***	(0.03;0.12)**			
3-comp. Weibull	8	0.83	0.14	-2,222	16	17^{+++}
		(0.25;0.66)***	(0.03;0.11)**			
2-comp. Lognormal	5	0.73	0.11	-2,232	17	16^{+++}
		(0.26;0.67)**	(0.03;0.12)*			

Double-Pareto Lognormal	4	1.08	0.17	-2,245	18	18^{+++}
		(0.26;0.66)***	(0.03;0.12)***			
4-comp. Exponential	7	1.18	0.16	-2,251	19	20^{+++}
		(0.26;0.66)***	(0.03;0.12)***			
Inv. Pareto-Lognormal-Pareto	4	1.27	0.18	-2,263	20	19^{+++}
		(0.25;0.66)***	(0.03;0.11)***			
Burr-Pareto	4	1.18	0.25	-2,284	22	22^{+++}
		(0.26;0.67)***	(0.03;0.12)***			
Burr	3	1.18	0.25	-2,284	21	21^{+++}
		(0.26;0.67)***	(0.03;0.12)***			
Inv. Pareto-Gamma-Pareto	4	1.65	0.28	-2,346	23	24^{+++}
		(0.25;0.65)***	(0.03;0.12)***			
Inv. Pareto-Weibull-Pareto	4	1.62	0.28	-2,348	24	25^{+++}
		(0.25;0.68)***	(0.03;0.12)***			
Gamma-Pareto	3	1.58	0.27	-2,355	26	26^{+++}
		(0.25;0.68)***	(0.03;0.12)***			
Weibull-Pareto	3	1.58	0.27	-2,355	27	27^{+++}
		(0.27;0.67)***	(0.03;0.11)***			
Exponential-Pareto	2	1.57	0.27	-2,355	25	23^{+++}
		(0.26;0.68)***	(0.03;0.12)***			
Inv. Pareto-Exponential-Pareto	3	1.57	0.27	-2,355	28	28^{+++}
		(0.26;0.67)***	(0.03;0.12)***			
3-comp. Gamma	8	1.01	0.20	-2,408	29	29^{+++}
		(0.26;0.68)***	(0.03;0.12)***			
3-comp. Exponential	5	1.44	0.26	-2,608	30	30^{+++}
		(0.25;0.65)***	(0.03;0.12)***			
Inv. Pareto-Lognormal	3	4.18	0.81	-2,875	31	31^{+++}
		(0.26;0.65)***	(0.03;0.12)***			
2-comp. Weibull	5	2.19	0.43	-2,918	32	32^{+++}
		(0.25;0.66)***	(0.03;0.12)***			
Lognormal-Pareto	3	3.25	0.64	-3,051	33	33^{+++}
		(0.25;0.67)***	(0.03;0.12)***			
Left-Pareto Lognormal	3	4.39	0.89	-3,103	34	34^{+++}
		(0.25;0.65)***	(0.03;0.11)***			
Right-Pareto Lognormal	3	3.51	0.73	-3,143	35	35^{+++}
		(0.26;0.66)***	(0.03;0.12)***			
Lognormal	2	3.96	0.88	-3,250	36	36^{+++}
		(0.25;0.65)***	(0.03;0.11)***			

2-comp. Gamma	5	3.35	0.71	-4,108	37	37^{+++}
		(0.25;0.65)***	(0.03;0.12)***			
5-comp. Fréchet	14	8.11	1.79	-4,863	38	40^{+++}
		(0.26;0.67)***	(0.03;0.12)***			
4-comp. Fréchet	11	8.35	1.80	-4,870	39	39^{+++}
		(0.25;0.66)***	(0.03;0.12)***			
3-comp. Fréchet	8	8.59	1.82	-4,881	40	38^{+++}
		(0.26;0.67)***	(0.03;0.12)***			
2-comp. Fréchet	5	9.55	1.92	-4,955	41	41^{+++}
		(0.25;0.67)***	(0.03;0.12)***			
2-comp. Exponential	3	5.75	1.20	-5,550	42	42^{+++}
		(0.25;0.67)***	(0.03;0.12)***			
Inv. Pareto-Weibull	3	9.91	2.42	-8,321	44	44^{+++}
		(0.26;0.65)***	(0.03;0.12)***			
Weibull	2	9.91	2.42	-8,321	43	43^{+++}
		(0.26;0.67)***	(0.03;0.11)***			
Inv. Pareto-Fréchet	3	10.04	2.61	-9,885	46	46^{+++}
		(0.26;0.68)***	(0.03;0.12)***			
Fréchet-Pareto	3	10.04	2.61	-9,885	47	47^{+++}
		(0.25;0.69)***	(0.03;0.13)***			
Fréchet	2	10.04	2.61	-9,885	45	45^{+++}
		(0.25;0.65)***	(0.03;0.11)***			
Inv. Pareto-Gamma	3	20.68	4.43	-17,309	48	48^{+++}
		(0.26;0.67)***	(0.03;0.11)***			
Gamma	2	20.72	4.44	-17,318	49	49^{+++}
		(0.25;0.67)***	(0.03;0.12)***			
Exponential	1	43.48	10.42	-41,128	50	50^{+++}
		(0.27;0.66)***	(0.03;0.12)***			
Inv. Pareto-Exponential	2	43.48	10.42	-41,128	51	51^{+++}
		(0.26;0.65)***	(0.03;0.11)***			
Pareto	2	49.14	9.43	-66,043	52	52^{+++}
		(0.26;0.65)***	(0.03;0.11)***			

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.

$^{+++}$, $^{++}$, $^{+}$ indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favour of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.

$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Table 11: Distribution fits to Portuguese domestic sales leaving out the first and last 1,000 observations in 2006.

		Goodness of fit		Information Criteria		
Distribution	Parms.	T_a^0	S_b^0	Loglike	R_{AIC}	R_{BIC}
4-comp. Lognormal	11	0.18	0.18	23,100	1	1
		(0.09;0.24)	(0.09;0.32)			
5-comp. Lognormal	14	0.21	0.20	23,093	2	2^{+++}
		(0.09;0.24)	(0.08;0.30)			
3-comp. Lognormal	8	0.25	0.28	22,844	3	3^{+++}
		(0.10;0.25)*	(0.09;0.31)*			
5-comp. Weibull	14	0.33	0.27	22,764	4	4^{+++}
		(0.09;0.25)***	(0.08;0.32)			
5-comp. Gamma	14	0.36	0.29	22,758	5	5^{+++}
		(0.09;0.24)***	(0.08;0.31)*			
4-comp. Gamma	11	0.40	0.30	22,724	6	7^{+++}
		(0.09;0.24)***	(0.08;0.31)*			
2-comp. Lognormal	5	0.29	0.26	22,695	7	6^{+++}
		(0.10;0.25)**	(0.09;0.31)			
4-comp. Weibull	11	0.40	0.29	22,691	8	8^{+++}
		(0.10;0.25)***	(0.08;0.31)*			
4-comp. Exponential	7	0.60	0.34	22,544	9	9^{+++}
		(0.10;0.25)***	(0.08;0.33)**			
5-comp. Exponential	9	0.59	0.36	22,541	10	10^{+++}
		(0.09;0.25)***	(0.08;0.32)**			
3-comp. Burr	11	0.30	0.38	22,477	11	11^{+++}
		(0.09;0.26)***	(0.08;0.32)**			
3-comp. Weibull	8	0.40	0.66	22,247	12	12^{+++}
		(0.09;0.24)***	(0.08;0.31)***			
5-comp. Fréchet	14	0.67	0.56	22,240	13	13^{+++}
		(0.09;0.25)***	(0.08;0.32)***			
3-comp. Gamma	8	0.56	0.88	22,132	14	14^{+++}
		(0.10;0.25)***	(0.09;0.32)***			
4-comp. Fréchet	11	0.68	0.66	22,056	15	16^{+++}
		(0.09;0.26)***	(0.08;0.32)***			
3-comp. Exponential	5	0.64	1.00	22,025	16	15^{+++}
		(0.10;0.25)***	(0.08;0.33)***			
3-comp. Fréchet	8	0.67	0.65	21,911	17	17^{+++}

		(0.09;0.25)***	(0.09;0.33)***			
2-comp. Burr	7	0.80	0.76	21,614	22	22^{+++}
		(0.09;0.25)***	(0.08;0.30)***			
Inv. Pareto-Burr-Pareto	5	0.80	0.76	21,614	21	21^{+++}
		(0.09;0.24)***	(0.08;0.30)***			
Inv. Pareto-Burr	4	0.80	0.76	21,614	19	19^{+++}
		(0.09;0.25)***	(0.08;0.32)***			
Burr	3	0.80	0.76	21,614	18	18^{+++}
		(0.10;0.25)***	(0.08;0.32)***			
Burr-Pareto	4	0.80	0.76	21,614	20	20^{+++}
		(0.09;0.24)***	(0.08;0.31)***			
5-comp. Burr	19	0.80	0.76	21,614	23	24^{+++}
		(0.09;0.25)***	(0.08;0.31)***			
Double-Pareto Lognormal	4	1.04	1.36	21,592	24	23^{+++}
		(0.10;0.25)***	(0.09;0.33)***			
Inv. Pareto-Lognormal-Pareto	4	1.18	1.51	21,179	25	25^{+++}
		(0.10;0.25)***	(0.09;0.33)***			
Inv. Pareto-Lognormal	3	2.48	3.35	20,614	26	26^{+++}
		(0.10;0.25)***	(0.09;0.33)***			
Right-Pareto Lognormal	3	2.02	3.25	20,585	27	27^{+++}
		(0.10;0.25)***	(0.08;0.33)***			
Lognormal-Pareto	3	1.85	3.01	20,494	28	28^{+++}
		(0.10;0.25)***	(0.09;0.32)***			
Left-Pareto Lognormal	3	2.49	3.70	20,423	29	29^{+++}
		(0.09;0.25)***	(0.08;0.33)***			
Lognormal	2	2.35	3.68	20,407	30	30^{+++}
		(0.10;0.25)***	(0.09;0.33)***			
Inv. Pareto-Fréchet-Pareto	4	1.46	2.06	20,193	31	31^{+++}
		(0.10;0.25)***	(0.08;0.32)***			
Inv. Pareto-Weibull-Pareto	4	1.95	2.55	19,520	33	33^{+++}
		(0.10;0.25)***	(0.08;0.32)***			
Weibull-Pareto	3	1.95	2.55	19,520	32	32^{+++}
		(0.09;0.25)***	(0.08;0.33)***			
Inv. Pareto-Gamma-Pareto	4	2.06	2.75	19,441	35	35^{+++}
		(0.10;0.25)***	(0.09;0.32)***			
Gamma-Pareto	3	2.06	2.75	19,441	34	34^{+++}
		(0.09;0.25)***	(0.09;0.34)***			
2-comp. Weibull	5	1.31	2.24	19,404	36	38^{+++}

		(0.09;0.25)***	(0.08;0.32)***			
Inv. Pareto-Exponential-Pareto	3	2.20	3.02	19,394	38	37+++
		(0.09;0.24)***	(0.08;0.31)***			
Exponential-Pareto	2	2.20	3.02	19,394	37	36+++
		(0.09;0.24)***	(0.08;0.30)***			
2-comp. Fréchet	5	1.49	2.25	19,272	39	39+++
		(0.10;0.25)***	(0.08;0.32)***			
2-comp. Gamma	5	2.30	3.50	16,675	40	40+++
		(0.09;0.25)***	(0.08;0.32)***			
2-comp. Exponential	3	3.73	5.90	13,268	41	41+++
		(0.10;0.26)***	(0.08;0.33)***			
Fréchet	2	7.68	12.96	-6,681	42	42+++
		(0.09;0.26)***	(0.08;0.33)***			
Fréchet-Pareto	3	7.68	12.96	-6,681	44	43.5+++
		(0.10;0.25)***	(0.08;0.32)***			
Inv. Pareto-Fréchet	3	7.68	12.96	-6,681	44	43.5+++
		(0.10;0.25)***	(0.09;0.33)***			
Inv. Pareto-Weibull	3	8.40	14.21	-8,737	46	46+++
		(0.10;0.26)***	(0.09;0.33)***			
Weibull	2	8.40	14.21	-8,738	45	45+++
		(0.10;0.24)***	(0.09;0.31)***			
Inv. Pareto-Gamma	3	15.94	24.26	-43,526	47	47+++
		(0.09;0.26)***	(0.08;0.34)***			
Gamma	2	15.94	24.27	-43,533	48	48+++
		(0.09;0.24)***	(0.08;0.30)***			
Exponential	1	32.58	56.49	-139,654	49	49+++
		(0.10;0.25)***	(0.08;0.32)***			
Inv. Pareto-Exponential	2	32.58	56.49	-139,654	50	50+++
		(0.10;0.25)***	(0.09;0.33)***			
Pareto	2	37.07	55.02	-214,535	51	51+++
		(0.09;0.25)***	(0.08;0.32)***			

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.

+++, ++, + indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favour of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.

$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Table 12: Out-of-sample and Cross-validation checks for selected distribution fits to Portuguese domestic sales in 2006.

Distribution	Parms.	Out-of-sample[a]	10-fold CV[b]	MCCV[c]
5-comp. Lognormal	14	16867	1276	6364
4-comp. Lognormal	11	16865	1275	6359
3-comp. Lognormal	8	16808	1264	6313
Double-Pareto Lognormal	4	16542	1243	6204
2-comp. Lognormal	5	16469	1240	6190
Inv. Pareto-Lognormal-Pareto	4	16342	1223	6104
Inv. Pareto-Lognormal	3	13259	919	4575
Lognormal-Pareto	3	12660	872	4350
Left-Pareto Lognormal	3	12089	805	4006
Right-Pareto Lognormal	3	11928	802	4001
Lognormal	2	11305	737	3669
Pareto	2		-43623	-218089

Notes: All distributions fitted using Maximum Likelihood.
[a] The out-of-sample test evaluates the distribution fit to Portuguese domestic sales in 2006 by means of log-liklihood for Portuguese domestic sales in 2007.
[b] The 10-fold CV displayes the average log-likelihood of the parameters obtained from the respective training samples (9 folds) evaluated on the test sample (remaining 1 fold). [b] The Monte Carlo Cross-Validatin displayes the average log-likelihood of the parameters obtained from the training samples (random sample of half of the original sample) evaluated on the test sample (remaining half of the original sample), repeated 20 times.

Table 13: Distribution fits to the U.S. Census 2000 city size distribution.

		Goodness of fit		Information Criteria		
Distribution	Parms.	T_a^0	S_b^0	Loglike	R_{AIC}	R_{BIC}
5-comp. Burr	19	0.22	0.02	-6,004	3	9^{+++}
		(0.33;0.87)	(0.02;0.09)			
3-comp. Burr	11	0.25	0.02	-6,006	1	2^{++}
		(0.33;0.85)	(0.02;0.10)			
4-comp. Burr	15	0.32	0.02	-6,008	2	5^{+++}
		(0.33;0.83)	(0.02;0.09)			
5-comp. Lognormal	14	0.58	0.05	-6,016	6	6^{+++}
		(0.32;0.87)	(0.02;0.10)			
4-comp. Lognormal	11	0.60	0.05	-6,016	5	4^{+++}
		(0.32;0.82)	(0.02;0.09)			
3-comp. Lognormal	8	0.62	0.05	-6,017	4	1
		(0.32;0.86)	(0.02;0.09)			
5-comp. Gamma	14	0.29	0.03	-6,033	7	10^{+++}
		(0.32;0.84)	(0.02;0.09)			
5-comp. Weibull	14	0.38	0.04	-6,037	9	12^{+++}
		(0.33;0.88)	(0.03;0.10)			
2-comp. Lognormal	5	0.71	0.05	-6,044	8	3^{+++}
		(0.33;0.82)	(0.02;0.09)			
2-comp. Burr	7	0.87	0.09	-6,056	10	7^{+++}
		(0.32;0.85)**	(0.02;0.09)*			
Right-Pareto Lognormal	3	1.33	0.17	-6,085	11	8^{+++}
		(0.31;0.84)***	(0.02;0.09)***			
Double-Pareto Lognormal	4	1.39	0.17	-6,085	12	11^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Inv. Pareto-Lognormal-Pareto	4	1.76	0.25	-6,135	14	14^{+++}
		(0.33;0.87)***	(0.02;0.09)***			
Lognormal-Pareto	3	1.75	0.25	-6,135	13	13^{+++}
		(0.33;0.86)***	(0.02;0.10)***			
4-comp. Weibull	11	0.69	0.08	-6,144	16	18^{+++}
		(0.32;0.82)	(0.02;0.09)*			
Inv. Pareto-Lognormal	3	1.90	0.27	-6,152	17	16^{+++}
		(0.33;0.84)***	(0.02;0.09)***			
Lognormal	2	1.89	0.27	-6,152	15	15^{+++}
		(0.32;0.85)***	(0.02;0.09)***			

Left-Pareto Lognormal	3	3.12	0.42	-6,152	18	17^{+++}
		(0.98;1.93)***	(0.14;0.29)***			
4-comp. Gamma	11	0.99	0.11	-6,163	19	19^{+++}
		(0.33;0.84)**	(0.02;0.09)**			
5-comp. Fréchet	14	1.73	0.15	-6,172	21	21^{+++}
		(0.33;0.85)***	(0.02;0.09)***			
4-comp. Fréchet	11	1.57	0.14	-6,174	20	20^{+++}
		(0.33;0.85)***	(0.02;0.09)***			
5-comp. Exponential	9	1.94	0.11	-6,260	22	23^{+++}
		(0.32;0.86)***	(0.02;0.10)**			
3-comp. Fréchet	8	1.61	0.17	-6,261	23	22^{+++}
		(0.32;0.83)***	(0.02;0.09)***			
4-comp. Exponential	7	1.79	0.14	-6,298	24	24^{+++}
		(0.32;0.84)***	(0.02;0.09)***			
Inv. Pareto-Burr	4	2.17	0.30	-6,370	26	26^{+++}
		(0.33;0.85)***	(0.03;0.09)***			
Inv. Pareto-Burr-Pareto	5	2.17	0.30	-6,370	28	28^{+++}
		(0.32;0.85)***	(0.02;0.10)***			
Burr-Pareto	4	2.17	0.30	-6,370	27	27^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Burr	3	2.17	0.30	-6,370	25	25^{+++}
		(0.32;0.84)***	(0.02;0.09)***			
3-comp. Weibull	8	1.71	0.18	-6,393	29	29^{+++}
		(0.32;0.84)***	(0.02;0.10)***			
Inv. Pareto-Fréchet-Pareto	4	3.05	0.40	-6,530	30	30^{+++}
		(0.33;0.83)***	(0.02;0.09)***			
3-comp. Gamma	8	2.37	0.25	-6,532	31	32^{+++}
		(0.33;0.85)***	(0.02;0.09)***			
2-comp. Fréchet	5	2.55	0.32	-6,538	32	31^{+++}
		(0.32;0.84)***	(0.02;0.09)***			
3-comp. Exponential	5	2.76	0.28	-6,633	33	33^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Inv. Pareto-Weibull-Pareto	4	3.60	0.48	-6,829	35	35^{+++}
		(0.32;0.86)***	(0.02;0.09)***			
Weibull-Pareto	3	3.60	0.48	-6,829	34	34^{+++}
		(0.32;0.88)***	(0.02;0.10)***			
Inv. Pareto-Gamma-Pareto	4	3.87	0.52	-6,848	37	39^{+++}
		(0.31;0.87)***	(0.02;0.09)***			

Gamma-Pareto	3	3.87	0.52	-6,848	36	37^{+++}
		(0.32;0.84)***	(0.02;0.10)***			
Inv. Pareto-Exponential-Pareto	3	3.96	0.54	-6,851	39	38^{+++}
		(0.32;0.82)***	(0.02;0.09)***			
Exponential-Pareto	2	3.96	0.54	-6,851	38	36^{+++}
		(0.32;0.85)***	(0.03;0.09)***			
2-comp. Weibull	5	2.96	0.32	-6,920	40	40^{+++}
		(0.32;0.87)***	(0.03;0.09)***			
Fréchet-Pareto	3	4.60	0.64	-7,404	42	42.5^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Inv. Pareto-Fréchet	3	4.60	0.64	-7,404	42	42.5^{+++}
		(0.33;0.85)***	(0.02;0.09)***			
Fréchet	2	4.60	0.64	-7,404	41	41^{+++}
		(0.32;0.84)***	(0.02;0.09)***			
2-comp. Gamma	5	4.23	0.54	-7,694	44	44^{+++}
		(0.32;0.86)***	(0.02;0.10)***			
2-comp. Exponential	3	7.16	0.84	-8,488	45	45^{+++}
		(0.33;0.84)***	(0.03;0.09)***			
Inv. Pareto-Weibull	3	8.31	1.13	-9,030	47	47^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Weibull	2	8.31	1.13	-9,030	46	46^{+++}
		(0.32;0.82)***	(0.02;0.09)***			
Inv. Pareto-Gamma	3	16.40	2.26	-13,169	48	49^{+++}
		(0.33;0.84)***	(0.02;0.09)***			
Gamma	2	16.42	2.26	-13,171	49	48^{+++}
		(0.32;0.87)***	(0.02;0.10)***			
Exponential	1	37.71	5.58	-25,359	50	50^{+++}
		(0.32;0.87)***	(0.02;0.09)***			
Inv. Pareto-Exponential	2	37.71	5.58	-25,359	51	51^{+++}
		(0.32;0.85)***	(0.02;0.09)***			
Pareto	2	41.69	5.06	-31,612	52	52^{+++}
		(0.33;0.83)***	(0.02;0.09)***			

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.

$^{+++}$, $^{++}$, $^{+}$ indicates the difference between this distribution's BIC and the first-ranked distribution in terms of BIC (ΔBIC) providing strong evidence in favour of the first-ranked distribution ($\Delta BIC > 10$), moderate evidence ($6 < \Delta BIC \leq 10$) and weak evidence ($2 < \Delta BIC \leq 6$) respectively.

$_a$ Values multiplied by 100 for expositional purpose, $_b$ Values divided by 1,000 for expositional purpose.

Table 14: Decomposition of procentual welfare gains from a reduction in variable trade costs $\tau^{ij} = 3 \rightarrow (\tau^{ij})' = 1$.

Distribution	Parms.	$ln\frac{U_i'}{U_i}$	$ln\frac{\tau_{ij}'}{\tau_{ij}}$	$ln\frac{M_i'}{M_i}$	$ln\frac{1-G(\omega_{ij}^*)'}{1-G(\omega_{ij}^*)}$	$ln\frac{\tilde{\omega}(\omega_{ij}^*)'}{\tilde{\omega}(\omega_{ij}^*)}$	$ln\frac{\lambda_{ij}'}{\lambda_{ij}}$
Pareto	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)***
Weibull	2	0.15	1.10	-0.16	0.12	1.35	-2.26
		(0.15;0.15)***	(1.10;1.10)	(-0.16;-0.16)***	(0.12;0.13)***	(1.30;1.41)***	(-2.32;-2.21)***
Inv. Pareto-Weibull	3	0.15	1.10	-0.16	0.12	1.35	-2.26
		(0.15;0.15)***	(1.10;1.10)	(-0.16;-0.16)***	(0.12;0.13)***	(1.30;1.41)***	(-2.32;-2.21)***
Left-Pareto Lognormal	3	0.16	1.10	-0.17	0.15	0.60	-1.51
		(0.16;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.58;0.62)***	(-1.53;-1.49)***
Inv. Pareto-Lognormal	3	0.17	1.10	-0.17	0.15	0.58	-1.49
		(0.16;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.56;0.60)***	(-1.51;-1.47)***
Lognormal	2	0.17	1.10	-0.17	0.15	0.53	-1.44
		(0.17;0.17)***	(1.10;1.10)	(-0.17;-0.17)***	(0.15;0.15)***	(0.51;0.55)***	(-1.46;-1.42)***
Right-Pareto Lognormal	3	0.18	1.10	-0.18	0.17	0.28	-1.19
		(0.18;0.19)**	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.23;0.33)**	(-1.24;-1.13)**
2-comp. Weibull	5	0.18	1.10	-0.13	0.11	0.57	-1.46
		(0.18;0.18)***	(1.10;1.10)	(-0.14;-0.13)***	(0.11;0.11)***	(0.56;0.59)***	(-1.48;-1.45)***
3-comp. Weibull	8	0.19	1.10	-0.19	0.18	0.25	-1.16
		(0.18;0.19)***	(1.10;1.10)	(-0.19;-0.19)***	(0.18;0.19)***	(0.25;0.26)***	(-1.17;-1.15)***
4-comp. Weibull	11	0.19	1.10	-0.18	0.17	0.22	-1.12
		(0.19;0.19)***	(1.10;1.10)	(-0.18;-0.17)***	(0.16;0.17)***	(0.21;0.23)***	(-1.13;-1.11)***
5-comp. Weibull	14	0.19	1.10	-0.18	0.18	0.22	-1.12
		(0.19;0.19)**	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.21;0.23)***	(-1.14;-1.11)***
Empirical	0	0.19	1.10	-0.18	0.18	0.20	-1.10
4-comp. Lognormal	11	0.19	1.10	-0.18	0.18	0.20	-1.10

		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.18;0.22)	(-1.13;-1.08)
5-comp. Lognormal	14	0.19	1.10	-0.19	0.18	0.20	-1.10
		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.19)	(0.17;0.22)	(-1.12;-1.07)
2-comp. Lognormal	5	0.19	1.10	-0.17	0.17	0.23	-1.13
		(0.19;0.19)	(1.10;1.10)	(-0.18;-0.17)***	(0.16;0.17)***	(0.22;0.25)***	(-1.15;-1.12)***
3-comp. Lognormal	8	0.19	1.10	-0.18	0.18	0.19	-1.09
		(0.19;0.19)	(1.10;1.10)	(-0.19;-0.18)	(0.17;0.18)	(0.16;0.22)	(-1.12;-1.06)
Lognormal-Pareto	3	0.22	1.10	-0.22	0.22	0.02	-0.90
		(0.20;0.21)***	(1.10;1.10)	(-0.22;-0.20)***	(0.20;0.22)***	(0.04;0.14)***	(-1.04;-0.93)***
Burr	3	-	1.10	-	-	-	-
		(0.20;0.21)***	(1.10;1.10)	(-0.21;-0.19)***	(0.19;0.21)***	(0.03;0.12)***	(-1.02;-0.92)***
2-comp. Burr	7	-	1.10	-	-	-	-
		(0.19;0.20)	(1.10;1.10)	(-0.20;-0.18)	(0.17;0.20)	(0.10;0.22)	(-1.12;-1.00)
3-comp. Burr	11	-	1.10	-	-	-	-
		(0.19;0.20)**	(1.10;1.10)	(-0.21;-0.18)	(0.18;0.21)	(0.08;0.20)	(-1.11;-0.97)
4-comp. Burr	15	-	1.10	-	-	-	-
		(0.19;0.21)**	(1.10;1.10)	(-0.21;-0.18)**	(0.18;0.21)**	(0.07;0.20)**	(-1.10;-0.96)**
5-comp. Burr	19	-	1.10	-	-	-	-
		(0.19;0.21)***	(1.10;1.10)	(-0.22;-0.19)***	(0.18;0.22)***	(0.05;0.19)***	(-1.09;-0.94)***
Burr-Pareto	4	-	1.10	-	-	-	-
		(0.20;0.21)***	(1.10;1.10)	(-0.21;-0.19)***	(0.19;0.21)***	(0.02;0.12)***	(-1.02;-0.91)***
Double-Pareto Lognormal	4	-	1.10	-	-	-	-
		(0.20;0.22)***	(1.10;1.10)	(-0.20;-0.19)***	(0.19;0.20)***	(0.02;0.09)***	(-0.98;-0.90)***
Fréchet	2	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.14;-0.08)***	(0.08;0.14)***	(0.00;0.01)***	(-0.89;-0.88)***
2-comp. Fréchet	5	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.15;-0.10)***	(0.10;0.15)***	(0.00;0.01)***	(-0.89;-0.88)***
3-comp. Fréchet	8	-	1.10	-	-	-	-

		(0.22;0.22)***	(1.10;1.10)	(-0.15;-0.11)**	(0.11;0.15)**	(0.00;0.01)***	(-0.89;-0.88)***
4-comp. Fréchet	11	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.15;-0.11)**	(0.11;0.15)**	(0.00;0.01)***	(-0.89;-0.88)***
5-comp. Fréchet	14	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.15;-0.10)***	(0.10;0.15)**	(0.00;0.01)***	(-0.89;-0.88)***
Fréchet-Pareto	3	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.14;-0.08)***	(0.08;0.14)***	(0.00;0.01)***	(-0.89;-0.88)***
Inv. Pareto-Burr	4	-	1.10	-	-	-	-
		(0.20;0.22)***	(1.10;1.10)	(-0.21;-0.19)***	(0.19;0.21)***	(0.02;0.11)***	(-1.00;-0.90)***
Inv. Pareto-Burr-Pareto	5	-	1.10	-	-	-	-
		(0.20;0.22)***	(1.10;1.10)	(-0.21;-0.19)***	(0.19;0.20)***	(0.02;0.11)***	(-1.00;-0.90)***
Inv. Pareto-Fréchet	3	-	1.10	-	-	-	-
		(0.22;0.22)***	(1.10;1.10)	(-0.14;-0.08)***	(0.08;0.14)**	(0.00;0.01)***	(-0.89;-0.88)***
Inv. Pareto-Fréchet-Pareto	4	-	1.10	-	-	-	-
		(0.21;0.22)***	(1.10;1.10)	(-0.19;-0.18)	(0.18;0.19)**	(0.01;0.07)***	(-0.96;-0.89)***
Inv. Pareto-Lognormal-Pareto	4	-	1.10	-	-	-	-
		(0.21;0.22)***	(1.10;1.10)	(-0.20;-0.18)	(0.18;0.20)***	(0.01;0.08)***	(-0.97;-0.89)***
Inv. Pareto-Weibull-Pareto	4	-	1.10	-	-	-	-
		(0.21;0.22)***	(1.10;1.10)	(-0.18;-0.16)**	(0.16;0.18)	(0.00;0.05)***	(-0.93;-0.88)***
Weibull-Pareto	3	-	1.10	-	-	-	-
		(0.21;0.22)***	(1.10;1.10)	(-0.18;-0.16)**	(0.16;0.18)	(0.00;0.05)***	(-0.93;-0.88)***

Notes: $ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i}$ indicates the log changes in real per-capita income due to an exogenous increase in variable trade costs τ_{ij} to τ'_{ij}. This is 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 ($m^{\sigma-1}_{\omega^{ij*}}$) and the bilateral trade share (λ^{ij}).

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.

Appendix B Fitting truncated data

This section extends the methodology of the main paper to allow fitting the distributions to truncated data. This allows us to single out and focus on tail performance while generalizing the proposed distributional fits to unrepresentative and/or truncated data. It also allows us to evaluates the ability of FMMs to accurately capture the tail of the empirical distribution.

B.1 (Inverse) Pareto

The (Inverse) Pareto distribution is a special distribution, being truncated from (above) below by definition.[1] This means that the (upper) lower truncation point lies within the parameter space of the distribution, and distribution fits can be optimized accordingly. The ML estimator as specified in equation 4 merely assumes the exogenously applied truncation points as the scale parameter.

Obtaining an accurate estimate for the (upper) lower bound is, however, vital to the accuracy of the estimated shape parameter $\hat{k}$. Choosing a (maximum) minimum too (high) low results in a biased shape parameter, as one will be fitting a power-law to non-power-law data. Choosing a value too (low) high, on the other hand, increases the statistical error and bias from finite-size effects on the shape parameter, as one discards legitimate data points. Moreover, it is widely documented that the minimum and shape parameters of the Pareto distribution exhibit a positive correlation (Eeckhout, 2004; di Giovanni and Levchenko, 2013; Head et al., 2014; Freund and Pierola, 2015; Bee and Schiavo, 2018).

In order to obtain an accurate estimate for the lower (upper) bound, therefore, we rely on a formal decision rule developed by Clauset et al. (2009).[2] For the ordered productivity set $\{x_b; b = 1, \ldots, B\}$, we evaluate every x_b as a potential (x_{max}) x_{min}, estimating the ML estimate of the power-law exponent k. We then use the Kolmogorov-Smirnov statistic to select the

[1] Fully truncated (both from below and above) Pareto distributions can be deduced from a truncated probability density function (see Eq. 2) and have been used in the economic literature (Helpman et al., 2008; Melitz and Redding, 2014; Feenstra, 2018).

[2] Alternative methodologies to determine the lower Pareto bound consist of (i) relying on a visual examination, looking for a 'kink' in the distribution above which the relationship between log-rank and log-size is approximately linear (di Giovanni and Levchenko, 2013; Bas et al., 2017), (ii) relying on export sales and assuming a truncation parameter equal to the minimum of sales (see, for instance, Freund and Pierola (2015)), (iii) determining the lower bound to ensure a Pareto parameter large enough to deliver finite moments when calibrating their theoretical models (Head et al., 2014; Bee and Schiavo, 2018), and (iv) estimating the lower bound assuming a mixed Lognormal-Pareto distribution (Malevergne et al., 2011; Bakar and Nadarajah, 2013; Nigai, 2017). Such methods are either subject to large measurement errors and inconsistencies or restrictive in their need to assume a distributional relation between the bulk and the tail of the distribution.

optimum (x_{max}) x_{min}. It is defined as the cutoff which minimizes the maximum absolute deviation of the empirical from the theoretical CDF:

$$T_{KS,\hat{x}_{max}} = \sup_{x \leq \hat{x}_{max}} \left| \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(x_b \leq \hat{x}_{max}) - G_{IP}(x; \hat{k}, \hat{x}_{max}) \right|$$
$$T_{KS,\hat{x}_{min}} = \sup_{x \geq \hat{x}_{min}} \left| \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(x_b \geq \hat{x}_{min}) - G_{P}(x; \hat{k}, \hat{x}_{min}) \right|, \tag{1}$$

where $\mathbb{I}_A$ is the indicator of event A.

B.2 Hump-shaped, piecewise composite and product distributions

Consisting of individual truncated densities, the estimation of piecewise composite distributions on truncated data is by its definition straightforward. Maximum likelihood methods for the remaining hump-shaped and product distributions can easily be adapted by truncating the distribution to be restricted within the data domain. The resulting truncated probability density function ($g^*(x)$) is then specified within the (exogenously determined) boundaries $x \in \left[c^l, c^u\right]$:

$$g^*(x) = \frac{g(x)}{G(c^u) - G(c^l)}. \tag{2}$$

B.3 FMM

The EM-algorithm can be adapted to fitting data only to truncated data within the (exogenously determined) boundaries $x \in \left[c^l, c^u\right]$. We specify the conditional densities

$$\begin{aligned} g(x|\mathbf{\Psi}, c^l \leq x \leq c^u) &= \frac{\sum_{i=1}^{I} \pi_i m_i(x|\boldsymbol{\theta_i})}{G(c^u|\mathbf{\Psi}) - G(c^l|\mathbf{\Psi})} \\ &= \sum_{i=1}^{I} \pi_i \frac{M_i(c^u|\boldsymbol{\theta_i}) - M_i(c^l|\boldsymbol{\theta_i})}{G(c^u|\mathbf{\Psi}) - G(c^l|\mathbf{\Psi})} \frac{m_i(x|\boldsymbol{\theta_i})}{M_i(c^u|\boldsymbol{\theta_i}) - M_i(c^l|\boldsymbol{\theta_i})} \\ &= \sum_{i=1}^{I} \eta_i m_i(x|\boldsymbol{\theta_i}, c^l \leq x \leq c^u), \end{aligned} \tag{3}$$

with $\eta_i > 0$, $\sum_{i=1}^{I} \eta_i = 1$ and M_i the component-specific Cumulative Distribution Function.

The Q-function becomes

$$\begin{aligned} Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s-1)}) &= E\left[logL(x|\boldsymbol{\Psi})|x, \boldsymbol{\Psi}^{(s-1)}\right] \\ &= \sum_{b=1}^{B}\sum_{i=1}^{I} \pi_{bi}^{(s)}\left[log(\eta_i) + log(m_i(x_b|\boldsymbol{\theta}_i, c^l \leq x_b \leq c^u))\right], \end{aligned} \tag{4}$$

where the posterior probability that x_b comes from the ith mixture is not affected by the truncation:

$$\pi_{bi}^{(s)} = \frac{\eta_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)}, c^l \leq x_b \leq c^u))}{\sum_{i=1}^{I}\eta_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)}), c^l \leq x_b \leq c^u)} = \frac{\pi_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)})}{\sum_{i=1}^{I}\pi_i^{(s-1)} m_i(x_b|\boldsymbol{\theta}_i^{(s-1)})}. \tag{5}$$

The M-step then again consists of maximizing the Q-function over the parameters $\boldsymbol{\Psi}$. Iterating over the E- and M-step until the algorithm converges provides us with distributions fitted to the truncated data.

Appendix C Robustness and Extensions

C.1 Extension to other distributions

The superior performance of FMMs is not limited to the Lognormal distribution. Appendix Table 7 displays the results of fits to the complete data expanding to FMMs of distributions often used in the economic literature such as the Exponential, Gamma, Weibull, Burr, and Fréchet distribution. Most of these mixtures are not able to match the performance of the Lognormal. Only the Burr distribution provides an equivalent fit to the PDF and CDF.[3]

Compared to Pareto-tailed combinations of distributions, we find that also mixtures of Weibull and Gamma can provide an improved distribution fit. Overall, the currently favored Double-Pareto Lognormal (Sager and Timoshenko, 2019) and Lognormal-Pareto (Nigai, 2017) distribution are ranked sixteenth and thirty-first, respectively, according to BIC, out of 52 considered distributions.

The consistent excellent performance of the Lognormal distribution can be motivated from two perspectives. From the perspective of overall fit, a mixture of (log-) normal distributions with sufficient components is assumed to be able to approach all distributions (McLachlan and Peel, 2000). From a generative perspective for individual components, the Lognormal distribution is the realization of applying the Central Limit Theorem (CLT) in the log domain: firm heterogeneity will approximately be Lognormal if it is the multiplicative product of many independent random variables. This corresponds with extensions of heterogeneous firms models à la Melitz (2003) that consider multi-dimensional firm heterogeneity, taking into consideration the product dimension (Bernard et al., 2009) or uncertainty in demand and/or supply (see for instance De Loecker (2011); Bas et al. (2017); Sager and Timoshenko (2019); Gandhi et al. (2020)).

C.2 Robustness

We scrutinize the robustness of our results with several additional analyses. First, we examine whether our results are not caused by sample selection. To this end, we restrict our dataset to the manufacturing sector only (see Appendix Table 10) and find the performance of FMMs to improve relative to Pareto-tailed distributions. Second, we inspect whether our results are not due to outliers in the tails of the distribution by discarding the 1,000 smallest and largest observations

[3]The Burr distribution fails to match higher moments of the data, however. See also section 6.

from our dataset. Results in Appendix Table 11 again confirm the superiority of FMMs.

The AIC reported in Table 1 is asymptotically equivalent to leave-one-out cross-validation (Stone, 1977). We perform a robustness check on the out-of-sample predictive accuracy of our results using (i) a Monte Carlo Cross-Validation (MCCV), (ii) $k-$fold cross-validation, and (iii) an out-of-sample test for model selection. The MCCV consists of partitioning the data $B = 20$ times into disjoint training and test subsets where the test subset is a fraction $\beta = 0.5$ of the overall data (Smyth, 1996). The k-fold cross-validation consists of partitioning the data into $k = 10$ disjoint subsets of the data (Grimm et al., 2017). The model is estimated based on $k-1$ partitions (training set). Then the model with unknown parameters fixed at their previously estimated values is applied to the remaining partition (the kth partition that was not part of the training sample, test set). This is repeated k times with each of the k potential configurations of the empirical data (Grimm et al., 2017). For both cross-validation procedures, we retain the log-likelihood for each iteration and show the resulting average log-likelihood. The out-of-sample test evaluates the distribution fit by the log-likelihood for Portuguese domestic sales in 2007, relying on the coefficients estimated on Portuguese domestic sales in 2006. The results of this exercise (see Online Appendix Table 12) confirm the main results and demonstrate that a mixture of Lognormals improve the model fit without over-fitting the data.

Finally, we also provide external validation, in line with Nigai (2017), by fitting the considered distributions to the U.S. Census 2000 city size distribution data. This dataset has been subject to an extensive debate in the city size literature, including the discussion between Eeckhout (2004, 2009) and Levy (2009).[4] Appendix Table 13 provides the test results, demonstrating that the city size distribution is neither Lognormal, Pareto, nor Pareto-tailed Lognormal. It is best approximated by a 2-component Lognormal distribution (according to the BIC). These results provide an overview of the city size literature until now and are in line with the findings of Kwong and Nadarajah (2019).

[4] The dataset is available at https://www.aeaweb.org/aer/data/sept09/20071478_data.zip.

Appendix D Motivation and identification of generative processes for mixture models

FMMs can be utilized in two ways. First, they can be used as a semi-parametric, flexible approximation of the overall distribution, which is the case in this paper. Second, they are model-based clustering methods when a certain distribution is imposed (Fop et al., 2018; Grün, 2018). While both applications rely on the idea that discrete subpopulations define the overall distribution, the semi-parametric approximation does not claim to identify these subpopulations. This appendix conceptualizes possible Data Generative Processes (DGPs) for FMMs based on theoretical and empirical work in the economics literature. We then elaborate on the identification difficulties/opportunities of the underlying mixture components in the context of productivity distributions.

D.1 Generative processes

Many economic models rely on the assumption that the firm size distribution originates from firm dynamics in productivity (see for instance Hopenhayn (1992); Luttmer (2007); Rossi-Hansberg and Wright (2007); Costantini and Melitz (2008); Arkolakis (2016)). In this section, we will use a simplified version of such productivity dynamics for explanatory purposes. Consider productivity dynamics specified as a first-order autoregressive process:

$$ln\omega_{bt} = c + \rho ln\omega_{bt-1} + \eta_{bt}, \tag{6}$$

where η_{bt} is a white noise process with zero mean and constant variance σ^2.

Some empirical evidence suggests that productivity dynamics, and therefore the resulting productivity distributions, are endogenous to exporting (De Loecker, 2013), importing (Kasahara and Rodrigue, 2008), innovation (Aw et al., 2011), management practices (Bloom and Reenen, 2011; Caliendo et al., 2020), ... Overall, there are "many sources of heterogeneity that support the idea of discrete subpopulations likely to differ in important characteristics ..." (Perline (2005),p.80). In the case of exporting, the endogenous evolution of productivity results in an exporting productivity premium. This can empirically be observed from the standard textbook comparison of cross-sectional productivity densities between exporting and non-exporting firms (see Figure 8).

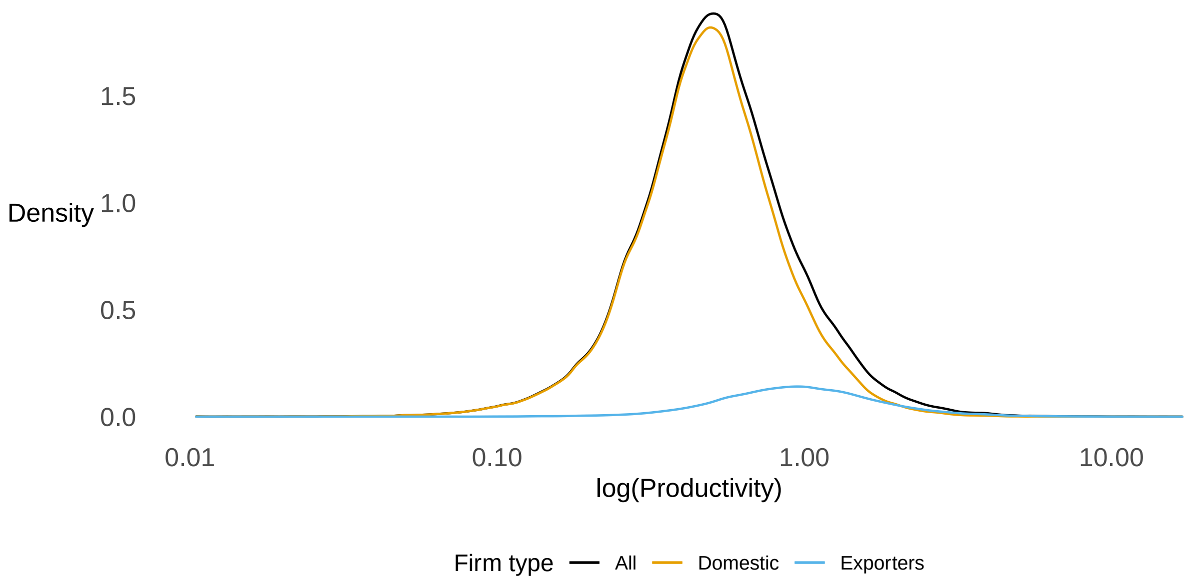

Figure 8: Productivity density of Portuguese firm productivity in 2006 for all, exporting- and non-exporting firms.
Notes: Productivity is measured as domestic sales (relative to the mean) to the power of $1/(\sigma - 1)$ with σ, the elasticity of substitution between varieties, set to four.

Building on equation 6, a simplified version of the empirical specification to identify such exporting productivity premium, and replicate Figure 8, is essentially a specifically parametrized FMM:

$$\begin{aligned} ln\omega_{bt} &= \alpha_0 + \beta_0 EXP_b + \alpha_1 ln\omega_{bt-1} + \beta_1 EXP_b \times ln\omega_{bt-1} + \eta_{bt} \\ &= EXP_b \left[\beta_0 + \beta_1 ln\omega_{bt-1}\right] + (1 - EXP_b)\left[\alpha_0 + \alpha_1 ln\omega_{bt-1}\right] + \eta_{bt}, \end{aligned} \tag{7}$$

with EXP_b a dummy variable that takes the value 1 when the firm b is an exporter and 0 otherwise.

Whereas the components are identified using an exporter dummy variable in this example, FMMs are a semi-parametric specification that remain agnostic about the (possibly multiple) determinants of the unobserved components and allow the data to determine these components:[5]

[5]Note that, for simplicity, we specify the variance to be constant between components. FMMs in the main analysis allow for the variance to differ between components.

$$ln\omega_{bt} = \sum_{i=1}^{I} \mathbb{I}_b^i \left[\beta_0^i + \beta_1^i ln\omega_{bt-1}\right] + \eta_{bt}. \tag{8}$$

D.2 Identification

As stated before, FMM's can focus on the semi-parametric, flexible approximation of the overall distribution or on model-based clustering. This paper purely focuses on the semi-parametric approximation. First, we take no à-priori stance on distributional specification.[6] Second, even if one is willing to assume distributional specification such as the Lognormal, the underlying components remain unidentifiable in the current setting. As the overall distribution is unimodal (see Figure 8), there is a large overlap between the underlying individual densities. These individual densities will therefore be poorly identified. Indeed, Figure 9 displays the posterior probability distribution for each component of the fitted 4-component Lognormal mixture from the main text. Whereas well-identified components have a large weight near zero and 1, average probabilities lie close to 0.25 in this case and are therefore not well identified. While the overall distribution can be closely approximated, the large overlap of individual densities results in a large uncertainty on which observation can be assigned to which density. Neither the parameter estimates used to characterize the clusters nor the partitions derived can therefore be uniquely determined, rendering the interpretation of results in terms of clustering futile (Follmann and Lambert, 1991; Hennig, 2000; Grün, 2018; Grün and Leisch, 2008).

Future research might resolve the identifiability problem relying on panel rather than cross-sectional data. The problem as specified now is a problem in levels (the cross-section), where it appears there is insufficient distance between different components for them to be identified. From empirical evidence, however, it can be deduced that the different components likely originate from differences in growth rates (Kasahara and Rodrigue, 2008; Aw et al., 2011; De Loecker, 2013; Caliendo et al., 2020). Tracking the growth rates of individual firms over time might allow for the

[6] The empirical evidence in this paper seems to favor a Lognormal specification. This can be motivated from two perspectives. From the perspective of overall fit, a mixture of (log-) normal distributions with sufficient components is assumed to be able to approach all distributions (McLachlan and Peel, 2000). From a generative perspective for individual components, the Lognormal distribution is the realization of applying the Central Limit Theorem (CLT) in the log domain: firm heterogeneity will approximately be Lognormal if it is the multiplicative product of many independent random variables. Whereas firm heterogeneity reduces to firm-level productivity in the Melitz (2003)-model, it has been argued to be multi-dimensional when taking into consideration, for instance, the product dimension (Bernard et al., 2009) or uncertainty in demand and/or supply (see (De Loecker, 2011; Bas et al., 2017; Sager and Timoshenko, 2019; Gandhi et al., 2020) . . .)

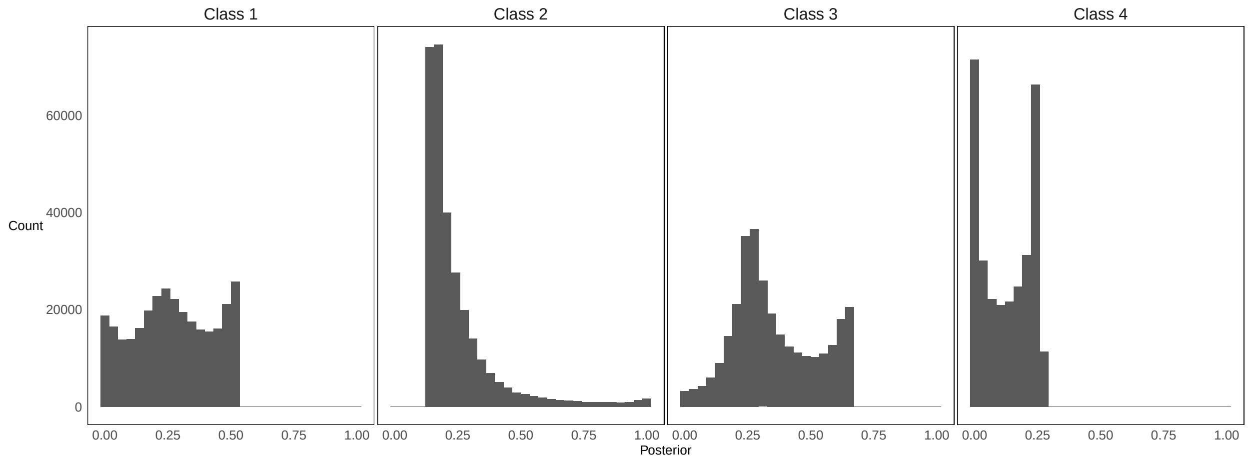

Figure 9: Posterior probability distribution for each component of the 4-components Lognormal mixture.

variation needed to identify the components of the overall distribution.

This observation can be easily illustrated using simulated data. Building on the example of the previous paragraph, imagine $ln\omega_{bt}$ follows an AR(1)-process with an exporting productivity premium of 20%:

$$ln\omega_{bt} = 1 + 1.2 \times EXP_b + 0.7 \times ln\omega_{bt-1} + \eta_{bt},$$

with $\eta_{bt} \sim \mathcal{N}(0, 0.3)$. We simulate this evolution for 200 exporters ($EXP_b = 1$) and 800 purely domestic businesses over 10 years.[7] The firm densities of the simulated data will look similar to Figure 8, with two densities largely overlapping but the exporter productivity density located on the right of domestic firms density.

If we fit, as in our main analysis, a FMM on the cross-sectional data of a selected (the first) year, we obtain a familiar posterior probability distribution (see Figure 10). Individual clusters are not well-identified. Exploiting the panel dimension of the data,[8] however, results in well-identified

[7]When simulating, we allow for a run-in period of 90 years.

[8]Specifically, the EM estimation procedure is adapted to take into account panel data. The component probabilities in our main analysis are specified over the complete data (eq. 7):

$$\pi_{bi}^{(s)} = E\left[z_{bi}|\omega_b, \mathbf{\Psi}^{(s-1)}\right] = \frac{\pi_i^{(s-1)} m_i(\omega_b|\boldsymbol{\theta}_i^{(s-1)})}{\sum_{i=1}^{I} \pi_i^{(s-1)} m_i(\omega_b|\boldsymbol{\theta}_i^{(s-1)})}.$$

When working with panel data, we adapt this specification to take into account the time dimension:

$$\pi_{bi}^{(s)} = E\left[z_{bi}|\omega_{bt}, \mathbf{\Psi}^{(s-1)}\right] = \frac{\pi_i^{(s-1)} \prod_{t=1}^{T} m_{it}(\omega_{bt}|\boldsymbol{\theta}_i^{(s-1)})}{\sum_{i=1}^{I} \pi_i^{(s-1)} \prod_{t=1}^{T} m_{it}(\omega_{bt}|\boldsymbol{\theta}_i^{(s-1)})}.$$

components. As can be observed in Figure 11, the posterior probabilities predominantly take the values zero or one. Once components are well-identified, one can try to determine which mechanisms motivate the existence of FMMs from a generative perspective.

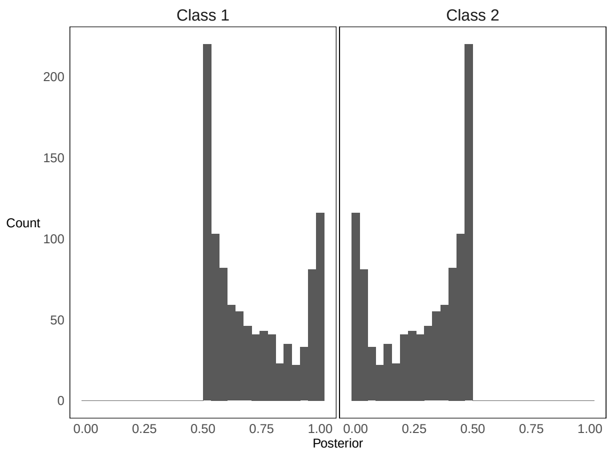

Figure 10: *Cross-sectional* posterior probability distribution for each component of the simulated 2-components normal mixture.

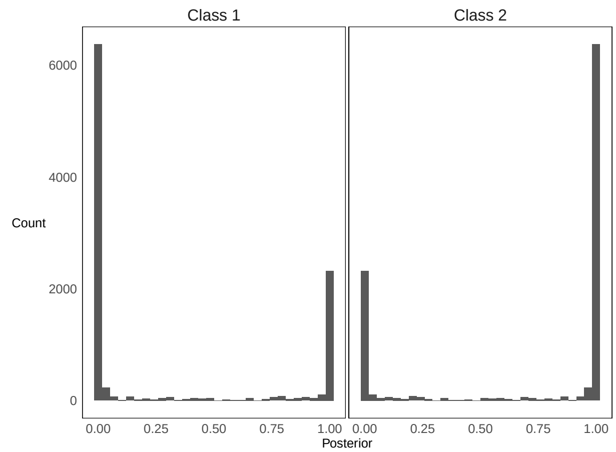

Figure 11: *Panel* posterior probability distribution for each component of the simulated 2-components normal mixture.

Note that the probabilities are specified to be constant over time, meaning that we do not allow for regime-switching in this exercise.

Appendix E Heterogeneous firms model

This appendix provides a detailed description of the heterogeneous firms models relied upon in the paper. We follow Dewitte (2020) in presenting a firm heterogeneous open economy model of Melitz (2003) with a finite number of firms. 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 equivalent to the Melitz (2003)- model.

E.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

$$U^j = \left(\sum_{i=1}^{I} \sum_{\varpi \in \Omega^i} q^{ij} \left(\varpi \right)^{\frac{\sigma - 1}{\sigma}} d\varpi \right)^{\frac{\sigma}{\sigma - 1}}, \tag{9}$$

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

$$\frac{q^{ij}(\varpi)}{Q^j} = \left[\frac{p^{ij}(\varpi)}{P^j} \right]^{-\sigma}, \tag{10}$$

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 ($b \in B$) that 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.[9] There is zero probability of firm death.[10] 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:

$$\begin{aligned}
\max_{q^{ij}} \pi^{ij} &= \max_{q^{ij}} \left[p^{ij} q^{ij} - \left(f^{ij} - \frac{\tau^{ij} q^{ij}}{\omega} \right) W^i \right] \\
&= \max_{q^{ij}} \left[\left(q^{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(Q^j\right)^{\frac{1}{\sigma}} P^j - \left(f^{ij} - \frac{\tau^{ij} q^{ij}}{\omega} \right) W^i \right],
\end{aligned} \tag{11}$$

results in an optimal quantity produced:

$$\begin{gathered}
\frac{\partial \pi^{ij}}{\partial q^{ij}} = 0 \\
\Leftrightarrow \\
\frac{\sigma-1}{\sigma} \left(q^{ij}\right)^{-\frac{1}{\sigma}} \left(Q^j\right)^{\frac{1}{\sigma}} P^j = \frac{\tau^{ij} W^i}{\omega} \\
\Leftrightarrow \\
q^{ij} = \left(\frac{\sigma}{\sigma-1} \frac{\tau^{ij} W^i}{\omega} \right)^{-\sigma} Q^j \left(P^j\right)^{\sigma}.
\end{gathered} \tag{12}$$

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

$$\begin{gathered}
\left(\frac{q^{ij}}{(Q^j)} \right)^{\frac{-1}{\sigma}} P^j = p^{ij} \\
p^{ij} = \frac{\sigma}{\sigma-1} \frac{\tau^{ij} W^i}{\omega}.
\end{gathered} \tag{13}$$

[9] As ω_b is the sole heterogeneity component identifying individual firms, we drop the subscript b in further derivations.

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

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

$$
\begin{aligned}
x^{ij} = p^{ij}q^{ij} &= \left(q^{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(Q^{j}\right)^{\frac{1}{\sigma}} P^{j} \\
&= \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma}
\end{aligned}
\tag{14}
$$

E.2 Operating decisions

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

$$
\begin{aligned}
\pi^{ij} = 0 &= p^{ij}q^{ij} - \left(f^{ij} - \frac{\tau^{ij}q^{ij}}{\omega^{ij*}}\right)W^{i}, \\
&= \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma} - f^{ij}W^{i} - \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{-\sigma} Q^{j}\left(P^{j}\right)^{\sigma}\frac{\tau^{ij}}{\omega^{ij*}}W^{i}, \\
&= \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma} - f^{ij}W^{i} - \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma}\left(\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma}, \\
&= \left(1-\frac{\sigma-1}{\sigma}\right)\left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma} - f^{ij}W^{i}, \\
&\Leftrightarrow \\
\sigma f^{ij}W^{i} &= \left(\frac{\sigma}{\sigma-1}\frac{\tau^{ij}W^{i}}{\omega^{ij*}}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma}.
\end{aligned}
\tag{15}
$$

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:

$$
\omega^{ij*} = \left(\frac{W^{i}}{W^{j}}\right)^{\frac{\sigma}{\sigma-1}} \left(\frac{f^{ij}}{f^{jj}}\right)^{\frac{1}{\sigma-1}} \left(\frac{\tau^{ij}}{\tau^{jj}}\right)\omega^{jj*}. \tag{16}
$$

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

$$\omega^{ij*} = \frac{\tau^{ij}}{\tau^{ii}} \left(\frac{P^j}{P^i}\right)^{\frac{\sigma}{1-\sigma}} \left(\frac{Q^i}{Q^j}\frac{f^{ij}}{f^{ii}}\right)^{\frac{1}{\sigma-1}} \omega^{ii*}. \tag{17}$$

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

- A high 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 high productivity firms 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:

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

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

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

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

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

$$\sum_{j=1}^{J} f^{ij} \left[\left(\omega^{ij*}\right)^{1-\sigma} m^{\sigma-1}_{\omega^{ij*}} - m^{0}_{\omega^{ij*}}\right] = f^{ie}, \tag{18}$$

where we denote by m^r_y the y-bounded, r-th sample moment of the productivity distribution. For the number of firms going to infinity, the law of large numbers kicks in such that we replace these sample moments with their continuous equivalent $\left(\mu^r(y) = \int_y^{\infty} \omega^r g(\omega) d\omega\right)$, providing us with the well-known continuous free-entry equation as specified by (Melitz, 2003).

Using the relation between productivity cutoffs (eq. 16), the free entry condition (eq. 18) determines a unique equilibrium values of these cutoffs.[11] 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; Dewitte, 2020).

E.3 Aggregation

Summing equation 14 across all active firms, we obtain an expression for aggregate trade between country i and j:

[11] Sufficient conditions for this equilibrium to exist are that the term in brackets of equation (18) 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 $(0, \infty)$.

$$X^{ij} = \left(\frac{\sigma}{\sigma-1}\tau^{ij}W^{i}\right)^{1-\sigma} Q^{j}\left(P^{j}\right)^{\sigma} M^{ie} m_{\omega^{ij*}}^{\sigma-1} \tag{19}$$

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

$$M^{i} = \left[1-G(\omega^{ii*})\right]M^{ie} = \frac{X^{i}}{\mathbb{E}\left[x^{i}\right]}. \tag{20}$$

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

$$\begin{aligned} M^{i} &= \frac{W^{i}L^{i}}{\sigma\left(\frac{f^{ie}}{1-G(\omega^{ii*})} + \sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})}f^{ij}\right)W^{i}} \\ &= \frac{L^{i}}{\sigma\left(\frac{f^{ie}}{1-G(\omega^{ii*})} + \sum_{j=1}^{J}\frac{1-G(\omega^{ij*})}{1-G(\omega^{ii*})}f^{ij}\right)}. \end{aligned} \tag{21}$$

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

$$\mathbb{W}^{i} = (P^{i})^{-1}. \tag{22}$$

The price index can be deduced from equation 19:

$$P^{j} = \left[\left(\frac{\sigma}{\sigma-1}\tau^{ij}W^{i}\right)^{1-\sigma}\frac{1}{\lambda_{ij}}\frac{M^{i}}{1-G(\omega^{ii*})}m_{\omega^{ij*}}^{\sigma-1}\right]^{\frac{1}{1-\sigma}}, \tag{23}$$

where we denote the share of expenditure by j on goods from i, the bilateral trade share, 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:

$$
\begin{aligned}
100 \times ln\frac{(\mathbb{W}^i)'}{\mathbb{W}^i} &= 100 \times -ln\frac{(P^i)'}{P^i} \qquad (24) \\
&= 100 \times -ln\frac{(P^j)'}{P^j} \\
&= 100 \times -\left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma-1}\left(ln\frac{(M^i)'}{M^i} - ln\frac{1-G(\omega^{ii*})'}{1-G(\omega^{ii*})} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}}\right)\right] \\
&= 100 \times -\left[ln\frac{(\tau^{ij})'}{(\tau^{ij})} - \frac{1}{\sigma-1}\left(ln\frac{(M^i)'}{M^i} - ln\frac{(m^{0}_{\omega^{ij*}})'}{m^{0}_{\omega^{ij*}}} + ln\frac{(m^{\sigma-1}_{\omega^{ij*}})'}{m^{\sigma-1}_{\omega^{ij*}}} - ln\frac{(\lambda^{ij})'}{\lambda^{ij}}\right)\right].
\end{aligned}
$$

E.4 Parametrization

To parametrize the previously described model, we need to parametrize two statistics related to the productivity distribution: the 0th and ($\sigma - 1$)the y-bounded moments of the *productivity* distribution (Nigai, 2017). As described in (Dewitte, 2020), 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, we can calculate the necessary analytical expressions using the distributional parameters from our empirical analysis to capture heterogeneity. This is the standard approach in the literature. Following (Nigai, 2017; Dewitte, 2020), we can also capture heterogeneity directly from the empirical, *finite* data. To compare GFT obtained assuming a parametric distribution and GFT obtained from the finite data, we perform a parametric bootstrap. This parametric bootstrap generates a range of finite sample estimates under the hypothesis that a certain parametric distribution generates the observed data (Dewitte, 2020).

E.4.1 Continuum of firms

When there is 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:

$$\mu_y^r = \int_y^\infty x^r g(x) dx. \tag{25}$$

The analytical expressions of these parametric implied population moments are gathered in Table 4 and 5 for all distributions considered. As bounded moments are not generally available, the mathematical elaboration on obtaining these expressions can be found in the section F.

E.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:

$$m_y^r = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(x > y\right) x^r. \tag{26}$$

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, we perform a parametric bootstrap. 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_y^r for $r = 0, 1$, as specified in equation 26 and corresponding GFT_B;

2. Estimate the parameters $\boldsymbol{\theta}$ of the distribution using MLE, calculate the parametric plug-in population moments as specified in equation 25, $\hat{\mu}^r(y|\hat{\boldsymbol{\theta}})$ for $r = 0, 1$, and corresponding $G\hat{F}T(\hat{\boldsymbol{\theta}})$;

3. $H_0 : GFT_B = G\hat{F}T(\hat{\boldsymbol{\theta}})$;

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 $(m_y^r)^*$ and calculate the corresponding GFT_B^*.[12]

[12]Note that we do not re-fit the parametric distribution to the bootstrap sample. The vastness of the dataset at our 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.

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

$$\hat{p}_l = \frac{1}{N+1}\left[\sum_{n=1}^{N} \mathbb{I}\left(GFT_B^* \geq GFT_B\right) + 1\right]; \qquad \hat{p}_r = \frac{1}{N+1}\left[\sum_{n=1}^{N} \mathbb{I}\left(GFT_B^* \leq GFT_B\right) + 1\right]. \tag{27}$$

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.[13] We address each issue below and argue that these complications have little influence on our results.

1. y can take values within the boundaries of the data but are not observed. We use the 'approxfun' interpolation function of the R base distribution to approximate the statistics for such lower bounds.[14] 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 ($y < x_{min}$):

$$\mu_y^r = \underbrace{\sum \mathbb{I}\left(y < x < x_{min}\right) x^r}_{\text{unobserved}} + \underbrace{\frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(x \geq x_{min}\right) x^r}_{observed}. \tag{28}$$

   The error arising from neglecting the unobserved part of the distribution is likely small as (i) the smallest observation x_{min} in our dataset 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

[13] We thank Gonzague Vannoorenberghe for pointing this out.
[14] All code available on request.

future profits (Impullitti et al., 2013). 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.

Appendix F Analytical expressions of μ_y^r

F.1 Pareto

$$\begin{aligned}\mu_y^r &= \int_y^\infty x^r \frac{kx_{min}^k}{x^{k+1}} dx \\ &= kx_{min}^k \frac{-y^{r-k}}{r-k} \qquad \text{if } k > r\end{aligned} \tag{29}$$

F.2 Inverse Pareto

$$\begin{aligned}\mu_y^r &= \int_y^{x_{max}} x^r \frac{kx_{max}^{-k}}{x^{-k+1}} dx \\ &= kx_{max}^{-k} \frac{x_{max}^{r+k} - y^{r+k}}{r+k}\end{aligned} \tag{30}$$

F.3 Lognormal

$$\begin{aligned}\mu_y^r &= \int_y^\infty x^r \frac{1}{xVar\sqrt{2\pi}} e^{-(lnx-\mu)^2/2Var^2} dx \\ &= \int_y^\infty e^{rlnx} \frac{1}{xVar\sqrt{2\pi}} e^{-(lnx-\mu)^2/2Var^2} dx\end{aligned} \tag{31}$$

Note that

$$\begin{aligned}rlnx - (lnx-\mu)^2/2Var^2 &= \frac{2Var^2 rlnx - (lnx)^2 - \mu^2 + 2\mu lnx}{2Var^2} \\ &= -\frac{(lnx)^2 - 2(Var^2 r + \mu)lnx + \left((Var^2 r + \mu)\right)^2 - (Var^2 r + \mu)^2 + \mu^2}{2Var^2} \\ &= -\frac{\left[lnx - \left(Var^2 r + \mu\right)\right]^2}{2Var^2} + \frac{(Var^2 r + \mu)^2 - \mu^2}{2Var^2} \\ &= -\frac{\left[lnx - \left(Var^2 r + \mu\right)\right]^2}{2Var^2} + \frac{r\left(rVar^2 + 2\mu\right)}{2}\end{aligned}$$

so that

$$
\begin{aligned}
\mu_y^r &= e^{\frac{r\left(rVar^2+2\mu\right)}{2}} \int_y^\infty \frac{1}{xVar\sqrt{2\pi}} e^{-\frac{\left[lnx-\left(Var^2r+\mu\right)\right]^2}{2Var^2}} dx \\
&\text{let } z = \frac{lnx-\left(rVar^2+\mu\right)}{Var},\ dz = \frac{dx}{xVar} \\
&= e^{\frac{r\left(rVar^2+2\mu\right)}{2}} \int_{\frac{lny-\left(rVar^2+\mu\right)}{Var}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dx \\
&= e^{\frac{r\left(rVar^2+2\mu\right)}{2}} \left[1-\Phi\left(\frac{lny-\left(rVar^2+\mu\right)}{Var}\right)\right]
\end{aligned} \tag{32}
$$

F.4 Weibull[15]

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r \frac{k}{s}\left(\frac{x}{s}\right)^{k-1} e^{-\left(\frac{x}{s}\right)^k} dx \\
&\text{let } z = \left(\frac{x}{s}\right)^k, dz = \frac{k}{s}\left(\frac{x}{s}\right)^{k-1} dx \\
&\text{s.t. } x = sz^{\frac{1}{k}} \\
&= \int_{\left(\frac{y}{s}\right)^k}^\infty s^r z^{\frac{r}{k}} e^{-z} dz \\
&= s^r \int_{\left(\frac{y}{s}\right)^k}^\infty z^{\left(\frac{r}{k}+1\right)-1} e^{-z} dz \\
&= s^r \Gamma\left(\frac{r}{k}+1, \left(\frac{y}{s}\right)^k\right)
\end{aligned} \tag{33}
$$

where $\Gamma(,)$ denotes the upper incomplete gamma function.

[15] The bounded moments of the exponential distribution are obtained setting k=1.

F.5 Fréchet

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r \frac{k}{s}\left(\frac{x}{s}\right)^{-1-k} e^{-\left(\frac{x}{s}\right)^{-k}} dx \\
&\qquad \text{let } z = \left(\frac{x}{s}\right)^{-k}, dz = \frac{-k}{s}\left(\frac{x}{s}\right)^{-k-1} dx \\
&\qquad \text{s.t. } x = s z^{-\frac{1}{k}} \\
&= -\int_{\left(\frac{y}{s}\right)^{-k}}^{0} s^r z^{-\frac{r}{k}} e^{-z} dz, \qquad \text{if } k > 0 \\
&= \int_0^{\left(\frac{y}{s}\right)^{-k}} s^r z^{-\frac{r}{k}} e^{-z} dz \\
&= s^r \int_0^{\left(\frac{y}{s}\right)^{-k}} z^{1-\left(\frac{r}{k}\right)-1} e^{-z} dz \\
&= s^r \left[1 - \Gamma\left(1 - \frac{r}{k}, \left(\frac{y}{s}\right)^{-k}\right)\right] \qquad \text{if } k > r
\end{aligned}
\tag{34}
$$

F.6 Burr

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r \frac{\frac{kc}{s}\left(\frac{x}{s}\right)^{c-1}}{\left(1+\left(\frac{x}{s}\right)^c\right)^{k+1}} dx \\
&\qquad \text{let } z = \left(\frac{x}{s}\right)^c, dz = \frac{c}{s}\left(\frac{x}{s}\right)^{c-1} dx \\
&\qquad \text{s.t. } x = s z^{\frac{1}{c}} \\
&= \int_{\left(\frac{y}{s}\right)^c}^\infty s^r z^{\frac{r}{c}} \frac{k}{(1+z)^{k+1}} dz, \qquad \text{if } c > 0 \\
&= s^r k \int_{\left(\frac{y}{s}\right)^c}^\infty z^{\frac{r}{c}} \frac{1}{(1+z)^{k+1}} dz \\
&= s^r k \int_{\left(\frac{y}{s}\right)^c}^\infty z^{\left(\frac{r}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz \\
&= s^r k \int_{\left(\frac{y}{s}\right)^c}^\infty z^{\left(\frac{r}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz \\
&= s^r k \left[\int_0^\infty z^{\left(\frac{r}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz - \int_0^{\left(\frac{y}{s}\right)^c} z^{\left(\frac{r}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz\right] \\
&\quad u = \frac{z}{1+z}, du = \frac{1}{(1+z)^2} \\
&\quad z = \frac{u}{1-u} \\
&= s^r k \left[\int_0^1 \left(\frac{u}{1-u}\right)^{\left(\frac{r}{c}+1\right)-1} \frac{1}{\left(1+\frac{u}{1-u}\right)^{k-1}} du - \int_0^{\frac{\left(\frac{y}{s}\right)^c}{1+\left(\frac{y}{s}\right)^c}} \left(\frac{u}{1-u}\right)^{\left(\frac{r}{c}+1\right)-1} \frac{1}{\left(1+\frac{u}{1-u}\right)^{k-1}} du\right] \\
&= s^r k \Bigg[\int_0^1 u^{\left(\frac{r}{c}+1\right)-1} (1-u)^{k-1-\left(\frac{r}{c}+1\right)+1} du \\
&\qquad\qquad - \int_0^{\frac{\left(\frac{y}{s}\right)^c}{1+\left(\frac{y}{s}\right)^c}} \int_0^{\left(\frac{y}{s}\right)^c} u^{\left(\frac{r}{c}+1\right)-1} (1-u)^{k-1-\left(\frac{r}{c}+1\right)+1} du\Bigg] \\
&= s^r k \left[\int_0^1 u^{\left(\frac{r}{c}+1\right)-1} (1-u)^{k-\left(\frac{r}{c}+1\right)} du - \int_0^{\frac{\left(\frac{y}{s}\right)^c}{1+\left(\frac{y}{s}\right)^c}} u^{\left(\frac{r}{c}+1\right)-1} (1-u)^{k-\left(\frac{r}{c}+1\right)} duz\right] \\
&= s^r k \left[\boldsymbol{B}\left(\frac{r}{c}+1, k-\frac{r}{c}\right) - \boldsymbol{B}\left(\frac{\left(\frac{y}{s}\right)^c}{1+\left(\frac{y}{s}\right)^c}; \frac{r}{c}+1, k-\frac{r}{c}\right)\right] \qquad \text{if } c > r, kc > r \qquad (35)
\end{aligned}
$$

where $\boldsymbol{B}(a,b)$ stands for the beta function, while $\boldsymbol{B}(x,a,b)$ stands for the lower incomplete beta function with upper bound x.

F.7 Generalized Gamma[16]

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r \frac{c}{s^k \Gamma(\frac{k}{c})} x^{k-1} e^{-\left(\frac{x}{s}\right)^c} dx \\
&\qquad \text{let } z = \left(\frac{x}{s}\right)^c, dz = \frac{c}{s}\left(\frac{x}{s}\right)^{c-1} dx \\
&\qquad \text{s.t. } x = s z^{\frac{1}{c}} \\
&= \int_{\left(\frac{y}{s}\right)^c}^\infty s^r \frac{z^{\frac{r}{c}}}{\Gamma(\frac{k}{c})} \left(\frac{s z^{\frac{1}{c}}}{s}\right)^{(k-1)-(c-1)} e^{-z} dz, \qquad \text{if } c > 0 \\
&= \frac{s^r}{\Gamma(\frac{k}{c})} \int_{\left(\frac{y}{s}\right)^c}^\infty z^{\frac{r+k}{c}-1} e^{-z} dz \\
&= \frac{s^r}{\Gamma(\frac{k}{c})} \Gamma\left(\frac{r+k}{c}, \left(\frac{y}{s}\right)^c\right) \qquad (36)
\end{aligned}
$$

F.8 Finite Mixture Model

The statistics for a Finite Mixture Model can easily be obtained from the calculated statistics for the underlying individual distributions on which the mixture consists. For a mixture of the form:

$$
g(x|\mathbf{\Psi}) = \sum_{i=1}^{I} \pi_i m_i(x|\boldsymbol{\theta}_i), \qquad \pi_i \geq 0, \quad \sum_{i=1}^{I} \pi_i = 1, \qquad (37)
$$

we obtain, due to its additivity and applying the sum rule in integration:

$$
\mu_y^r = \int_y^\infty x^r g(x|\mathbf{\Psi}) dx = \int_y^\infty x^r \sum_{i=1}^{I} \pi_i m_i(x|\boldsymbol{\theta_i}) dx = \sum_{i=1}^{I} \pi_i \int_y^\infty x^r m_i(x) dx = \sum_{i=1}^{I} \pi_i (\mu_i)_y^r. \qquad (38)
$$

F.9 Piecewise composite

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r g(x|\boldsymbol{\theta}) dx \\
&= \begin{cases}
\frac{\alpha_1}{1+\alpha_1+\alpha_2} \frac{(\mu_1)_y^r - (\mu_1)_{c_1}^r}{M_1(c_1)} + \frac{1}{1+\alpha_1+\alpha_2} \frac{(\mu_2)_{c_1}^r - (\mu_2)_{c_2}^r}{M_2(c_2) - M_2(c_1)} + \frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{(\mu_3)_y^r}{1 - M_3(c_2)} & \text{if } \; 0 < y \leq c_2 \\
\frac{1}{1+\alpha_1+\alpha_2} \frac{(\mu_2)_y^r - (\mu_2)_{c_2}^r}{M_2(c_2) - M_2(c_1)} + \frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{(\mu_3)_{c_2}^r}{1 - M_3(c_2)} & \text{if } \; c_1 < y \leq c_2 \\
\frac{\alpha_2}{1+\alpha_1+\alpha_2} \frac{(\mu_3)_y^r}{1 - M_3(c_2)} & \text{if } \; c_2 < y < \infty
\end{cases} \qquad (39)
\end{aligned}
$$

[16] The bounded moments of the Gamma distribution are obtained setting c=1.

F.10 Right-Pareto Lognormal

$$
\begin{aligned}
\mu_y^r &= \int_y^\infty x^r k_2 x^{-k_2-1} e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)dx \\
&= k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \int_y^\infty x^{\sigma-k_2-2}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)dx \\
&\qquad dv = x^{\sigma-k_2-2}dx, v = \frac{x^{\sigma-k_2-1}}{\sigma-k_2-1} \\
&\qquad u = \Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right), du = d\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right) \\
&= k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \left[\frac{x^{\sigma-k_2-1}}{\sigma-k_2-1}\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right)\right]_y^\infty \\
&\quad - k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \int_y^\infty \frac{x^{\sigma-k_2-1}}{\sigma-k_2-1} d\Phi\left(\frac{lnx-\mu-k_2Var^2}{Var}\right) \\
&= k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \left[0 - \frac{x_{ij^*}^{\sigma-k_2-1}}{\sigma-k_2-1}\Phi\left(\frac{lny-\mu-k_2Var^2}{Var}\right)\right] \\
&\quad - k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \int_y^\infty \frac{x^{\sigma-k_2-1}}{\sigma-k_2-1}\frac{1}{xVar\sqrt{2\pi}} e^{-\frac{\left[lnx-\mu-k_2Var^2\right]^2}{2Var^2}} dx
\end{aligned} \tag{40}
$$

The last integral resembles the bounded moment condition of the Lognormal distribution solved earlier with moment $(r-k_2)$ and mean $(\mu+k_2Var^2)$ so that

$$
\begin{aligned}
\mu_y^r = &-k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \frac{x_{ij^*}^{\sigma-k_2-1}}{\sigma-k_2-1}\Phi\left(\frac{lny-\mu-k_2Var^2}{Var}\right) \\
&- \frac{k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}}}{r-k_2} e^{\frac{(r-k_2)\left((r-k_2)Var^2+2(\mu+k_2Var^2)\right)}{2}} \Bigg[1 \\
&\qquad -\Phi\left(\frac{lny-\left((r-k_2)Var^2-(\mu+k_2Var^2)\right)}{Var}\right)\Bigg]
\end{aligned} \tag{41}
$$

Note that

$$e^{k_2\mu+\frac{k_2^2 Var^2}{2}+\frac{(r-k_2)\left((r-k_2)Var^2+2(\mu+k_2 Var^2)\right)}{2}}$$

$$e^{\frac{2k_2\mu+k_2^2 Var^2+(r-k_2)\left[rVar^2+2\mu+k_2 Var^2\right]}{2}}$$

$$e^{\frac{2k_2\mu+k_2^2 Var^2+r^2 Var^2+2\mu r+k_2 r Var^2-k_2 r Var^2-2\mu k_2+k_2^2 Var^2}{2}}$$

$$e^{\frac{r^2 Var^2+2\mu r}{2}}$$

so that we get

$$\begin{aligned}\mu_y^r = & -k_2 e^{k_2\mu+\frac{k_2^2 Var^2}{2}} \frac{x_{ij^*}^{\sigma-k_2-1}}{\sigma-k_2-1}\Phi\left(\frac{lny-\mu-k_2 Var^2}{Var}\right)\\ & -\frac{k_2}{r-k_2}e^{\frac{r^2 Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2-\mu}{Var}\right)\end{aligned} \tag{42}$$

F.11 Left-Pareto Lognormal

$$\begin{aligned}\mu_y^r &= \int_y^\infty x^r x^{k_1-1}e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\Phi^c\left(\frac{lnx-\mu+k_1 Var^2}{Var}\right)dx\\ &= k_1 e^{k_2\mu+\frac{k_2^2 Var^2}{2}}\left[\left(\frac{-(y)^{\sigma-k_2-1}}{\sigma-k_2-1}\right)-e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\int_y^\infty x^{\sigma-2+k_1}\Phi\left(\frac{lnx-\mu+k_1 Var^2}{Var}\right)dx\right]\\ &= -k_1 e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\frac{x_{ij^*}^{\sigma+k_1-1}}{\sigma+k_1-1}\Phi^c\left(\frac{lny-\mu+k_1 Var^2}{Var}\right)\\ &\quad+\frac{k_1}{r+k_1}e^{\frac{r^2 Var^2+2\mu r}{2}}\Phi^c\left(\frac{lny-rVar^2+\mu}{Var}\right)\end{aligned} \tag{43}$$

F.12 Double-Pareto Lognormal

$$\begin{aligned}
\mu_y^r = {} & \frac{k_2 k_1}{k_2 + k_1} \int_y^\infty x^r x^{-k_2 - 1} e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \Phi \left(\frac{lnx - \mu - k_2 Var^2}{Var} \right) dx \\
& + \frac{k_2 k_1}{k_2 + k_1} \int_y^\infty x^r x^{k_1 - 1} e^{-k_1 \mu + \frac{k_1^2 Var^2}{2}} \Phi^c \left(\frac{lnx - \mu + k_1 Var^2}{Var} \right) dx \\
& = -\frac{k_2 k_1}{k_2 + k_1} e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \frac{x_{ij^*}^{\sigma - k_2 - 1}}{\sigma - k_2 - 1} \Phi \left(\frac{lny - \mu - k_2 Var^2}{Var} \right) \\
& \quad - \frac{k_2 k_1}{k_2 + k_1} \frac{1}{r - k_2} e^{\frac{r^2 Var^2 + 2\mu r}{2}} \Phi^c \left(\frac{lny - rVar^2 - \mu}{Var} \right) \\
& - \frac{k_2 k_1}{k_2 + k_1} e^{-k_1 \mu + \frac{k_1^2 Var^2}{2}} \frac{x_{ij^*}^{\sigma + k_1 - 1}}{\sigma + k_1 - 1} \Phi \left(\frac{lny - \mu + k_1 Var^2}{Var} \right) \\
& \quad - \frac{k_2 k_1}{k_2 + k_1} \frac{1}{r + k_1} e^{\frac{r^2 Var^2 + 2\mu r}{2}} \Phi^c \left(\frac{lny - rVar^2 - \mu}{Var} \right)
\end{aligned} \tag{44}$$

Appendix References

Arkolakis, C. (2016). A Unified Theory of Firm Selection and Growth. *The Quarterly Journal of Economics 131*(1), 89–155.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review 101*(4), 1312–44.

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.

Bernard, A. B., S. J. Redding, and P. K. Schott (2009). Products and productivity. *The Scandinavian Journal of Economics 111*(4), 681–709.

Bloom, N. and J. V. Reenen (2011). Chapter 19 - human resource management and productivity. Volume 4 of *Handbook of Labor Economics*, pp. 1697–1767. Elsevier.

Caliendo, L., G. Mion, L. D. Opromolla, and E. Rossi-Hansberg (2020). Productivity and Organization in Portuguese Firms. *Journal of Political Economy forthcoming.*

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.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

Dewitte, R. (2020). From Heavy-tailed Micro to Macro: On the characterization of firm-level heterogeneity and its aggregation properties. MPRA Paper 103170, 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.

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

Follmann, D. A. and D. Lambert (1991). Identifiability of finite mixtures of logistic regression models. *Journal of Statistical Planning and Inference 27*(3), 375–381.

Fop, M., T. B. Murphy, et al. (2018). Variable selection methods for model-based clustering. *Statistics Surveys 12*, 18–65.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Grün, B. (2018). Model-based clustering. arXiv preprint 1807.01987, arXiv.

Grün, B. and F. Leisch (2008). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Journal of classification 25*(2), 225–247.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of classification 17*(2).

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Impullitti, G., A. A. Irarrazabal, and L. D. Opromolla (2013). A theory of entry into and exit from export markets. *Journal of International Economics 90*(1), 75 – 90.

Kasahara, H. and J. Rodrigue (2008). Does the use of imported intermediates increase productivity? plant-level evidence. *Journal of Development Economics 87*(1), 106–118.

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

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

McLachlan, G. J. and D. Peel (2000). *Finite mixture models*. New York: Wiley Series in Probability and Statistics.

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.

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

Rossi-Hansberg, E. and M. L. J. Wright (2007). Establishment size dynamics in the aggregate economy. *American Economic Review 97*(5), 1639–1666.

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.