Gains From Parameters: Heterogeneity in the Productivity Distribution and International Trade*

Ruben Dewitte†, Michel Dumont‡, Glenn Rayp§, Peter Willemé¶

19th March 2019

Abstract

Parametric approximations of the productivity distribution determine multiple research fields, including the literature on firm-driven aggregate growth, trade and granularity. Strikingly, most distributions considered up till now are the result of a homogeneous dynamic productivity process. This stands in stark contrast with evidence of productivity evolving endogenously with firm-level characteristics such as exporting and innovation This paper claims that heterogeneity in productivity can be captured more adequately when allowing for heterogeneity in its parametric approximation. To this purpose, we introduce Finite Mixture Modelling (FMM). FMMs build on the existence of discrete subpopulations in the data. As such, they allow for a very flexible fit to the heterogeneity distribution and can be supported by a very general specification of firm dynamics. We establish their excellent empirical performance on the population of Portuguese firms. A Gains From Trade exercise demonstrates the need to combine multiple distributions to obtain representative welfare impacts.

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

JEL Codes: L11, F11, F12

---

*We thank conference participants at the 2018 ETSG conference.

†Ghent University, Department of Economics

‡Federal Planning Bureau

§Ghent University, Department of Economics

¶Federal Planning Bureau

1 Introduction

Ever since Melitz (2003), firm-level heterogeneity is predominantly defined in terms of productivity. This heterogeneity in productivity is then usually captured by means of a single parametric distribution, such as the Pareto or the Lognormal distribution (Chaney, 2008; Head et al., 2014). The importance of such parametric approximation can be appraised by its large influence on various research fields. For instance, firm-level dynamics in aggregate growth models are determined by the parametric approximation of the stationary productivity distribution (Luttmer, 2007; Arkolakis, 2016; Rossi-Hansberg and Wright, 2007). Also, the propagation of firm-level volatility to the aggregate level is strongly dependent on a Pareto right tail of the productivity distribution (di Giovanni and Levchenko, 2012; Gabaix, 2011; Carvalho and Grassi, ming). Lastly, it is recognized in the trade literature that differences in the choice of the productivity distribution significantly affects Gains From Trade (Head et al., 2014; Aw et al., 2011; Nigai, 2017) and alters the channels through which trade operates (Fernandes et al., 2018; Bas et al., 2017; Arkolakis et al., 2012; Melitz and Redding, 2015).

Strikingly, most distributions considered are the result of a homogeneous dynamic productivity process, i.e. the evolution of productivity is independent of firm-level characteristics other than productivity itself. This stands in stark contrast with evidence of the productivity evolution being endogenous, impacted by firm-level characteristics such as exporting (De Loecker, 2013) and innovation (Aw et al., 2011). It is not unreasonable to assume that heterogeneous dynamic processes will lead to heterogeneity in the resulting productivity distributions, an idea already hinted at by Luttmer (2007) and Costantini and Melitz (2008).

This paper addresses the question whether heterogeneity in productivity can be captured more adequately when allowing for heterogeneity in its parametric approximation. To this purpose, we introduce Finite Mixture Modelling (FMM). A FMM is essentially a weighted sum of an à 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 renders them favorable both from an empirical and theoretical point of view.

We schowcase the excellent empirical performance of FMMs on the universe of Portuguese firm's productivity in 2006. Being the first to have access to such representative dataset on the firm size distribution allows us evaluate the performance of parametric distributions for the complete as well as to focus on both the *left*- and right- tail of the productivity distribution. Moreover, it insulates us from erroneous conclusions due to truncated or unrepresentative data in the left tail of the distribution (Perline, 2005). We compare FMMs with a large range of alternative distributional forms and find all statistical tests (Log-likelihood, Aikaike and Bayesian Information Criteria, Kolmogorov-Smirnov test statistic, . . .) to favour FMMs for the complete as well as for the left- and right-tail of the distribution.

From a theoretical point of view, a FMM is a generalization of existing distributions and their underlying generative process. An often-raised concern of the Pareto distribution, for instance, is its far-reaching simplification of reality (see f.i. Bas et al. (2017); Melitz and Redding (2015)). This paper shows that some Pareto-tailed distributions are simply exogenously imposed and rigidly structured restrictions of FMMs. In the same vein, Luttmer (2007) acknowledges the simplicity of implementing FMMs when generalizing his homogeneous firm dynamics to allow for productivity evolving heterogeneously accross industries. Overall, FMMs alleviate the danger of misspecification and/or oversimplification when imposing homogeneity restrictions.

A Gains From Trade (GFT) exercise demonstrates the importance of correctly approximating the heterogeneity distribution in the class of heterogeneous firms models à la Melitz (2003) and underlines the straightforward implementation of FMMs into such models. A vast literature finds that different parametric approximations result in significant differences in GFT (Head et al., 2014; Fernandes et al., 2018; Bee and Schiavo, 2017; Nigai, 2017) and alters the channels through which trade operates (Fernandes et al., 2018; Bas et al., 2017; Arkolakis et al., 2012; Melitz and Redding, 2015). We integrate a large range of new distributions, including but not limited to FMMs, into the GFT literature. We find that one-component distributions do not deliver representative GFT, while increasing the components of individual distributions results in GFT closer to the truth.

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

2 Literature Review

We provide an overview of the size distribution literature related to the literature on firm size/productivity distributions. This overview explains why the Pareto distribution can only match the tail of size distributions while hump-shaped distributions such as the Lognormal or the Weibull distribution cannot match both the tail and the bulk of the distribution simultaneously. Size distributions are best approximated by a combination of distributions, which we find to exist in three forms: piecewise, multiplicative and finite mixture distributions. We argue that FMMs are preferable both from an empirical and theoretical point of view due to their flexible, semi-parametric nature.

The **Pareto distribution** has been dominating the class of heterogeneous firms models (Melitz, 2003). Even though the Melitz (2003)-model is not restricted to this distributional choice, its convenience lead to a widespread reliance on the Pareto distribution for social welfare and economic policy analysis.[1] The fit of a Pareto distribution is usually evaluated using its CDF, which follows a straight line on a log-log scale with the shape parameter as coefficient (k):

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

Figure 1 compares a fitted Pareto survival function (=1−CDF) with the empirical survival function of Portuguese productivity[2] on a log-log scale for the complete dataset (upper panel). It is immediately apparent that the Pareto distribution is not a good fit to the complete distribution due to the existence of a hump in the middle (see also the Probability Density Function (PDF) in Appendix Figure A2). Nevertheless, both the left- (lower left panel) and right tail (lower right panel) showcase linearity of the CDF and survival function respectively on a log-log scale, in line with Pareto behaviour in the tails of the distribution.[3]

[FIGURE 1]

Just as every bend line looks straight when you zoom in close enough though, so too does firms sales only appear to be straight when truncated sufficiently.[4] Consequently, this apparent straight line behaviour of the tails can 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 distribution.[5] Proof of which is the the

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

[2]Productivity is computed as demeaned domestic sales to the power of 0.33. See Section 4 for more details.

[3]The Inverse Pareto distribution is specified as

$$G_{IP}(\omega; \omega_{max}, k_1) = 1 - \left(\frac{\omega_{max}}{\omega}\right)^{-k}, \qquad \omega \leq \omega_{max}.$$

[4]Note that the influential paper of Axtell (2001) does not rely on truncated data but unintentionally favours the Pareto distribution due to data binning (Virkar and Clauset, 2014) and methodological choices (Clauset et al., 2009) characteristic of that time. See Dewitte (2019) for a note outlining the reasons why we should not rely on Axtell (2001) as evidence in favour of Pareto anymore.

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

performance of the Lognormal distribution in the lower panels of Figure 1.

Alternative **hump-shaped distributions** such as the Weibull (Bee and Schiavo, 2017) and Lognormal (Fernandes et al., 2018; Head et al., 2014; Eeckhout, 2004) are claimed to provide a better fit to complete size distributions. In the firm size literature, this claim is usually supported by comparing with a limited number of alternative distributions (mosty Pareto) using the low-powered R-squared (Clauset et al., 2009). Even though hump-shaped distributions such as the Logormal can adequately fit the tail or the bulk of the empirical distribution, they cannot do so simultaneously. This is easily observable from the upper panel of Figure 1 where the single Lognormal distribution is not able to match the right the tail of the complete productivity distribution while matching the bulk satisfactorily.

As single distributions were not capable of matching both the bulk and the tail(s) of the productivity distribution simultaneously, the literature resorted to combinations of distributions. Such combinations are found to exist in three forms. We start by discussing the **piecewise or composite distributions** with probability density specified as:

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

where $\forall i \in I : m_i^*(\omega) = \frac{m_i(\omega)}{\int_{c_{i-1}}^{c_i} m_i(\omega) d\omega}$ is the truncated probability density function (PDF) of $m_i(\omega)$. For this distribution to be well-behaved, additional differentiability and continuity conditions are imposed on the truncation points (Bakar et al., 2015). While these composite distributions can be formed from many individual parametric distributions (Bakar et al., 2015), they are mostly focused on Pareto-tailed Lognormal distributions. The Pareto-tailed Lognormal has been applied in the city size literature (Ioannides and Skouras, 2013; Luckstead and Devadoss, 2017), while the Pareto-Right-tailed Lognormal version was applied by

Nigai (2017) to the Melitz (2003)-model to calculate GFT. A disadvantage of the composite distributions is the ambiguity surrounding their generative process. For instance, it is yet unclear which firm dynamics could explain the existence of hard cutoffs that separate the Lognormal from the Pareto distribution.

Alternatively, we can rely on **multiplicative distributions** to fit both the bulk and the tail of size distributions. Reed and Jorgensen (2004) proposed to multiply a Lognormal with a (Double-)Pareto distribution, resulting in the so-called Double-Pareto Lognormal distribution. It is found to approximate well city size distributions (Reed, 2002; Giesen et al., 2010). Its fit to firm size distributions has, to our knowledge, not been investigated yet. Moreover, a generative process for this distribution exists is applicable to heterogeneous firms models (Arkolakis, 2016). Interestingly, the Double-Pareto Lognormal distribution can be seen as a structured infinite mixture of Lognormal distributions with fixed mixing parameter firm age T (Reed, 2002).

Therefore, this Double-Pareto Lognormal distribution, being based on a single, pre-specified and infinite-dimensional mixing parameter, carries the danger of misspecification and/or oversimplification. After all, firm age is not the only possible mixing parameter we can think of. It has already been argued that firm dynamics, and therefore the resulting firm size distributions, could be endogenous to exporting (De Loecker, 2013), innovation (Costantini and Melitz, 2008; Aw et al., 2011; Atkeson and Burstein, 2010), industrial linkages (Luttmer, 2007), ... Overall, there are "many sources of heterogeneity that support the idea of discrete subpopulations likely to differ in important characteristics ..." (Perline (2005),p.80).

Mixture distributions are flexible distribution that allow to capture such unknown sources of underlying heterogeneity. FMMs are essentially a weighted sum of I individual densities $m_i(\cdot)$:

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

where I represents the number of components or discrete subpopulations, π_i is the prior probability of belonging to component i, $\boldsymbol{\theta_i}$ the component-specific parameter vector of density $m_i(\cdot)$ and $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{K-1}, \theta_1, \ldots, \theta_i)$ is the vector of all model parameters (McLachlan and Peel, 2000). FMMs are also referred to as latent-class models given that the number of components, and thus also the mixing parameter itself, don't have to be specified à priori but are determined by the data. As such, a finite mixture model resembles a semi-parametric approach ideal to fully capture the heterogeneity of size distributions, as demonstrated by Kwong and Nadarajah (2019) on city size distributions and Miljkovic and Grün (2016) on actuarial loss data.[6] The generative process of a FMM corresponds to a simple combination of the generative processes of the underlying individual densities and is therefore easily integrable into existing models of firm dynamics (see f.i. Luttmer (2007)). In the same vein, FMM can easily be integrated in the class of heterogeneous firms models à la Melitz (2003).[7]

3 Methodology

From the previous section, it became clear that the literature on firm size distributions lacks a clear statistical framework that differentiates between a sufficiently large number of distributions over a consistent data range. In this section, we establish a methodology that uniformly fits the Pareto, hump-shaped and combinations of distributions (composite, multiplicative and mixture) both to the tails and the entire productivity distribution. We then present the statistical tests that will be used to differentiate between the fitted distributions.

[6] A semi-parametric approach is to be favoured over a nonparametric approach in the case of heavy-tailed distributions such as firm size. This because the heavy tails renders nonparametric procedures less efficient (Clauset et al., 2009).

[7] A unique equilibrium is ensured in the Melitz (2003)-model if the zero-profit cutoff monotonically decreases with productivity. This corresponds with the sufficient condition of $\frac{g(\omega)\omega}{1-G(\omega)}$ increasing to infinity on $(0, \infty)$. Finite mixtures of individual densities conform this sufficient condition also ensure a unique equilibrium in the Melitz (2003)-model, as is demonstrated in section 6.

3.1 Distribution fitting

We rely on augmented maximum likelihood (ML) over all firms $f \in F$ to fit all considered distributions to the data: (Inverse) Pareto,, hump-shaped and combinations of distributions (composite, multiplicative and mixture).[8],[9] In the case of FMMs, ML is wrapped in an Expectation-Maximization (EM) algorithm to capture the underlying heterogeneity of the distribution. We develop on these estimation methods in order to fit these distributions both to complete and truncated data. This will allow us to single out and focus on tail performance as well as generalizing the estimation methods to unrepresentative/truncated data.

3.1.1 (Inverse) Pareto

Complete data The ML estimator for the shape parameter k over all firms $f \in F$ can easily be obtained as

$$k_{IP} = \left[\frac{1}{F} \sum_{f=1}^{F} ln \frac{\omega_{max}}{\omega_f} \right]^{-1}, \qquad k_P = \left[\frac{1}{F} \sum_{f=1}^{F} ln \frac{\omega_f}{\omega_{min}} \right]^{-1}.$$

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

[8]Popular fitting techniques in the firm size literature rely on the minimization of squared errors between a log-linearization of the theoretical and empirical PDF's/CDF's (Bas et al., 2017; di Giovanni and Levchenko, 2013; Nigai, 2017; Axtell, 2001; Bee and Schiavo, 2017; Head et al., 2014; Freund and Pierola, 2015). 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 what extent the standard errors obtained from these methods are valid (Clauset et al., 2009). Maximum likelihood methods do not suffer from such problems.

[9]See Appendix Tables A7,A8 and A9 for an overview of the distribution specifications.

Truncated data If we want to fit the (Inverse) Pareto distribution to just the tails, we need to adapt the ML-procedure accordingly. The truncation has no influence on the above-specified ML-estimator for the shape parameter k , as the truncation points are parameters themselves. We do have to adapt the estimation method for the truncation points though, as we don't want to specify them arbitrarily.

There is no consensus on the practices to determine the (maximum) minimum of the (Inverse) Pareto-distribution. Some rely on visual techniques, 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). Some use export sales, and assume as such a truncation parameter equal to the minimum of sales, f.i. Freund and Pierola (2015). Some determine their minimum to ensure a Pareto parameter large enough to deliver finite moments when calibrating their theoretical models (Head et al., 2014; Bee and Schiavo, 2017). Others estimate the minimum, assuming a mixed Log-normal/Pareto distribution (Bakar and Nadarajah, 2013; Nigai, 2017; Malevergne et al., 2011). Such methods are however, either subject to possibly large measurement errors and inconsistencies or restrictive in their need to assume a distributional relation both for the bulk and the tail of the distribution.

Obtaining an accurate estimate for the lower bound is, however, vital to the accuracy of the estimated shape parameter $\hat{k}$. Choosing a minimum too low results in a biased shape parameter, as we will be fitting a power-law to non-power-law data. Choosing a value too high, on the other hand, increases the statistical error and bias from finite size effects on the shape parameter, as we throw legitimate data points away. Moreover, it is widely documented that the minimum and shape parameter of the (Truncated) Pareto distribution exhibit a positive correlation Head et al. (2014); Bee and Schiavo (2017); Freund and Pierola (2015); Eeckhout (2004); di Giovanni and Levchenko (2013).

In order to obtain an accurate estimate for the lower bound, we rely on a formal decision rule developed by Clauset et al. (2009). For the ordered productivity set $\{\omega_f; f = 1, \ldots, F\}$, we evaluate every ω_f as a potential ω_{min} (ω_{max}), estimating the ML estimate of the power-law exponent k. We then use the Kolmogorov-Smirnov statistic to select the optimum ω_{min} (ω_{max}). It is defined as the cut-off which minimizes the quantity

$$T_{KS,\hat{\omega}_{max}} = \sup_{\omega \leq \hat{\omega}_{max}} |S(\omega) - G_P(\omega; \hat{k}, \hat{\omega}_{max})|$$

$$T_{KS,\hat{\omega}_{min}} = \sup_{\omega \geq \hat{\omega}_{min}} |S(\omega) - G_{IP}(\omega; \hat{k}, \hat{\omega}_{min})|$$

where $S(\omega)$ is the CDF of the observed values for $\omega_f \geq \omega_{min}$ ($\omega_f \leq \omega_{max}$).

3.1.2 Pareto-tailed distributions

We focus on Pareto-Lognormal combinations of distributions as an example of Pareto-tailed distributions. This because it is the only distribution we know of that already has parametric specifications with Pareto tails on both ends of the distribution and has already been researched extensively in the academic literature.

From the composite distributions, we consider the Pareto-tailed Lognormal distribution as proposed by Luckstead and Devadoss (2017). This distribution consists of an Inverse Pareto left tail, a Lognormal body and a Pareto right tail. Limiting the boundaries of the composite distribution allows us to consider a Paret-Right-tailed (Ioannides and Skouras, 2013) and Left-tailed Lognormal respectively. The ML-estimator of these distributions has no closed form solution and needs to be approached numerically (Luckstead and Devadoss, 2017).

From the multiplicative distributions, we consider the Double-Pareto Lognormal distribution (Reed and Jorgensen, 2004). This distribution is the result of multiplying a Double Pareto, used by o.a. 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. Also for this distribution, the ML-estimator of these distributions has no closed form solution and needs to be approached numerically (Reed and Jorgensen, 2004).

3.1.3 FMM

FMMs can be estimated on many distributions. We restrict ourselves to a range of distributions that are commonly used in the economic literature and which ensure a unique equilibrium in the Melitz (2003)-model: the Lognormal, Weibull, Fréchet, Gamma, Burr and Exponential distribution.

Complete data The maximum likelihood estimation of a FMM (see eq. 1) is not straightforward. This because the number of components i is unobserved. The log-likelihood function can be written as

$$logL(\omega|\mathbf{\Psi}) = \sum_{f=1}^{F}\sum_{i=1}^{I} z_{fi}\left[log(\pi_i) + log(m_i(\omega|\theta_i))\right] \tag{2}$$

where z_{fi} is an unobserved component indicator equal to one if ω_f originated from subpopulation i and zero otherwise. Two steps need to be taken iteratively in order to be able to maximize this equation. The E-step of the s-th iteration consists of determining the conditional expectation of eq. 2 given the observed data and the current parameter estimates from iteration $s-1$:

$$\begin{aligned} Q(\mathbf{\Psi}|\mathbf{\Psi}^{(s-1)}) &= E\left[logL(\omega|\mathbf{\Psi})|\omega, \mathbf{\Psi}^{(s-1)}\right] \\ &= \sum_{f=1}^{F}\sum_{i=1}^{I} \pi_{fi}^{(s)}\left[log(\pi_i) + log(m_i(\omega|\boldsymbol{\theta}_i))\right] \end{aligned} \tag{3}$$

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

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

The M-step then, consists of maximizing the Q-function over the parameters Ψ:

$$\mathbf{\Psi}^{(s)} = \max_{\mathbf{\Psi}} Q(\mathbf{\Psi}|\mathbf{\Psi}^{(s-1)}) \tag{4}$$

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).

Truncated data The EM-algorithm can easily be adapted to fitting data only to truncated data within the boundaries $\omega \in \left[\omega^l, \omega^u\right]$. We specify the conditional densities

$$\begin{aligned}
g(\omega|\mathbf{\Psi}, \omega^l \leq \omega \leq \omega^u) &= \frac{\sum_{i=1}^{I} \pi_i m_i(\omega|\boldsymbol{\theta_i})}{G(\omega^u|\mathbf{\Psi}) - G(\omega^l|\mathbf{\Psi})} \\
&= \sum_{i=1}^{I} \pi_i \frac{M_i(\omega^u|\boldsymbol{\theta_i}) - M_i(\omega^l|\boldsymbol{\theta_i})}{G(\omega^u|\mathbf{\Psi}) - G(\omega^l|\mathbf{\Psi})} \frac{m_i(\omega|\boldsymbol{\theta_i})}{M_i(\omega^u|\boldsymbol{\theta_i}) - M_i(\omega^l|\boldsymbol{\theta_i})} \\
&= \sum_{i=1}^{I} \eta_i m_i(\omega|\boldsymbol{\theta_i}, \omega^l \leq \omega \leq \omega^u)
\end{aligned} \tag{5}$$

with $\eta_i > 0$, $\sum_{i=1}^{I} \eta_i = 1$. The Q-function becomes

$$\begin{aligned}
Q(\mathbf{\Psi}|\mathbf{\Psi}^{(s-1)}) &= E\left[logL(\omega|\mathbf{\Psi})|\omega, \mathbf{\Psi}^{(s-1)}\right] \\
&= \sum_{f=1}^{F} \sum_{i=1}^{I} \pi_{fi}^{(s)} \left[log(\eta_i) + log(m_i(\omega|\boldsymbol{\theta}_i, \omega^l \leq \omega \leq \omega^u))\right]
\end{aligned} \tag{6}$$

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

$$\pi_{fi}^{(s)} = \frac{\eta_i^{(s-1)} m_i(\omega_f|\boldsymbol{\theta}_i^{(s-1)}, \omega^l \leq \omega \leq \omega^u))}{\sum_{i=1}^{I} \eta_i^{(s-1)} m_i(\omega_f|\boldsymbol{\theta}_i^{(s-1)}), \omega^l \leq \omega \leq \omega^u)} = \frac{\pi_i^{(s-1)} m_i(\omega_f|\boldsymbol{\theta}_i^{(s-1)})}{\sum_{i=1}^{I} \pi_i^{(s-1)} m_i(\omega_f|\boldsymbol{\theta}_i^{(s-1)})}.$$

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

3.2 Distribution selection

We rely on multiple distinct criteria to differentiate between the distributions. First, we consider whether the proposed parametric distribution is a sufficiently good fit to the data. Then, we differentiate between distributions using information criteria.

Goodness of fit We use the Kolmogorov-Smirnov (KS) test to evaluate whether the proposed parametric distribution is a good fit for the empirical distribution. The KS-statistic quantifies the largest distance between the empirical distribution $S(\omega)$ and the parametric distribution $G(\omega; \cdot)$:

$$T_{KS} = \sup |S(\omega) - G(\omega; \cdot)|$$

In the same vein, we calculate the cumulative sum of distances between the empirical and parametric distribution to get an idea of the overall performance of the parametric distribution:

$$\sum_{f=1}^{F} KS_f = \sum_{f=1}^{F} |S(\omega) - G(\omega; \cdot)|.$$

Information Criteria We differentiate between distributions based on the negative log-likelihood, the Akaike or Bayesian Information Criteria. The Negative Log-Likelihood equals

$NLL = -\sum_{f=1}^{F} lng(\omega_f;\cdot)$. We can differentiate between two distributions is based on the ratio of their likelihoods:

$$LR = \sum_{f=1}^{F} ln \frac{g_1(\omega_f;\cdot)}{g_2(\omega_f,\cdot)} \tag{7}$$

with $g_{1,2}$ the probability densities of the rivaling distributions. If these distributions are *nested*, the test statistic amounts to minus two times this ratio, which follows a chi-squared distribution with 1 degree of freedom Wilks (1938).If these distributions are *non-nested*, the test statistic will be the sample average of this ratio, standardized by a consistent estimate of its standard deviation. The null hypothesis states that both classes of distributions are equally far (in the Kullback-Leibler divergence/relative entropy sense) from the true distribution. If this is true, our test statistic will follow (asymptotically) a Gaussian distribution with mean zero. 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 $-\infty$. In case of differentiation between Pareto and the one-component Lognormal, the likelihood ratio test is the "Universally Most Powerful Unbiased"-test (UMPU) (Malevergne et al., 2011; del Castillo and Puig, 1999).

The Aikaike Information criteria augments the log-likelihood information for the number of parameters and is defined as $AIC = 2np - 2ln(L)$ with np the number of parameters and $ln(L)$ the log likelihood.

The Bayesian Information criteria then, is defined as $BIC = npln(F) - 2ln(L)$. Differentiation between distributions relies on relative distance of the BIC's: $\Delta BIC = BIC_1 - BIC_2$. The value of ΔBIC implies strong evidence in favor of distribution 1 if $B > 10$, moderate evidence if $6 < B \leq 10$ and weak evidence if $2 < B \leq 6$ (Kass and Raftery, 1995).

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, Pareto, ... 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 Portuguese firms that has been used already by a.o. (Fernandes and Ferreira, 2017; Fonseca et al., 2018; Dias et al., 2016; Carreira and Teixeira, 2016; Bastos et al., 2018).[10] It contains data both on firm-level sales and number of employees. Moreover, each firm has a unique identification number that allows us to cross-reference this dataset with a dataset on international trade.

A possible limitation of our dataset is that its subject is a very special economy: Portugal. The firm size distributions of Portugal were earlier the object of study by Cabral and Mata (2003), relying 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 suggest, using international comparisons of the firm size distribution for the manufacturing sector, that Portugal is not very different from other countries in this respect. Compared to Cabral and Mata (2003), our dataset is not limited to the manufacturing sector.

We reduce our dataset disregarding individual companies[11] and calculate our main dependant variable from the positive domestic sales of 299,935 Portuguese firms in 2006. Relying on domestic rather than total sales corrects for the impact of international trade on the firm size distribution (di Giovanni et al., 2011). We demean domestic sales (r) and take it to the power of 0.33, which corresponds to the value of the elasticity of the substitution parameter of four ($\sigma = 4$), to obtain a measure of productivity (ω) à la Melitz (2003): $\omega \sim Cr^{\frac{\sigma}{\sigma-1}}$ up to a constant (C) (Neary et al., 2015; Nigai, 2017).[12] We acknow-

[10] A comparison between SCIE and the OECD SBDS database proves the full coverage of firms in our dataset for the Portuguese economy (see Table A1).

[11] Disregarding individual companies renders our dataset more comparable with earlier datasets used to evaluate firm size distributions such as the ORBIS database used by Nigai (2017). Also, it results in a unimodal density distribution of firms sales as to a bimodal distribution if all firms are considered (See Figure A1) that would favour our proposed FMM.

[12] See Appendix B for more information on the relation between firm-level productivity and sales based on

ledge that measuring productivity this way does not allow for non-constant markups (Melitz and Ottaviano, 2008) or measurement error and heterogeneity in input use (Dewitte et al., 2018). It is, however, in line with the current literature on firm size distributions, allowing straightforward comparison of the results while our methodology can easily be transferred to alternative measurements of productivity.

5 Results

We fit the considered distributions to Portuguese productivity in the year 2006. We focus initially on mixtures of Lognormal distributions, as it allows for a straightforward comparison with Pareto-tailed Lognormal distributions and limit our discussion to up to 4-component FMMs, proving to be sufficient for our main message. We show that our results can be extended to other distributions and are robust to sample selection and outliers. They also stand tall in an external validation on city size distributions.

Table 1 displays the distributions fits both to the complete and tails of Portuguese firm-level productivity in 2006. We start with focusing on the complete distribution. As the distributions in the table are ordered according to their Negative Log-likelihood ranking (R_{NLL}), we immediately observe single parametric distributions residing in the bottom. This demonstrates the need of combining distributions to adequately capture heterogeneity in productivity. The Pareto distribution, for instance, is a really bad fit to the distribution with a Log-Likelihood, a KS-test statistic and cumulative KS-deviation up to 100 times bigger than the best-fitting mixture of Lognormals. Concerning the multi-component distributions, we see an increasing fit with an increasing number of parameters. These Log-likelihood rankings are not reversed when parameter correction ($R_{AIC,BIC}$) is applied. The mixtures of Lognormals behave significantly better than the best-fitting Pareto-Lognormal combination (designated in grey) according to the BIC criteria and has a KS-test statistic and cumulative

the Melitz (2003)-model. Notice that the unknown constant C is irrelevant for GFT calculations (Bee and Schiavo, 2017).

KS-deviation that is approximately 50% smaller. Notice also that this paper introduces no less than four types of distributions to the firm size literature that provide a better fit than the currently assumed best-fitting Right-Pareto-tailed Lognormal distribution (Nigai, 2017).

[TABLE 1]

Allowing for heterogeneity in distributions clearly provides a better fit to the complete distribution, but what about the tails? This is mostly interesting from the Pareto point of view, which is often claimed to be the best fit to the tails of size distributions. To investigate the performance, we fitted the (Inverse) Pareto to the (left) right tail of the distribution using the methods described in the methodology section (Section 3). We recovered the best-fitting truncation point for the (Inverse) Pareto distribution. We reduced our dataset according to these truncation parameters and fitted truncated mixtures of Lognormals to both tails of the distribution. This methodology puts the Pareto distribution twice in the advantage. First, it is free from a parametric specification for the bulk of the distribution and second, the truncation parameter is chosen in function of the best-fitting (Inverse) Pareto distribution. Nevertheless, it seems that mixtures of Lognormals are more adequate to approximate the tails of the distribution. Moreover, being able to find a better fit by increasing the number of parameters rejects the hypothesis of the CDF following a straight line on a log-log scale, as discussed in Section 1.

Figure 2 provides more insight into the numerical results of Table 1. It plots both the absolute KS-deviation (left) and cumulative absolute KS-deviation (right) between the parametric and empirical distribution. The figure shows the large errors related to the one-component Lognormal distribution. Augmenting the Lognormal distribution with a Pareto right-tail as Nigai (2017) improves the fit only marginally. While it does provide a slightly better fit in the right tail of the distribution, this comes at the cost of a worse fit to the left-tail of the distribution and an almost as-bad fit to the bulk of the distribution as the one-component Lognormal distribution. The best-fitting Pareto-tailed Lognormal, the Double-Pareto Lognormal, does a better job at fitting the distribution. However, it clearly lags behind in comparison with the best-fitting FMM which only displays marginal errors both in the bulk and the tails of the data.

[FIGURE 2]

5.1 Extension to other distributions

The superior fit of FMMs is not limited to the Lognormal distribution, and can be expanded to the Burr and/or Weibull distribution. Table A2 displays the test 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. Overall, most of these distributions are not able to match the performance of the Lognormal. Only the Burr distribution provides an even better fit to the data. Compared to Pareto-Lognormal combinations, we find that also mixtures of Weibull distributions are able to provide a better fit.

5.2 Robustness

We check whether our results are not the result of sample selection. We restricted our dataset to the manufacturing sector only (see Table A4) and found the performance of FMMs to improve relative to Pareto-Lognormal distributions. This result underwrites the asymptotic nature of Pareto-Lognormal distributions.

We also examined whether our results are not due to outliers in the tails of the distribution. Therefore, we reduced our dataset by discarding the first and last 1,000 observations of the dataset. Results in Table A5 confirm the superiority of FMMs.

We validate our results externally 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 famous discussion between Eeckhout (2009) and Levy (2009).[13] Table A3 showcases the test results, demonstrating that the city size distribution

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

is neither Lognormal, nor Pareto or Pareto-tailed Lognormal. It is best approximated by a FMM of Burr distributions. These results summarise the city size literature up till now and are in line with the findings of Kwong and Nadarajah (2019).

6 Gains From Trade implications[14]

Finding the best parametric approximation of firm size distributions is of critical importance to calculating Gains From Trade. A vast literature finds that alternative parametric approximations result in significant differences in GFT (Head et al., 2014; Bas et al., 2017; Fernandes et al., 2018; Bee and Schiavo, 2017; Nigai, 2017) and alters the channels through which trade operates (Fernandes et al., 2018; Bas et al., 2017; Arkolakis et al., 2012; Melitz and Redding, 2015). In this section, we integrate the fitted distributions fits from the previous section into the Melitz (2003)-framework. This allows us to perform a GFT exercise along the lines of Melitz and Redding (2015) and Bee and Schiavo (2017) and investigate the importance of providing a good fit to the productivity distribution.

Our setup is a a two-country symmetric heterogenous firms model, similar to Melitz and Redding (2015) (see Appendix B for more details). We parametrize our model to represent two symmetric countries i and j, and choose labor in one country as the numeraire, so that $W_i = W_j = 1$. The elasticity of substitution σ is set to four. We calibrate the structural parameters in order to match stylized facts in our dataset, similar to (Head et al., 2014; Melitz and Redding, 2015; Bee and Schiavo, 2017). The variable trade costs $\bar{\tau}_{ij}$ amount to 1.76 when calibrated to match the average fraction of exports in firms sales (15.54%)[15]. We choose fixed entry costs f^e to deliver a entry rate of 0.5 (Head et al., 2014) and fixed exporting costs $\bar{f}_{ij}$ to match the export participation rate (7.32%). Domestic fixed costs are set to one $f_{ii} = 1$, as the quantity of relevance to calculate GFT is the ratio between domestic and exporting costs f_{ij}/f_{ii}. This calibration setup ensures equal export participation at the

[14]See Appendix B for a full workout of the model.

[15] $\frac{\tau^{1-\sigma_s}}{1+\tau^{1-\sigma_s}} = \frac{\sum_{j\neq i} X_{ijs}}{X_{is}}$, assuming a CES demand system and symmetry across countries.

initial situation, disregarding the distribution (Bee and Schiavo, 2017).

At last, we need to parametrize the model according to the parametric approximation of the heterogeneity distribution. We calculated the necessary analytical expressions linked to all distribution considered in the empirical analysis, details of which can be found in Appendix B. We found the Melitz (2003) to generalize easily to FMMs due to its additivity of individual densities. We recover the distributional parameters from our empirical analysis when possible. As is well known for the case of the Pareto distribution, the Melitz (2003)-model imposes parameter restrictions for heavy-tailed distributions to allow for a unique finite equilibrium. This issue is usually resolved by fitting the Pareto distribution to only a small, left-truncated part the dataset (Bee and Schiavo, 2017; Bas et al., 2017; di Giovanni and Levchenko, 2013). In our case, the parameter restrictions applies to multiple heavy-tailed distributions, so that we decided on an alternative solution. We fit the distributions to the data under the additional parameter restrictions necessary to keep the model solution finite.[16] The influence of introducing these parameter restrictions on the distributional fit can be observed in Appendix Table A6.

We compute the GFT by evaluating the change in welfare, measured by total utility U, away from its calibrated state due to an exogenous increase in variable/fixed trade costs τ/f to τ'/f':

$$\text{GFT} = \frac{U - U'}{U'} \text{ x } 100. \tag{8}$$

We let variable trade costs vary between 3 and 1 relative to the calibrated variable trade costs $\bar{\tau}_{ij}$ and investigate the influence of doubling and halving fixed trade costs relative to their calibrate state $\bar{f}_{ijs}$ on the procentual welfare loss/gain in Table 2. The distributions are ordered according to their Negative Log-likelihood. As such, we can interpret the first-ranked four-component Burr distribution as providing the closest values to the truth. Results are

[16]See Appendix Table B1, B2 and B3 for the necessary parameter restrictions for each distribution.

similar to earlier reported calibration exercises (see f.i. Bee and Schiavo (2017)). Both for changes in fixed and variable trade costs, we observe large changes in procentual gains/losses depending on the distributional fit. A change of the variable trade costs from 3 to the calibrated variable trade costs of $\bar{\tau}_{ij=}$1.76, for instance, implies a welfare gain of 3.69% with a 4-component Lognormal distribution, while it only results in a 1.77% welfare gain under the 1-component Lognormal distribution. These differences can not be linked directly to the overall fit to the distribution, though. For instance, the Pareto distribution is able to approximate GFT from the 4-component Burr distribution quite closely with a welfare gain of 4.03%, despite not being able to adequately match the complete firm size distribution. This is in line with earlier results reported by f.i. Fernandes and Ferreira (2017). Most likely a good fit to the right tail of the distribution dominates the GFT outcome. However, even though though this is the case for GFT, providing a good fit to the complete firm size distribution allows for better approximations to f.i. the trade elasticity (Bas et al., 2017). Increasing the number of components of FMMs brings the GFT closer to the true value, seemingly converging with an increasing number of parameters.

[TABLE 2]

7 Conclusion

This paper provides evidence that heterogeneity in productivity can be captured most adequately by Finite Mixture Models. Building on the existence of discrete subpopulations, FMMs generalize existing theories on firm dynamics. Moreover, their flexible, semi-parametric nature results in improved empirical performance compared to currently considered distributions in the firm size literature. A Gains From Trade exercise demonstrates the need to combine multiple distributions to obtain representative welfare impacts.

Throughout our research, we found the firm size literature to be lacking a clear statistical framework that differentiates between a sufficiently large number of distributions over a consistent data range. The road to our conclusion therefore resulted in three main contributions to the literature. First, we introduce multiple new distribution types to the literature and establish a uniform statistical framework to fit and compare all these distributions. Second, we rely on a dataset covering the universe of Portuguese firms in 2006, allowing for the first time to evaluate the complete as well as to focus both on the *left*- and right- tail of the productivity distribution. Third, we provide analytical expressions for all distributions considered and implement them into a quantitative evaluation of a heterogeneous firms model à la Melitz (2003). This allows us to generalize earlier findings of the literature on different distribution types significantly affecting the resulting GFT.

Even though our results provide strong evidence in favour of FMMs, we take no stance on distribution type nor on the mixing parameter defining its underlying discrete subpopulations. It is clear that the two are closely interconnected, and therefore not easily identifiable. Future research will be necessary to take a closer look at the sources of distinctive distributions within the population of firms.

The idea of FMMs opens many new venues for research. For instance, the estimation of firm heterogeneity, productivity, usually relies a first-order Markov process that is identical for the complete population. Concurrently, however, it is recognized that firms can differ in their productivity evolution depending on o.a. firm age Olley and Pakes (1996), exporting behaviour De Loecker (2013), innovation behavior Aw et al. (2011) .... 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, it would provide the opportunity to investigate more specifically the sources of discrete subpopulations within sales.

References

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

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, June). 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.

Bee, M. and S. Schiavo (2017). Powerless: gains from trade when firm productivity is not pareto distributed. *Review of World Economics*.

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

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. and B. Grassi (Forthcoming). Large firm dynamics and the business cycle. *American Economic Review*.

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

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

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

del Castillo, J. and P. Puig (1999). The Best Test of Exponentiality against Singly Truncated Normal Alternatives. *Journal of the American Statistical Association 94*(446), 529–532.

Dewitte, R. (2019). The gravity equation in international trade: a note. mimeo.

Dewitte, R., G. Rayp, and B. Merlevede (2018). The distribution of productivity: Measurement and applications. mimeo.

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.

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.

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

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

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.

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

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

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

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.

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, Mar). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Phys. Rev. E 83*, 036111.

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 G. I. P. Ottaviano (2008, 01). Market Size, Trade, and Productivity. *The Review of Economic Studies 75*(1), 295–316.

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.

Neary, P., M. Mrázová, 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.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

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 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, December). Establishment size dynamics in the aggregate economy. *American Economic Review 97*(5), 1639–1666.

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

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics 9*(1), 60–62.

Figures

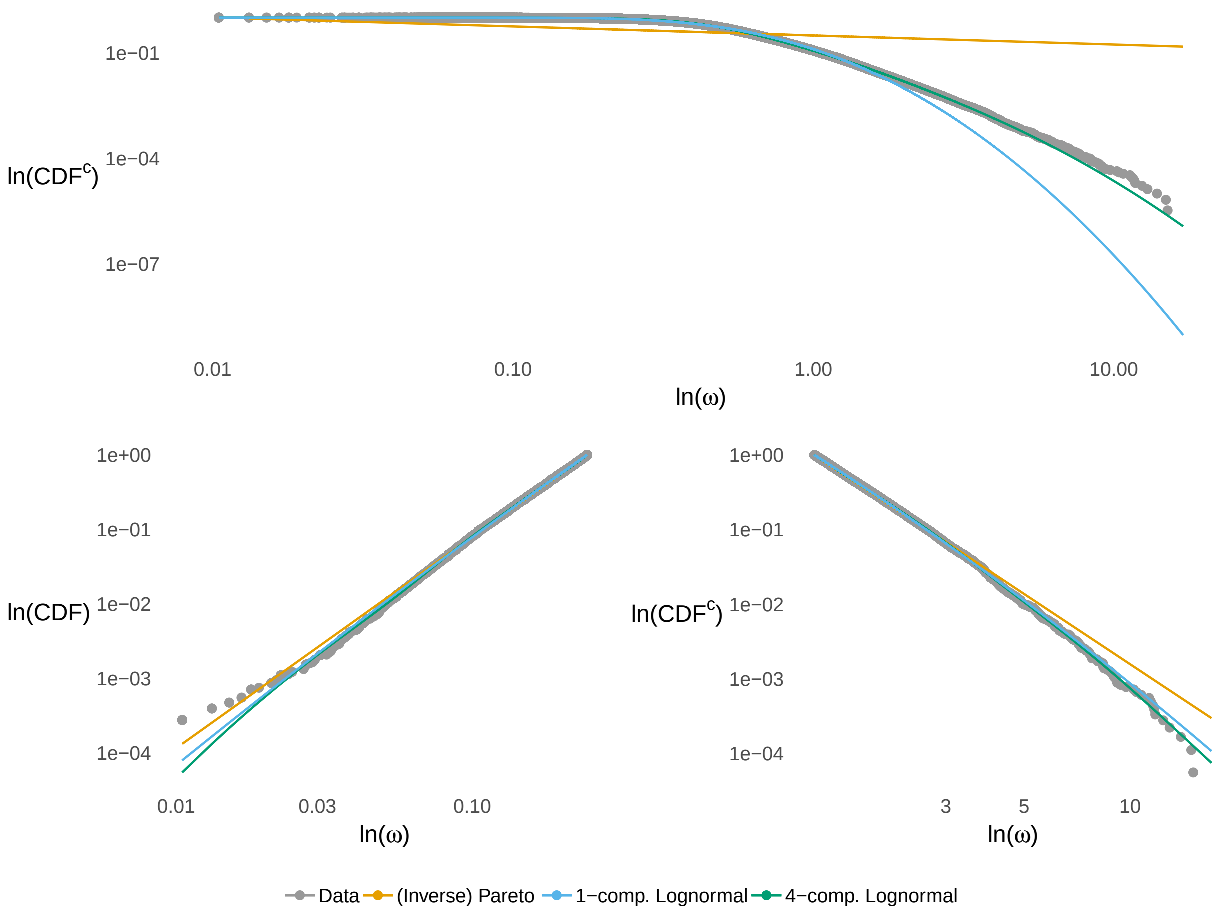

Figure 1: Empirical survival function of Portuguese firm productivity in 2006 (upper panel) on a log-log scale with fitted (Inverse) Pareto and Finite Mixtures of the Lognormal distributions reveal the capacity of FMMs to fit the complete distribution. Focusing on the left tail (lower left panel) right tail (lower right panel) showcases the difficulty to differentiate between distributions fitted to the tail only. **Notes:** Productivity, ω, is measured as domestic sales (relative to the mean) to the power of $1/(\sigma - 1)$ with $\sigma = 4$. (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.

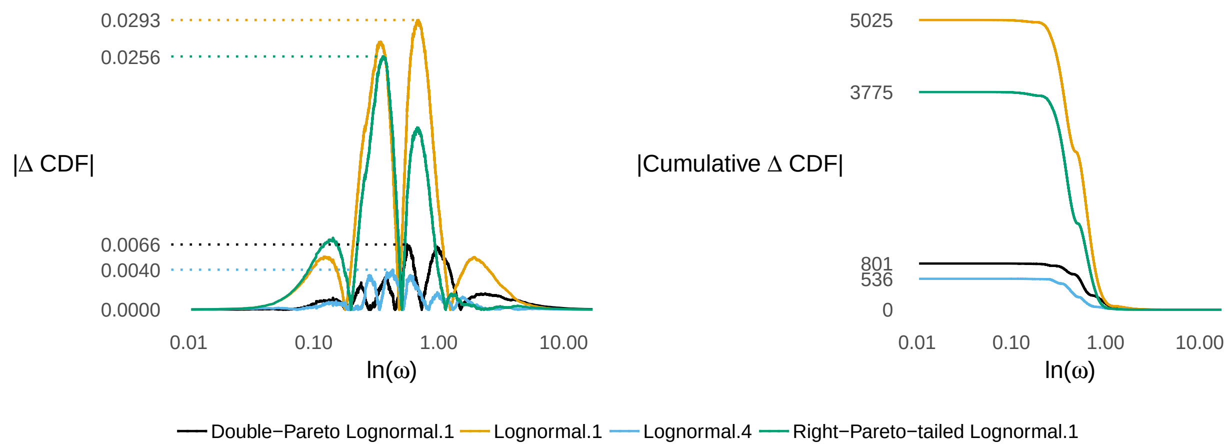

Figure 2: Absolute deviations (left) and cumulative absolute deviations (right) of 1-, and 4-component Lognormal, Double-Pareto Lognormal and Right-Pareto-tailed Lognormal distributions from the empirical distribution (left) reveal the superior performance of multi-component FMMs over the complete range of the data – Portugal 2006.

Tables

Table 1: Distribution fits to Portuguese productivity in 2006 demonstrate the superior fit of FMMs to the complete as well as to the tails of the distribution.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike[1]	R_{AIC}	R^2_{BIC}
			Complete data (F= 299,935)				
Lognormal	4	11	0.0040	536	-58325	1	1^{+++}
Lognormal	3	8	0.0046	622	-58352	2	2^{+++}
Double-Pareto Lognormal	1	4	0.0066	801	-58538	3	3
Lognormal	2	5	0.0072	914	-58588	4	4
Pareto-tailed Lognormal	1	4	0.0081	1014	-58737	5	5
Left-Pareto-tailed Lognormal	1	3	0.0302	4264	-61769	9	9
Right-Pareto-tailed Lognormal	1	3	0.0256	3776	-62247	10	10
Left-Pareto Lognormal	1	3	0.0323	4906	-62909	11	11
Right-Pareto Lognormal	1	3	0.0282	4376	-62939	12	12
Lognormal	1	2	0.0293	5025	-63595	13	13
Pareto	1	1	0.4834	68179	-507194	25	25
			Left tail (F=25,549)				
Lognormal	4	11	0.0049	28	46135***	1	4
Lognormal	3	8	0.0063	42	46132***	2	3
Lognormal	2	5	0.0075	66	46122	4	2
Lognormal	1	2	0.0102	98	46121*	3	1
Inverse Pareto	1	-	0.0079	93	46118	-	-
			Right tail (F=18,217)				
Lognormal	4	11	0.0042	23	-7556***	4	4
Lognormal	2	5	0.0048	29	-7558***	2	2
Lognormal	3	8	0.0043	25	-7558***	3	3
Lognormal	1	2	0.0069	42	-7559***	1	1
Pareto	1	-	0.0086	78	-7571	-	-

Notes: The FMM are fitted to the tails with truncation parameters determined by the best-fitting (Inverse) Pareto distribution. The distributions indicated in grey are the best-fitting (paretly) Pareto distributions to the Log-likelihood.

[1] ***,** ,* indicates significance at the 0.01,0.05,0.1 level respectively for the p-value of the Log-likelihood ratio test **in the tails** between FMM and the (Inverse) Pareto distribution.

[2] +++ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) **for the complete dataset** is greater than 10: $\Delta BIC > 10$, ++ indicates that $6 < \Delta\ BIC \leq 10$ and + indicates that $2 < \Delta\ BIC \leq 6$.

Table 2: Procentual welfare loss/gain from changing variable and fixed trade costs relative to their calibrated state.

			Δ Variable trade costs		Δ fixed trade costs	
Distribution	Comp.	Par.	$3 \rightarrow \bar{\tau}_{ij}$	$\bar{\tau}_{ij} \rightarrow 1$	$2\bar{f}_{ij} \rightarrow \bar{f}_{ij}$	$\bar{f}_{ij} \rightarrow 0.5\bar{f}_{ij}$
Burr	4	15	3.69	17.67	0.25	0.25
Burr	3	11	4.35	18.81	0.05	0.05
Lognormal	4	11	2.84	16.13	0.53	0.52
Lognormal	3	8	2.82	16.09	0.54	0.52
Burr	1	3	4.48	19.03	0.01	0.01
Double-Pareto Lognormal	1	4	4.48	19.02	0.01	0.01
Lognormal	2	5	3.07	16.45	0.49	0.45
Pareto-tailed Lognormal	1	4	2.28	15.29	0.62	0.67
Weibull	3	8	2.77	16.09	0.53	0.54
Gamma	4	11	3.15	16.56	0.48	0.42
Gamma	3	8	3.17	16.58	0.47	0.42
Gamma	2	5	2.78	15.77	0.63	0.54
Left-Pareto-tailed Lognormal	1	3	1.64	14.23	0.67	0.80
Right-Pareto-tailed Lognormal	1	3	2.91	16.38	0.49	0.51
Left-Pareto Lognormal	1	3	1.61	14.18	0.67	0.80
Right-Pareto Lognormal	1	3	2.52	15.64	0.55	0.59
Lognormal	1	2	1.77	14.44	0.68	0.78
Weibull	2	5	2.02	15.10	0.89	0.87
Gamma	1	2	1.01	13.48	0.62	0.92
Weibull	1	2	0.88	13.45	0.61	0.96
Exponential	1	1	2.05	15.32	0.82	0.85
Exponential	4	7	2.05	15.32	0.82	0.85
Exponential	3	5	2.05	15.32	0.82	0.85
Exponential	2	3	2.05	15.32	0.82	0.85
Burr	2	7	3.52	17.38	0.30	0.30
Pareto	1	1	4.03	18.42	0.16	0.16

Notes: Distributions are ordered according to their NLL (See Table A6). $\bar{\tau}_{ij}$ and $\bar{f}_{ij}$ represent the variable and fixed trade costs calibrated to match the export intensity and participation rate of the data respectively.

A Additional Figures and table

A.1 Figures

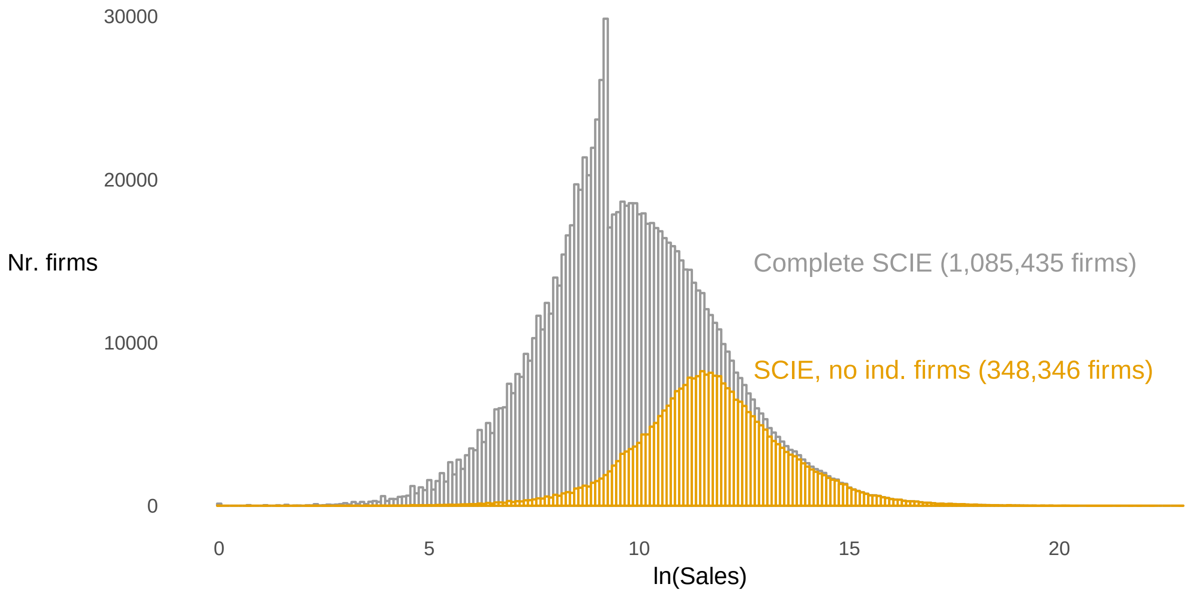

Figure A1: A density comparison of the SCIE dataset with and without individual companies shows that (i) individual companies represent a dominant part of the number of firms and (ii) including individual companies results in a bimodal distribution of firm sales.

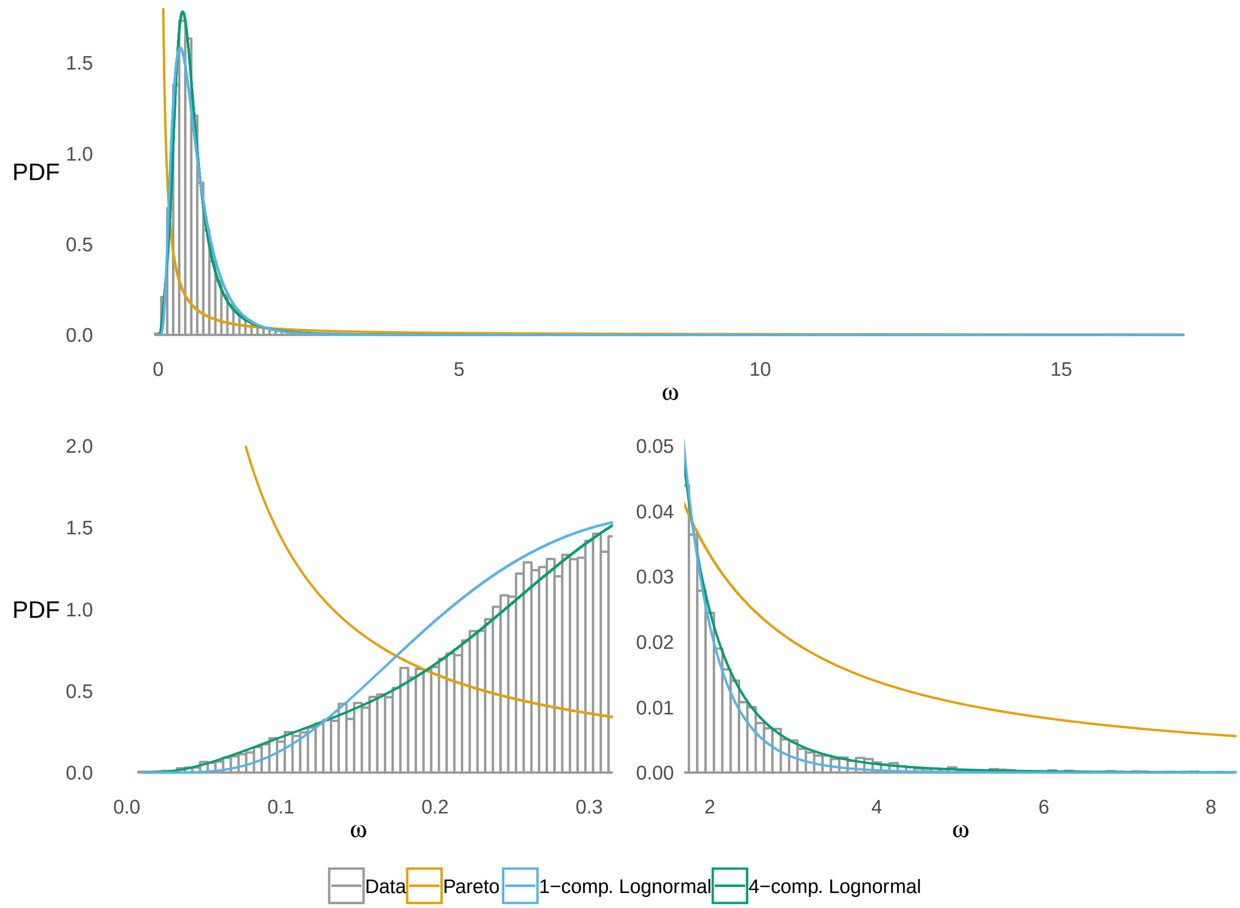

Figure A2: Empirical probability density function of Portuguese firm productivity in 2006 (upper panel) with fitted Pareto and Finite Mixtures of the Lognormal distribution reveal the capacity of FMMs to fit the complete distribution. Focusing on the left- (lower left panel) and right tail (lower right panel) showcases the difficulty of the Pareto and one-component Lognormal distributions to match the bulk and tails of the distribution simultaneously. **Notes:** Productivity, ω, is measured as domestic sales (relative to the mean) to the power of $1/(\sigma-1)$ with $\sigma = 4$. Distributions are fitted using maximum likelihood methods (cf. infra) to the complete dataset.

A.2 Tables

Table A1: Coverage ratio of SCIE vs OECD SDBS database

	Number of Enterprises						Total Employment						Turnover					
Industry	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, as well as absent industries are due to missing information from SBDS even though the data is available in our SCIE database.

Table A2: Distribution fits to the Portuguese firms sales distribution in 2006 demonstrate the dependence of FMMs superior performance on its underlying density.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike	R_{AIC}	R^1_{BIC}
Lognormal	4	11	0.0040	536	-58325	1	1^{+++}
Lognormal	3	8	0.0046	622	-58352	2	2^{+++}
Burr	1	3	0.0073	949	-58517	3	3^{+++}
Double-Pareto Lognormal	1	4	0.0066	801	-58538	4	4
Lognormal	2	5	0.0072	914	-58588	5	5
Pareto-tailed Lognormal	1	4	0.0081	1014	-58737	6	6
Weibull	3	8	0.0070	969	-59114	7	7
Gamma	4	11	0.0145	2314	-59486	8	8
Gamma	3	8	0.0151	2395	-59544	9	9
Gamma	2	5	0.0181	2804	-60183	10	10
Left-Pareto-tailed Lognormal	1	3	0.0302	4264	-61769	11	11
Right-Pareto-tailed Lognormal	1	3	0.0256	3776	-62247	12	12
Left-Pareto Lognormal	1	3	0.0323	4906	-62909	13	13
Right-Pareto Lognormal	1	3	0.0282	4376	-62939	14	14
Lognormal	1	2	0.0293	5025	-63595	15	15
Weibull	2	5	0.0211	3227	-64526	16	16
Fréchet	2	5	0.0693	11661	-74493	17	17
Fréchet	4	11	0.0666	11216	-74796	18	18
Fréchet	3	8	0.0673	12596	-77240	19	19
Gamma	1	2	0.0678	10582	-77462	20	20
Weibull	1	2	0.0918	16516	-100680	21	21
Fréchet	1	2	0.0891	16720	-103876	22	22
Exponential	1	1	0.2442	33681	-152462	23	23
Exponential	4	7	0.2442	33681	-152462	26	26
Exponential	3	5	0.2442	33681	-152462	25	25
Exponential	2	3	0.2442	33681	-152462	24	24
Burr	2	7	0.3222	45902	-218443	27	27
Pareto	1	1	0.4834	68179	-507194	28	28

Notes: The distribution indicated in grey is the best-fitting combination of Pareto and Lognormal according to BIC. [1] $^{+++}$ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) is greater than 10:$\Delta BIC > 10$, $^{++}$ indicates that $6 < \Delta\ BIC \leq 10$ and $^{+}$ indicates that $2 < \Delta\ BIC \leq 6$.

Table A3: Distribution fits to the U.S. Census 2000 city size distribution confirm the supremacy of FMM.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike	R_{AIC}	R^1_{BIC}
Lognormal	4	11	0.0066	50	-6018	2	3^{+++}
Lognormal	3	8	0.0069	51	-6019	1	1^{+++}
Lognormal	2	5	0.0079	57	-6046	3	2^{+++}
Right-Pareto Lognormal	1	3	0.0132	167	-6085	4	4
Double-Pareto Lognormal	1	4	0.0132	167	-6085	5	5
Right-Pareto-tailed Lognormal	1	3	0.0175	253	-6135	6	6
Pareto-tailed Lognormal	1	4	0.0175	253	-6135	7	7
Lognormal	1	2	0.0189	268	-6152	8	8
Left-Pareto-tailed Lognormal	1	3	0.0189	268	-6152	9	9
Left-Pareto Lognormal	1	3	0.0189	268	-6152	10	10
Fréchet	4	11	0.0161	142	-6183	11	11
Weibull	4	11	0.0136	131	-6189	12	12
Fréchet	3	8	0.0162	174	-6262	13	13
Exponential	4	7	0.0186	136	-6298	14	14
Burr	1	3	0.0218	296	-6370	15	15
Weibull	3	8	0.0174	185	-6393	16	16
Fréchet	2	5	0.0256	320	-6538	17	17
Exponential	3	5	0.0276	276	-6633	18	18
Weibull	2	5	0.0303	323	-6920	19	19
Gamma	4	11	0.0635	442	-7384	20	21
Fréchet	1	2	0.0459	641	-7404	21	20
Gamma	3	8	0.0645	454	-7433	22	22
Gamma	2	5	0.0850	723	-8486	23	24
Exponential	2	3	0.0716	841	-8488	24	23
Weibull	1	2	0.0838	1130	-9030	25	25
Gamma	1	2	0.1686	2071	-13893	26	26
Pareto	1	1	0.4176	5059	-31612	27	27
Exponential	1	1	0.3770	5575	-Inf	28	28

Notes: The distribution indicated in grey is the best-fitting combination of Pareto and Lognormal according to BIC. [1] $^{+++}$ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) is greater than 10:$\Delta BIC > 10$, $^{++}$ indicates that $6 < \Delta\ BIC \leq 10$ and $^{+}$ indicates that $2 < \Delta\ BIC \leq 6$

Table A4: Limiting the data to the manufacturing sector reveals an improving fit of FMMs relative to Pareto-Lognormal distributions, underwriting the asymptotic nature of Pareto-Lognormal distributions.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike	R_{AIC}	R^1_{BIC}
Burr	4	15	0.0030	27	-8840	3	5^{+++}
Burr	3	11	0.0029	25	-8842	2	2^{+++}
Burr	2	7	0.0028	39	-8845	1	1^{+++}
Lognormal	4	11	0.0033	33	-8846	4	3^{+++}
Lognormal	3	8	0.0044	56	-8863	5	4^{+++}
Weibull	4	11	0.0056	77	-8875	6	6^{+++}
Weibull	3	8	0.0085	149	-8960	7	8^{++}
Gamma	3	8	0.0132	194	-8963	8	10
Lognormal	2	5	0.0081	123	-8972	9	7^{+++}
Double-Pareto Lognormal	1	4	0.0108	169	-8984	10	9
Pareto-tailed Lognormal	1	4	0.0127	176	-9001	11	11
Gamma	4	11	0.0144	274	-9010	12	13
Burr	1	3	0.0118	253	-9022	13	12
Gamma	2	5	0.0162	331	-9102	14	14
Left-Pareto-tailed Lognormal	1	3	0.0418	806	-9614	15	15
Weibull	2	5	0.0220	432	-9657	16	16
Right-Pareto-tailed Lognormal	1	3	0.0326	592	-9817	17	17
Left-Pareto Lognormal	1	3	0.0439	890	-9841	18	18
Right-Pareto Lognormal	1	3	0.0351	735	-9882	19	19
Lognormal	1	2	0.0396	876	-9988	20	20
Fréchet	4	11	0.0781	1890	-11844	21	22
Fréchet	2	5	0.0963	2037	-11857	22	21
Fréchet	3	8	0.0749	1912	-11873	23	23
Gamma	1	2	0.0790	1664	-11913	24	24
Weibull	1	2	0.0991	2422	-15059	25	25
Fréchet	1	2	0.1003	2613	-16623	26	26
Exponential	1	1	0.2542	4663	-22334	27	27
Exponential	4	7	0.2542	4663	-22334	30	30
Exponential	3	5	0.2542	4663	-22334	29	29
Exponential	2	3	0.2542	4663	-22334	28	28
Pareto	1	1	0.4914	9427	-72781	31	31

Notes: The distribution indicated in grey is the best-fitting combination of Pareto and Lognormal according to BIC. [1] $^{+++}$ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) is greater than 10:$\Delta BIC > 10$, $^{++}$ indicates that $6 < \Delta\ BIC \leq 10$ and $^{+}$ indicates that $2 < \Delta\ BIC \leq 6$

Table A5: Distribution fits to the dataset without possible outliers confirms the superiority of FMMs.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike	R_{AIC}	R^1_{BIC}
Lognormal	3	8	0.0026	285	-47487	1	1^{+++}
Lognormal	4	11	0.0025	243	-47505	2	2^{+++}
Weibull	4	11	0.0041	284	-47640	3	3^{+++}
Weibull	3	8	0.0044	705	-48070	4	4^{+++}
Lognormal	2	5	0.0073	958	-48153	5	5^{+++}
Burr	1	3	0.0080	763	-48702	6	6^{+++}
Double-Pareto Lognormal	3	4	0.0104	1364	-48724	7	7^{+++}
Left-Pareto-tailed Lognormal	2	3	0.0248	3346	-49702	8	8^{+++}
Right-Pareto Lognormal	2	3	0.0202	3254	-49730	9	9
Right-Pareto-tailed Lognormal	2	3	0.0185	3011	-49822	10	10
Pareto-tailed Lognormal	3	4	0.0185	3004	-49829	11	11
Left-Pareto Lognormal	2	3	0.0249	3705	-49893	12	12
Lognormal	1	2	0.0235	3680	-49908	13	13
Weibull	2	5	0.0136	2281	-50912	14	14
Fréchet	2	5	0.0150	2250	-51042	15	15
Fréchet	1	2	0.0768	12960	-76997	16	16
Weibull	1	2	0.0841	14213	-79053	17	17
Pareto	1	1	0.3707	55018	-284851	18	18

Notes: Outliers are identified as the first and last 1,000 observations of the dataset. The distribution indicated in grey is the best-fitting combination of Pareto and Lognormal according to BIC. [1] $^{+++}$ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) is greater than 10:$\Delta BIC > 10$, $^{++}$ indicates that $6 < \Delta\ BIC \leq 10$ and $^{+}$ indicates that $2 < \Delta\ BIC \leq 6$

Table A6: Restricted Distribution fits demonstrate the performance loss for distributions limited by the Melitz-model.

			Goodness of fit		Information Criteria		
Distribution	Comp.	Par.	T_{KS}	$\sum_{f=1}^{F} KS_f$	Loglike	R_{AIC}	R^1_{BIC}
Burr[2]	4	15	0.0030	180	-58225	1	3^{+++}
Burr[2]	3	11	0.0029	248	-58243	3	2^{+++}
Burr[2]	2	7	0.0031	273	-58244	2	1^{+++}
Lognormal	4	11	0.0040	536	-58325	4	4^{+++}
Lognormal	3	8	0.0046	622	-58352	5	5^{+++}
Burr[2]	1	3	0.0091	1055	-58526	6	6^{+++}
Lognormal	2	5	0.0072	914	-58588	7	7^{+++}
Double-Pareto Lognormal[2]	1	4	0.0117	1337	-58660	8	8
Weibull	3	8	0.0070	969	-59114	9	9
Gamma	4	11	0.0145	2314	-59486	10	10
Gamma	3	8	0.0151	2395	-59544	11	11
Gamma	2	5	0.0181	2804	-60183	12	12
Left-Pareto-tailed Lognormal	1	3	0.0302	4264	-61769	13	13
Left-Pareto Lognormal	1	3	0.0323	4906	-62909	14	14
Right-Pareto Lognormal[2]	1	3	0.0282	4376	-62939	15	15
Lognormal	1	2	0.0293	5025	-63595	16	16
Pareto-tailed Lognormal	1	4	0.0522	8530	-64497	17	17
Weibull	2	5	0.0211	3227	-64526	18	18
Right-Pareto-tailed Lognormal[2]	1	3	0.0737	12161	-70060	19	19
Fréchet[2]	2	5	0.0693	11661	-74493	20	20
Fréchet[2]	4	11	0.0666	11216	-74796	21	21
Fréchet[2]	3	8	0.0673	12596	-77240	22	22
Gamma	1	2	0.0678	10582	-77462	23	23
Weibull	1	2	0.0918	16516	-100680	24	24
Fréchet[2]	1	2	0.0891	16720	-103876	25	25
Exponential	1	1	0.2442	33681	-152462	26	26
Exponential	4	7	0.2442	33681	-152462	29	29
Exponential	3	5	0.2442	33681	-152462	28	28
Exponential	2	3	0.2442	33681	-152462	27	27
Pareto	1	1	0.9950	149942	-3078109	30	30

Notes: The distribution indicated in grey is the best-fitting combination of Pareto and Lognormal according to BIC. [1] $^{+++}$ indicates that the difference between BIC of the distribution and the best-scoring Pareto-Lognormal combination (grey) is greater than 10:$\Delta BIC > 10$, $^{++}$ indicates that $6 < \Delta\ BIC \leq 10$ and $^{+}$ indicates that $2 < \Delta\ BIC \leq 6$. , [2]These distributions have been restricted in their parameters in order to provide finite solutions to the Melitz-model (See Appendix B).

Table A7: Overview of all distributions considered

Distribution	**Abbreviation**	**Support**	**Parameters**	**Change in parameters from power transformation** $a\omega^b$
Pareto	P	$[\omega_{min}, \infty[$	k, ω_{min}	$kb, \left(\frac{\omega_{min}}{a}\right)^{\frac{1}{b}}$
Inverse Pareto	IP	$[0, \omega_{max}]$	k, ω_{max}	$kb, \left(\frac{\omega_{max}}{a}\right)^{\frac{1}{b}}$
Lognormal	LN	$[0, \infty[$	μ, Var	$\frac{\mu - lna}{b}, \frac{Var}{b}$
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	$GG\left(bk, b, \left(\frac{s}{a}\right)^{\frac{1}{b}}\right)$
Double-Pareto Lognormal	DPLN	$[0, \infty[$	k_1, μ, Var, k_2	$\frac{k_1}{b}, b\mu + log(a), Var, \frac{k_2}{b}$
Left-Pareto Lognormal	LPLN	$[0, \infty[$	k_1, μ, Var	$\frac{k_1}{b}, b\mu + log(a), Var$
Right-Pareto Lognormal	RPLN	$[0, \infty[$	μ, Var, k_2	$b\mu + log(a), Var, \frac{k_2}{b}$
Pareto-tailed Lognormal	PTLN	$[0, \infty[$	μ, Var, η, τ	$\frac{\mu - lna}{b}, \frac{Var}{b}, \left(\frac{\eta}{a}\right)^{\frac{1}{b}}, \left(\frac{\tau}{a}\right)^{\frac{1}{b}}$
Pareto-left-tail Lognormal	PLTLN	$[0, \infty[$	μ, Var, η	$\frac{\mu - lna}{b}, \frac{Var}{b}, \left(\frac{\eta}{a}\right)^{\frac{1}{b}}$
Pareto-right-tail Lognormal	PRTLN	$[0, \infty[$	μ, Var, τ	$\frac{\mu - lna}{b}, \frac{Var}{b}, \left(\frac{\tau}{a}\right)^{\frac{1}{b}}$

Table A8: Overview of the probability and cumulative density functions of one-component distributions considered.

Distribution	**PDF**	**CDF**
P	$\frac{k\omega_{min}^{k}}{\omega^{k+1}}$	$1-\left(\frac{\omega_{min}}{\omega}\right)^{k}$
IP	$\frac{k\omega_{max}^{k}}{\omega^{-k+1}}$	$1-\left(\frac{\omega_{max}}{\omega}\right)^{-k}$
LN	$\frac{1}{\omega Var\sqrt{2\pi}}e^{-(ln\omega-\mu)^2/2Var^2}$	$\Phi\left(\frac{ln\omega-\mu}{Var}\right)$
W	$\frac{k}{s}\left(\frac{\omega}{s}\right)^{k-1}e^{-\left(\frac{\omega}{s}\right)^{k}}$	$1-e^{-\left(\frac{\omega}{s}\right)^{k}}$
Exp	$\frac{1}{s}e^{-\frac{\omega}{s}}$	$1-e^{-\frac{\omega}{s}}$
B	$\frac{\frac{kc}{s}\left(\frac{\omega}{s}\right)^{c-1}}{\left(1+\left(\frac{\omega}{s}\right)^{c}\right)^{k+1}}$	$1-\frac{1}{\left(1+\left(\frac{\omega}{s}\right)^{c}\right)^{k}}$
F	$\frac{k}{s}\left(\frac{\omega}{s}\right)^{-1-k}e^{-\left(\frac{\omega}{s}\right)^{-k}}$	$e^{-\left(\frac{\omega}{s}\right)^{-k}}$
GG[1]	$\frac{c}{s^k\Gamma(\frac{k}{c})}\omega^{k-1}e^{-\left(\frac{\omega}{s}\right)^{c}}$	$\frac{1}{\Gamma(\frac{k}{c})}\gamma(\frac{k}{c},\left(\frac{\omega}{s}\right)^{c})$
G[1]	$\frac{1}{s^k\Gamma(k)}\omega^{k-1}e^{-\frac{\omega}{s}}$	$\frac{1}{\Gamma(k)}\gamma(k,\frac{\omega}{s})$

Notes: [1] $\Gamma(x)$ stands for the Gamma function, while $\gamma(s,x)$ stands for the lower incomplete Gamma function with upper bound x.

Table A9: Overview of the probability and cumulative density functions of Pareto-Lognormal combinations considered.

Distribution	**PDF**	**CDF**
DPLN[1]	$\frac{k_2 k_1}{k_2 + k_1}\left[\omega^{-k_2-1}e^{k_2\mu+\frac{k_2^2 Var^2}{2}}\Phi\left(\frac{ln\omega-\mu-k_2Var^2}{Var}\right) + \omega^{k_1-1}e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\Phi^c\left(\frac{ln\omega-\mu+k_1Var^2}{Var}\right)\right]$	$\Phi\left(\frac{ln\omega-\mu}{Var}\right) - \frac{1}{k_2+k_1}\left[k_1\omega^{-k_2}e^{k_2\mu+\frac{k_2^2 Var^2}{2}}\Phi\left(\frac{ln\omega-\mu-k_2Var^2}{Var}\right) - k_2\omega^{k_1}e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\Phi^c\left(\frac{ln\omega-\mu+k_1Var^2}{Var}\right)\right]$
LPLN[1]	$k_1\omega^{k_1-1}e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\Phi^c\left(\frac{ln\omega-\mu+k_1Var^2}{Var}\right)$	$\Phi\left(\frac{ln\omega-\mu}{Var}\right) - \omega^{k_1}e^{-k_1\mu+\frac{k_1^2 Var^2}{2}}\Phi^c\left(\frac{ln\omega-\mu+k_1Var^2}{Var}\right)$
RPLN[1]	$k_2\omega^{-k_2-1}e^{k_2\mu+\frac{k_2^2 Var^2}{2}}\Phi\left(\frac{ln\omega-\mu-k_2Var^2}{Var}\right)$	$\Phi\left(\frac{ln\omega-\mu}{Var}\right) - \omega^{-k_2}e^{k_2\mu+\frac{k_2^2 Var^2}{2}}\Phi\left(\frac{ln\omega-\mu-k_2Var^2}{Var}\right)$
PTLN[1,2]	$\begin{cases} cb\left(\eta,\tau\right)\omega^{\beta-1} & \text{if } \omega < \eta \\ \frac{b(\eta,\tau)}{(2\pi)^{1/2}\omega Var}e^{-\frac{(log(\omega)-\mu)^2}{2Var^2}}, & \text{if } \eta \leq \omega \leq \tau \\ ab\left(\eta,\tau\right)\omega^{-\zeta-1}, & \text{if } \tau < \omega \end{cases}$	$\begin{cases} \frac{cb(\eta,\tau)\omega^{\beta}}{\beta} & \text{if } \omega < \eta \\ \ldots + \frac{b(\eta,\tau)}{2}\left(2\Phi\left(\frac{log(q)-\mu}{Var}\right)-1\right) - \frac{b(\eta,\tau)}{2}\left(2\Phi\left(\frac{log(\eta)-\mu}{Var}\right)-1\right), & \text{if } \eta \leq \omega \leq \tau \\ \ldots - ab\left(\eta,\tau\right)\frac{\omega^{-\zeta}-\tau^{-\zeta}}{\zeta} & \text{if } \tau < \omega \end{cases}$
PLTLN[1,2]	$\begin{cases} cb\left(\eta,\infty\right)\omega^{\beta-1} & \text{if } \omega < \eta \\ \frac{b(\eta,\infty)}{(2\pi)^{1/2}\omega Var}e^{-\frac{(log(\omega)-\mu)^2}{2Var^2}}, & \text{if } \eta \leq \omega \end{cases}$	$\begin{cases} \frac{cb(\eta,\infty)\omega^{\beta}}{\beta} & \text{if } \omega < \eta \\ \ldots + \frac{b(\eta,\infty)}{2}\left(2\Phi\left(\frac{log(q)-\mu}{Var}\right)-1\right) - \frac{b(\eta,\infty)}{2}\left(2\Phi\left(\frac{log(\eta)-\mu}{Var}\right)-1\right), & \text{if } \eta \leq \omega \end{cases}$
PRTLN[1,2]	$\begin{cases} \frac{b(0,\tau)}{(2\pi)^{1/2}\omega Var}e^{-\frac{(log(\omega)-\mu)^2}{2Var^2}}, & \text{if } \omega \leq \tau \\ ab\left(0,\tau\right)\omega^{-\zeta-1}, & \text{if } \tau < \omega \end{cases}$	$\begin{cases} \frac{b(0,\tau)}{2}\left(2\Phi\left(\frac{log(q)-\mu}{Var}\right)-1\right) - \frac{b(0,\tau)}{2}\left(2\Phi\left(\frac{log(\eta)-\mu}{Var}\right)-1\right), & \text{if } \omega \leq \tau \\ \ldots - ab\left(0,\tau\right)\frac{\omega^{-\zeta}-\tau^{-\zeta}}{\zeta} & \text{if } \tau < \omega \end{cases}$

Notes: [1] Φ and Φ^c stand for the standard normal and complementary standard normal cdfs. [2] $\beta = -\frac{log(\eta)-\mu}{Var^2}$, $\zeta = \frac{log(\tau)-\mu}{Var^2}$, $a = \frac{\tau^{\zeta+1}}{(2\pi)^{1/2}\tau Var}e^{-\frac{(log(\tau)-\mu)^2}{2Var^2}}$, $c = \frac{\eta^{1-\beta}}{(2\pi)^{1/2}\eta Var}e^{-\frac{(log(\eta)-\mu)^2}{2Var^2}}$, $b\left(\eta,\tau\right) = \left(\frac{\eta}{\beta(2\pi)^{1/2}Var}e^{-\frac{(log(\eta)-\mu)^2}{2Var^2}} + \Phi\left(\frac{\tau-\mu}{Var}\right) - \Phi\left(\frac{\eta-\mu}{Var}\right) + \frac{\tau}{\zeta(2\pi)^{1/2}Var}e^{-\frac{(log(\tau)-\mu)^2}{2Var^2}}\right)^{-1}$

B Heterogeneous firms model

We present a multi-country, multi-sector generalization of the workhorse heterogeneous firm model of Melitz and Redding (2015). We then calibrate this model under to investigate the importance of variations in the parametrization of the productivity distribution.

B.1 Model

Setup Imagine a world with N countries, each country $i \in N$ populated by L_i identical households. Each household supplies inelastically one unit of labor, earning wage W_i. They make their consumptions choices according to a nested Cobb-Douglas-CES utility function:

$$U_i = \prod_{s=1}^{S} \left(\sum_{i=1}^{N} \int_0^{M_{ijs}} q_{ijs} \left(\nu_{is}\right)^{\frac{\sigma_s - 1}{\sigma_s}} \, d\nu_{is} \right)^{\frac{\sigma_s}{\sigma_s - 1} \alpha_{js}},$$

where S denotes the number of industries, M_{ijs} is the number of firms from country i serving market j in industry s, q_{ijs} is the quantity of an industry s variety from country i consumed in country j, α_{is} is the fraction of country j income spend on industry s varieties and $\sigma_s > 1$ the elasticity of substitution between industry s varieties.

Sectoral equilibrium Profit maximization of the firm results in an equilibrium price as a constant markup over marginal costs $p_{ijs} = \frac{\sigma_s}{\sigma_s - 1} \frac{\tau_{ijs} W_i}{\omega}$, resulting in the revenue of a firm from i in j with productivity $\omega \sim g_{is}(\omega)$:

$$r_{ijs}(\omega) = \left(\frac{\sigma_s}{\sigma_s - 1} \frac{\tau_{ijs} W_i}{\omega} \right)^{1-\sigma_s} P_{js}^{\sigma_s - 1} \alpha_s W_j L_j,$$

with P_{js} the CES price index in country j and sector s. Only firms that are productive enough can survive $\omega \geq \omega^*_{iis}$ and/or export $\omega \geq \omega^*_{ijs}$ defined under the zero-profit condition:

$$\sigma_s W_i f_{ijs} = \left(\frac{\sigma_s}{\sigma_s - 1} \frac{\tau_{ijs} W_i}{\omega^*_{ijs}} \right)^{1-\sigma_s} P_{js}^{\sigma_s - 1} \alpha_s W_j L_j.$$

Moreover, free entry implies that in equilibrium, the expected value of entry must be equal to the sunk cost of entry:

$$\sum_{j=1}^{N} f_{ijs} \int_{\omega_{ijs}^*}^{\infty} \left[\left(\frac{\omega}{\omega_{ijs}^*} \right)^{\sigma_s - 1} - 1 \right] g_{is}(\omega) d\omega = \sum_{j=1}^{N} f_{ijs} J(\omega_{ijs}^*) = f_{is}^e.$$

$S * N$ free-entry equations and $S * N$ cut-off conditions determine the $2(S * N)$ cut-off conditions in function of wages, fixed costs and distributional parameters. These cut-offs, then, allow us to determine all firm performance measures (relative to wages).

General equilibrium Imposing an equilibrium on the labour market (total demand = total supply) in each country allows us to obtain an expression for the mass of entering firms that only depends on labor supply:

$$L_{is} = M_{is}^e \sum_{j=1}^{N} \sigma_s f_{ijs} \omega_{ijs}^{1-\sigma_s} \int_{\omega_{ijs}^*}^{\infty} \omega^{\sigma_s - 1} g_{is}(\omega) d\omega = M_{is}^e \sum_{j=1}^{N} \sigma_s f_{ijs} \left(\omega_{ijs}^* \right)^{1-\sigma_s} I(\omega_{ijs}^*).$$

where the industry labour shares are obtained from cost-minimization

$$L_{is} = \alpha_s L_i.$$

An equilibrum on the goods market is obtained from imposing trade balance

$$W_i L_i = \sum_{s=1}^{S} \sum_{j=1}^{N} \frac{X_{jis}}{X_{js}} \alpha_s W_j L_j,$$

where X_{jis} denotes total sales by firms from j in i for sector s

$$X_{ijs} = M_{is}^e \sigma_s W_i f_{ijs} \left(\omega_{ijs}^* \right)^{1-\sigma_s} I(\omega_{ijs}^*).$$

B.2 Parametrization

We parametrize our model to represent two symmetric countries i and j, and choose labor in one country as the numeraire, so that $W_i = W_j = 1$. The elasticity of substitution σ is set to four. We calibrate the structural parameters in order to match stylized facts in our dataset, similar to (Head et al., 2014; Melitz and Redding, 2015; Bee and Schiavo, 2017). The variable trade costs τ are calibrated to match the average fraction of exports in firms sales (0.1554)[17]. We choose fixed entry costs f^e to deliver a entry rate of 0.5 (Head et al., 2014) and fixed exporting costs f_{ij} to match the export participation rate (0.0732). Domestic fixed costs are set to one $f_{ii} = 1$, as the quantity of relevance to calculate GFT is the ratio between domestic and exporting costs f_{ij}/f_{ii}. This calibration setup ensures equal export participation at the initial situation, disregarding the distribution (Bee and Schiavo, 2017).

At last, we need to parametrize two statistics related to the productivity distribution, the distributional parameters of which are recovered from our empirical analysis:

$$I(\omega_{ijs}^*) = \int_{\omega_{ijs}^*}^{\infty} (\omega)^{\sigma_s - 1} g_{is}(\omega) d\omega,$$

and

$$\begin{aligned}
J(\omega_{ijs}^*) &= \int_{\omega_{ijs}^*}^{\infty} \left[\left(\frac{\omega}{\omega_{ijs}^*} \right)^{\sigma_s - 1} - 1 \right] g_{is}(\omega) d\omega \\
&= \left(\omega_{ijs}^*\right)^{1-\sigma_s} I(\omega_{ijs}^*) - \int_{\omega_{ijs}^*}^{\infty} g_{is}(\omega) d\omega \\
&= \left(\omega_{ijs}^*\right)^{1-\sigma_s} I(\omega_{ijs}^*) - \int_{0}^{\infty} g_{is}(\omega) d\omega + \int_{0}^{\omega_{ijs}^*} g_{is}(\omega) d\omega \\
&= \left(\omega_{ijs}^*\right)^{1-\sigma_s} I(\omega_{ijs}^*) - 1 + G_{is}\left(\omega_{ijs}^*\right).
\end{aligned}$$

The first statistic $I(\omega_{ijs}^*)$ is essentially the $\sigma_s - 1$ moment of the productivity distribution truncated at ω_{ijs}. The analytical expressions of this statistic for each distribution are gathered in tables B1, B2 and B3. The calculation of these statistics can be found in the ensuing subsection. The second statistic, then, can easily be calculated with knowledge of the first statistic and the CDF of the productivity distribution. Notice also these statistics easily generalize to FMM of the available distributions due to its additivity and applying the sum rule in integration:

$$I(\omega_{ijs}^*) = \int_{\omega_{ijs}^*}^{\infty} (\omega)^{\sigma_s - 1} \sum_{k=1}^{K} m_{is}^k(\omega) d\omega = \sum_{k=1}^{K} \int_{\omega_{ijs}^*}^{\infty} (\omega)^{\sigma_s - 1} m_{is}^k(\omega) d\omega = \sum_{k=1}^{K} I^k(\omega_{ijs}^*),$$

[17] $\frac{\tau^{1-\sigma_s}}{1+\tau^{1-\sigma_s}} = \frac{\sum_{j \neq i} X_{ijs}}{X_{is}}$, assuming a CES demand system and symmetry across countries.

and

$$J(\omega^*_{ijs}) = \left(\omega^*_{ijs}\right)^{1-\sigma_s} I(\omega^*_{ijs}) - 1 + G_{is}\left(\omega^*_{ijs}\right) = \left(\omega^*_{ijs}\right)^{1-\sigma_s} \sum_{k=1}^{K} I^k(\omega^*_{ijs}) - 1 + \sum_{k=1}^{K} M^k_{is}\left(\omega^*_{ijs}\right).$$

Table B1: Expression of the statistic $I\left(\omega_{ijs}\right)$ for the one-component distributions considered.

Distribution	$I(\omega^*_{ijs})$	**Additional parameter restrictions**[1]
P	$\frac{k\omega_{min}^k}{\sigma_s-1-k}-\left(\omega^*_{ijs}\right)^{\sigma_s-1-k}$	$k>\sigma_s-1$
LN	$e^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[1-\Phi\left(\frac{ln\omega^*_{ijs}-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)\right]$	-
W[3]	$s^{\sigma_s-1}\Gamma\left(\frac{\sigma_s-1}{k}+1,\left(\frac{\omega_{ijs}}{s}\right)^k\right)$	-
Exp[3]	$s^{\sigma_s-1}\Gamma\left(\sigma_s+1,\frac{\omega_{ijs}}{s}\right)$	-
B[2]	$s^{\sigma_s-1}k\left[\boldsymbol{B}\left(\frac{\sigma_s-1}{c}+1,k-\frac{\sigma_s-1}{c}\right)-\boldsymbol{B}\left(\frac{\left(\frac{\omega_{ijs}}{s}\right)^c}{1+\left(\frac{\omega_{ijs}}{s}\right)^c};\frac{\sigma_s-1}{c}+1,k-\frac{\sigma_s-1}{c}\right)\right]$	$c>\sigma_s-1, kc>\sigma_s-1$
F[3]	$s^{\sigma_s-1}\left[1-\Gamma\left(1-\frac{\sigma_s-1}{k},\left(\frac{\omega_{ijs}}{s}\right)^{-k}\right)\right]$	$k>\sigma_s-1$
GG[3]	$\frac{s^{\sigma_s-1}}{\Gamma\left(\frac{k}{c}\right)}\Gamma\left(\frac{\sigma_s-1+k}{c},\left(\frac{\omega_{ijs}}{s}\right)^c\right)$	-
G[3]	$\frac{s^{\sigma_s-1}}{\Gamma(k)}\Gamma\left(\sigma_s-1+k,\frac{\omega_{ijs}}{s}\right)$	-

Notes: [1]Additional parameter restrictions represent parameter restrictions needed to keep the statistic, and therefore the model solutions, finite.[2]$\boldsymbol{B}(a,b)$ stands for the beta function, while $\boldsymbol{B}(x,a,b)$ stands for the lower incomplete beta function with upper bound x. [3]$\Gamma(x)$ stands for the Gamma function, while $\Gamma(s,x)$ stands for the upper incomplete Gamma function with lower bound x.

Table B2: Expression of the statistic $I\left(\omega_{ijs}\right)$ for the Pareto-Lognormal combinations considered.

Distribution	$I(\omega^*_{ijs})$	**Additional parameter restrictions**[1]
DPLN	$-\frac{k_2k_1}{k_2+k_1}e^{k_2\mu+\frac{k_2^2Var^2}{2}}\frac{\omega_{ijs^*}^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{ln\omega^*_{ijs}-\mu-k_2Var^2}{Var}\right)$ $-\frac{k_2k_1}{k_2+k_1}\frac{1}{\sigma_s-1-k_2}e^{\frac{(\sigma_s-1)^2Var^2+2\mu(\sigma_s-1)}{2}}\Phi^c\left(\frac{ln\omega^*_{ijs}-(\sigma_s-1)Var^2-\mu}{Var}\right)$ $-\frac{k_2k_1}{k_2+k_1}e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\frac{\omega_{ijs^*}^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{ln\omega^*_{ijs}-\mu+k_1Var^2}{Var}\right)$ $+\frac{k_2k_1}{k_2+k_1}\frac{1}{\sigma_s-1+k_1}e^{\frac{(\sigma_s-1)^2Var^2+2\mu(\sigma_s-1)}{2}}\Phi^c\left(\frac{ln\omega^*_{ijs}-(\sigma_s-1)Var^2-\mu}{Var}\right)$	$k_2>\sigma_s-1$
LPLN	$-k_1e^{-k_1\mu+\frac{k_1^2Var^2}{2}}\frac{\omega_{ijs^*}^{\sigma_s+k_1-1}}{\sigma_s+k_1-1}\Phi^c\left(\frac{ln\omega^*_{ijs}-\mu+k_1Var^2}{Var}\right)$ $+\frac{k_1}{\sigma_s-1+k_1}e^{\frac{(\sigma_s-1)^2Var^2+2\mu(\sigma_s-1)}{2}}\Phi^c\left(\frac{ln\omega^*_{ijs}-(\sigma_s-1)Var^2+\mu}{Var}\right)$	-
RPLN	$-k_2e^{k_2\mu+\frac{k_2^2Var^2}{2}}\frac{\omega_{ijs^*}^{\sigma_s-k_2-1}}{\sigma_s-k_2-1}\Phi\left(\frac{ln\omega^*_{ijs}-\mu-k_2Var^2}{Var}\right)$ $-\frac{k_2}{\sigma_s-1-k_2}e^{\frac{(\sigma_s-1)^2Var^2+2\mu(\sigma_s-1)}{2}}\Phi^c\left(\frac{ln\omega^*_{ijs}-(\sigma_s-1)Var^2+\mu}{Var}\right)$	$k_2>\sigma_s-1$

Notes: [1]Additional parameter restrictions represent parameter restrictions needed to keep the statistic, and therefore the model solutions, finite.

Table B3: Expression of the statistic $I\left(\omega_{ijs}\right)$ for the Pareto-Lognormal combinations considered.

Distribution	$I(\omega^*_{ijs})$	**Additional parameter restrictions**[1]
PTLN[2,3]	$cb\left(\eta,\tau\right)\frac{\eta^{\sigma_s-1+\beta}-\left(\omega^*_{ijs}\right)^{\sigma_s-1+\beta}}{\sigma_s-1+\beta} + b\left(\eta,\tau\right)e^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[\Phi\left(\frac{ln\tau-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)-\Phi\left(\frac{ln\eta-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)\right] + ab\left(\eta,\tau\right)\frac{-\left(\tau\right)^{\sigma_s-1-\zeta}}{\sigma_s-1-\zeta}$	$\zeta>\sigma_s-1$
PLTLN[2,3]	$cb\left(\eta,\infty\right)\frac{\eta^{\sigma_s-1+\beta}-\left(\omega^*_{ijs}\right)^{\sigma_s-1+\beta}}{\sigma_s-1+\beta} + b\left(\eta,\infty\right)e^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[1-\Phi\left(\frac{ln\eta-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)\right]$	-
PRTLN[2,3]	$b\left(0,\tau\right)e^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[\Phi\left(\frac{ln\tau-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)-\Phi\left(\frac{ln\omega^*_{ijs}-\left((\sigma_s-1)Var^2+\mu\right)}{Var}\right)\right] + ab\left(0,\tau\right)\frac{-\left(\tau\right)^{\sigma_s-1-\zeta}}{\sigma_s-1-\zeta}$	$\zeta>\sigma_s-1$

Notes: [1]Additional parameter restrictions represent parameter restrictions needed to keep the statistic, and therefore the model solutions, finite. [2]Φ and Φ^c stand for the standard normal and complementary standard normal cdfs. [3]$\beta=-\frac{log(\eta)-\mu}{Var^2}$, $\zeta=\frac{log(\tau)-\mu}{Var^2}$, $a=\frac{\tau^{\zeta+1}}{(2\pi)^{1/2}\tau Var}e^{-\frac{(log(\tau)-\mu)^2}{2Var^2}}$, $c=\frac{\eta^{1-\beta}}{(2\pi)^{1/2}\eta Var}e^{-\frac{(log(\eta)-\mu)^2}{2Var^2}}$, $b\left(\eta,\tau\right)=\left(\frac{\eta}{\beta(2\pi)^{1/2}Var}e^{-\frac{(log(\eta)-\mu)^2}{2Var^2}}+\Phi\left(\frac{\tau-\mu}{Var}\right)-\Phi\left(\frac{\eta-\mu}{Var}\right)+\frac{\tau}{\zeta(2\pi)^{1/2}Var}e^{-\frac{(log(\tau)-\mu)^2}{2Var^2}}\right)^{-1}$

B.2.1 Pareto

$$I(\omega_{ijs}^*) = \int_{\omega_{ijs}^*}^{\infty} (\omega)^{\sigma_s - 1} \frac{k\omega_{min}^k}{\omega^{k+1}} d\omega$$
$$= k\omega_{min}^k \frac{-\left(\omega_{ijs}^*\right)^{\sigma_s - 1 - k}}{\sigma_s - 1 - k} \qquad \text{if } k > \sigma_s - 1$$

B.2.2 LogNormal

$$I(\omega_{ijs}^*) = \int_{\omega_{ijs}^*}^{\infty} (\omega)^{\sigma_s - 1} \frac{1}{\omega Var\sqrt{2\pi}} e^{-(ln\omega - \mu)^2/2Var^2} d\omega$$
$$= \int_{\omega_{ijs}^*}^{\infty} e^{(\sigma_s - 1)ln\omega} \frac{1}{\omega Var\sqrt{2\pi}} e^{-(ln\omega - \mu)^2/2Var^2} d\omega$$

Note that

$$(\sigma_s - 1)ln\omega - (ln\omega - \mu)^2 / 2Var^2 = \frac{2Var^2(\sigma_s - 1)ln\omega - (ln\omega)^2 - \mu^2 + 2\mu ln\omega}{2Var^2}$$
$$= -\frac{(ln\omega)^2 - 2(Var^2(\sigma_s - 1) + \mu)ln\omega + \left((Var^2(\sigma_s - 1) + \mu)\right)^2 - (Var^2(\sigma_s - 1) + \mu)^2 + \mu^2}{2Var^2}$$
$$= -\frac{\left[ln\omega - \left(Var^2(\sigma_s - 1) + \mu\right)\right]^2}{2Var^2} + \frac{(Var^2(\sigma_s - 1) + \mu)^2 - \mu^2}{2Var^2}$$
$$= -\frac{\left[ln\omega - \left(Var^2(\sigma_s - 1) + \mu\right)\right]^2}{2Var^2} + \frac{(\sigma_s - 1)\left((\sigma_s - 1)Var^2 + 2\mu\right)}{2}$$

so that

$$I(\omega_{ijs}^*) = e^{\frac{(\sigma_s - 1)\left((\sigma_s - 1)Var^2 + 2\mu\right)}{2}} \int_{\omega_{ijs}^*}^{\infty} \frac{1}{\omega Var\sqrt{2\pi}} e^{-\frac{\left[ln\omega - \left(Var^2(\sigma_s - 1) + \mu\right)\right]^2}{2Var^2}} d\omega$$
$$\text{let } z = \frac{ln\omega - ((\sigma_s - 1)Var^2 + \mu)}{Var}, \; dz = \frac{d\omega}{\omega Var}$$
$$= e^{\frac{(\sigma_s - 1)\left((\sigma_s - 1)Var^2 + 2\mu\right)}{2}} \int_{\frac{ln\omega_{ijs}^* - ((\sigma_s - 1)Var^2 + \mu)}{Var}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} d\omega$$
$$= e^{\frac{(\sigma_s - 1)\left((\sigma_s - 1)Var^2 + 2\mu\right)}{2}} \left[1 - \Phi\left(\frac{ln\omega_{ijs}^* - ((\sigma_s - 1)Var^2 + \mu)}{Var}\right)\right]$$

B.2.3 Weibull[18]

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s-1} \frac{k}{s} \left(\frac{\omega}{s}\right)^{k-1} e^{-\left(\frac{\omega}{s}\right)^k} d\omega \\
&\quad \text{let } z = \left(\frac{\omega}{s}\right)^k, dz = \frac{k}{s}\left(\frac{\omega}{s}\right)^{k-1} d\omega \\
&\quad \text{s.t. } \omega = s z^{\frac{1}{k}} \\
&= \int_{\left(\frac{\omega_{ijs}}{s}\right)^k}^{\infty} s^{\sigma_s-1} z^{\frac{\sigma_s-1}{k}} e^{-z} dz \\
&= s^{\sigma_s-1} \int_{\left(\frac{\omega_{ijs}}{s}\right)^k}^{\infty} z^{\left(\frac{\sigma_s-1}{k}+1\right)-1} e^{-z} dz \\
&= s^{\sigma_s-1} \Gamma\left(\frac{\sigma_s-1}{k}+1, \left(\frac{\omega_{ijs}}{s}\right)^k\right)
\end{aligned}
$$

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

B.2.4 Fréchet

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s-1} \frac{k}{s} \left(\frac{\omega}{s}\right)^{-1-k} e^{-\left(\frac{\omega}{s}\right)^{-k}} d\omega \\
&\quad \text{let } z = \left(\frac{\omega}{s}\right)^{-k}, dz = \frac{-k}{s}\left(\frac{\omega}{s}\right)^{-k-1} d\omega \\
&\quad \text{s.t. } \omega = s z^{-\frac{1}{k}} \\
&= -\int_{\left(\frac{\omega_{ijs}}{s}\right)^{-k}}^{0} s^{\sigma_s-1} z^{-\frac{\sigma_s-1}{k}} e^{-z} dz, \qquad \text{if } k > 0 \\
&= \int_0^{\left(\frac{\omega_{ijs}}{s}\right)^{-k}} s^{\sigma_s-1} z^{-\frac{\sigma_s-1}{k}} e^{-z} dz \\
&= s^{\sigma_s-1} \int_0^{\left(\frac{\omega_{ijs}}{s}\right)^{-k}} z^{1-\left(\frac{\sigma_s-1}{k}\right)-1} e^{-z} dz \\
&= s^{\sigma_s-1} \left[1 - \Gamma\left(1 - \frac{\sigma_s-1}{k}, \left(\frac{\omega_{ijs}}{s}\right)^{-k}\right)\right] \qquad \text{if } k > \sigma_s - 1
\end{aligned}
$$

[18] You can easily obtain the truncated moment of the exponential distribution setting k=1.

B.2.5 Burr

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s - 1} \frac{\frac{kc}{s}\left(\frac{\omega}{s}\right)^{c-1}}{\left(1+\left(\frac{\omega}{s}\right)^c\right)^{k+1}} d\omega \\
&\qquad \text{let } z = \left(\frac{\omega}{s}\right)^c, dz = \frac{c}{s}\left(\frac{\omega}{s}\right)^{c-1} d\omega \\
&\qquad \text{s.t. } \omega = s z^{\frac{1}{c}} \\
&= \int_{\left(\frac{\omega^*_{ijs}}{s}\right)^c}^{\infty} s^{\sigma_s - 1} z^{\frac{\sigma_s - 1}{c}} \frac{k}{(1+z)^{k+1}} dz, \qquad \text{if } c > 0 \\
&= s^{\sigma_s - 1} k \int_{\left(\frac{\omega^*_{ijs}}{s}\right)^c}^{\infty} z^{\frac{\sigma_s - 1}{c}} \frac{1}{(1+z)^{k+1}} dz \\
&= s^{\sigma_s - 1} k \int_{\left(\frac{\omega^*_{ijs}}{s}\right)^c}^{\infty} z^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz \\
&= s^{\sigma_s - 1} k \int_{\left(\frac{\omega^*_{ijs}}{s}\right)^c}^{\infty} z^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz \\
&= s^{\sigma_s - 1} k \left[\int_0^{\infty} z^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} \frac{1}{(1+z)^{k+1}} dz - \int_0^{\left(\frac{\omega_{ijs}}{s}\right)^c} z^{\left(\frac{\sigma_s - 1}{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^{\sigma_s - 1} k \left[\int_0^1 \left(\frac{u}{1-u}\right)^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} \frac{1}{\left(1+\frac{u}{1-u}\right)^{k-1}} du - \int_0^{\frac{\left(\frac{\omega_{ijs}}{s}\right)^c}{1+\left(\frac{\omega_{ijs}}{s}\right)^c}} \left(\frac{u}{1-u}\right)^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} \frac{1}{\left(1+\frac{u}{1-u}\right)^{k-1}} du \right. \\
&= s^{\sigma_s - 1} k \left[\int_0^1 u^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} (1-u)^{k-1-\left(\frac{\sigma_s - 1}{c}+1\right)+1} du - \int_0^{\frac{\left(\frac{\omega_{ijs}}{s}\right)^c}{1+\left(\frac{\omega_{ijs}}{s}\right)^c}} \int_0^{\left(\frac{\omega_{ijs}}{s}\right)^c} u^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} (1-u)^{k-1-(\sigma_s} \right. \\
&= s^{\sigma_s - 1} k \left[\int_0^1 u^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} (1-u)^{k-\left(\frac{\sigma_s - 1}{c}+1\right)} du - \int_0^{\frac{\left(\frac{\omega_{ijs}}{s}\right)^c}{1+\left(\frac{\omega_{ijs}}{s}\right)^c}} u^{\left(\frac{\sigma_s - 1}{c}+1\right)-1} (1-u)^{k-\left(\frac{\sigma_s - 1}{c}+1\right)} duz\right] \\
&= s^{\sigma_s - 1} k \left[\boldsymbol{B}\left(\frac{\sigma_s - 1}{c}+1, k - \frac{\sigma_s - 1}{c}\right) - \boldsymbol{B}\left(\frac{\left(\frac{\omega_{ijs}}{s}\right)^c}{1+\left(\frac{\omega_{ijs}}{s}\right)^c}; \frac{\sigma_s - 1}{c}+1, k - \frac{\sigma_s - 1}{c}\right)\right] \\
&\qquad \text{if } c > \sigma_s - 1, kc > \sigma_s - 1
\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.

B.2.6 Generalized Gamma[19]

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s - 1} \frac{c}{s^k \Gamma(\frac{k}{c})} \omega^{k-1} e^{-\left(\frac{\omega}{s}\right)^c} d\omega \\
&\qquad \text{let } z = \left(\frac{\omega}{s}\right)^c, dz = \frac{c}{s}\left(\frac{\omega}{s}\right)^{c-1} d\omega \\
&\qquad \text{s.t. } \omega = s z^{\frac{1}{c}} \\
&= \int_{\left(\frac{\omega_{ijs}}{s}\right)^c}^{\infty} s^{\sigma_s - 1} \frac{z^{\frac{\sigma_s - 1}{c}}}{\Gamma(\frac{k}{c})} \left(\frac{s z^{\frac{1}{c}}}{s}\right)^{(k-1)-(c-1)} e^{-z} dz, \qquad \text{if } c > 0 \\
&= \frac{s^{\sigma_s - 1}}{\Gamma(\frac{k}{c})} \int_{\left(\frac{\omega_{ijs}}{s}\right)^c}^{\infty} z^{\frac{\sigma_s - 1 + k}{c} - 1} e^{-z} dz \\
&= \frac{s^{\sigma_s - 1}}{\Gamma(\frac{k}{c})} \Gamma\left(\frac{\sigma_s - 1 + k}{c}, \left(\frac{\omega_{ijs}}{s}\right)^c\right)
\end{aligned}
$$

B.2.7 Right-Pareto Lognormal

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s - 1} k_2 \omega^{-k_2 - 1} e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right) d\omega \\
&= k_2 e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \int_{\omega^*_{ijs}}^{\infty} \omega^{\sigma_s - k_2 - 2} \Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right) d\omega \\
dv &= \omega^{\sigma_s - k_2 - 2} d\omega, v = \frac{\omega^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \\
u &= \Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right), du = d\Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right) \\
&= k_2 e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \left[\frac{\omega^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right)\right]_{\omega^*_{ijs}}^{\infty} \\
&\qquad - k_2 e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \int_{\omega^*_{ijs}}^{\infty} \frac{\omega^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} d\Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right) \\
&= k_2 e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \left[0 - \frac{\omega_{ijs^*}^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \Phi\left(\frac{ln\omega^*_{ijs} - \mu - k_2 Var^2}{Var}\right)\right] \\
&\qquad - k_2 e^{k_2 \mu + \frac{k_2^2 Var^2}{2}} \int_{\omega^*_{ijs}}^{\infty} \frac{\omega^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \frac{1}{\omega Var \sqrt{2\pi}} e^{-\frac{\left[ln\omega - \mu - k_2 Var^2\right]^2}{2 Var^2}} d\omega
\end{aligned}
$$

[19]You can easily obtain thetruncated moments of the Gamma distribution setting c=1.

The last integral resembles the truncated moment condition of the Lognormal distribution solved earlier with moment $((\sigma_s - 1) - k_2)$ and mean $(\mu + k_2 Var^2)$ so that

$$
\begin{aligned}
I(\omega^*_{ijs}) = & -k_2 e^{k_2\mu + \frac{k_2^2 Var^2}{2}} \frac{\omega_{ijs^*}^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \Phi\left(\frac{ln\omega^*_{ijs} - \mu - k_2 Var^2}{Var}\right) \\
& - \frac{k_2 e^{k_2\mu + \frac{k_2^2 Var^2}{2}}}{\sigma_s - 1 - k_2} e^{\frac{((\sigma_s-1)-k_2)\left(((\sigma_s-1)-k_2)Var^2 + 2(\mu + k_2 Var^2)\right)}{2}} \left[1 - \Phi\left(\frac{ln\omega^*_{ijs} - (((\sigma_s - 1) - k_2)Var^2 - (\mu
\end{aligned}
$$

Note that

$$
e^{k_2\mu + \frac{k_2^2 Var^2}{2} + \frac{((\sigma_s-1)-k_2)\left(((\sigma_s-1)-k_2)Var^2 + 2(\mu + k_2 Var^2)\right)}{2}}
$$

$$
e^{\frac{2k_2\mu + k_2^2 Var^2 + ((\sigma_s-1)-k_2)\left[(\sigma_s-1)Var^2 + 2\mu + k_2 Var^2\right]}{2}}
$$

$$
e^{\frac{2k_2\mu + k_2^2 Var^2 + (\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1) + k_2(\sigma_s-1)Var^2 - k_2(\sigma_s-1)Var^2 - 2\mu k_2 + k_2^2 Var^2}{2}}
$$

$$
e^{\frac{(\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1)}{2}}
$$

so that we get

$$
\begin{aligned}
I(\omega^*_{ijs}) = & -k_2 e^{k_2\mu + \frac{k_2^2 Var^2}{2}} \frac{\omega_{ijs^*}^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1} \Phi\left(\frac{ln\omega^*_{ijs} - \mu - k_2 Var^2}{Var}\right) \\
& - \frac{k_2}{\sigma_s - 1 - k_2} e^{\frac{(\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1)}{2}} \Phi^c\left(\frac{ln\omega^*_{ijs} - (\sigma_s - 1)Var^2 - \mu}{Var}\right)
\end{aligned}
$$

B.2.8 Left-Pareto Lognormal

$$
\begin{aligned}
I(\omega^*_{ijs}) &= \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s - 1} \omega^{k_1 - 1} e^{-k_1\mu + \frac{k_1^2 Var^2}{2}} \Phi^c\left(\frac{ln\omega - \mu + k_1 Var^2}{Var}\right) d\omega \\
&= k_1 e^{k_2\mu + \frac{k_2^2 Var^2}{2}} \left[\left(\frac{-(\omega^*_{ijs})^{\sigma_s - k_2 - 1}}{\sigma_s - k_2 - 1}\right) - e^{-k_1\mu + \frac{k_1^2 Var^2}{2}} \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s - 2 + k_1} \Phi\left(\frac{ln\omega - \mu + k_1 Var^2}{Var}\right) d\omega\right]
\end{aligned}
$$

$$I(\omega^*_{ijs}) = -k_1 e^{-k_1\mu + \frac{k_1^2 Var^2}{2}} \frac{\omega_{ijs^*}^{\sigma_s+k_1-1}}{\sigma_s+k_1-1} \Phi^c\left(\frac{ln\omega^*_{ijs} - \mu + k_1 Var^2}{Var}\right)$$
$$+ \frac{k_1}{\sigma_s-1+k_1} e^{\frac{(\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1)}{2}} \Phi^c\left(\frac{ln\omega^*_{ijs} - (\sigma_s-1)Var^2 + \mu}{Var}\right)$$

B.2.9 Double-Pareto Lognormal

$$I(\omega^*_{ijs}) = \frac{k_2 k_1}{k_2+k_1} \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s-1} \omega^{-k_2-1} e^{k_2\mu + \frac{k_2^2 Var^2}{2}} \Phi\left(\frac{ln\omega - \mu - k_2 Var^2}{Var}\right) d\omega$$
$$+ \frac{k_2 k_1}{k_2+k_1} \int_{\omega^*_{ijs}}^{\infty} (\omega)^{\sigma_s-1} \omega^{k_1-1} e^{-k_1\mu + \frac{k_1^2 Var^2}{2}} \Phi^c\left(\frac{ln\omega - \mu + k_1 Var^2}{Var}\right) d\omega$$
$$= -\frac{k_2 k_1}{k_2+k_1} e^{k_2\mu + \frac{k_2^2 Var^2}{2}} \frac{\omega_{ijs^*}^{\sigma_s-k_2-1}}{\sigma_s-k_2-1} \Phi\left(\frac{ln\omega^*_{ijs} - \mu - k_2 Var^2}{Var}\right)$$
$$- \frac{k_2 k_1}{k_2+k_1} \frac{1}{\sigma_s-1-k_2} e^{\frac{(\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1)}{2}} \Phi^c\left(\frac{ln\omega^*_{ijs} - (\sigma_s-1)Var^2 - \mu}{Var}\right)$$
$$- \frac{k_2 k_1}{k_2+k_1} e^{-k_1\mu + \frac{k_1^2 Var^2}{2}} \frac{\omega_{ijs^*}^{\sigma_s+k_1-1}}{\sigma_s+k_1-1} \Phi\left(\frac{ln\omega^*_{ijs} - \mu + k_1 Var^2}{Var}\right)$$
$$- \frac{k_2 k_1}{k_2+k_1} \frac{1}{\sigma_s-1+k_1} e^{\frac{(\sigma_s-1)^2 Var^2 + 2\mu(\sigma_s-1)}{2}} \Phi^c\left(\frac{ln\omega^*_{ijs} - (\sigma_s-1)Var^2 - \mu}{Var}\right)$$

B.2.10 Pareto-Tailed Lognormal distribution[20]

We start with the case where $\omega_{ijs} < \eta$, such that:

$$I(\omega^*_{ijs}) = \int_{\omega^*_{ijs}}^{\eta} (\omega)^{\sigma_s-1} cb\omega^{\beta-1} d\omega + \int_{\eta}^{\tau} (\omega)^{\sigma_s-1} \frac{b}{(2\pi)^{1/2}\omega Var} e^{-\frac{(log(\omega)-\mu)^2}{2Var^2}} d\omega + \int_{\tau}^{\infty} (\omega)^{\sigma_s-1} ab\omega^{-\zeta-1} d\omega.$$

We notice that each individual truncated moment condition has already been worked out in previous subsections (for the Pareto and Lognormal distributions). We can therefore easily deduce the analytical solution

[20] The Pareto-Right- and Left-Tailed Lognormal distribution follow straightforwardly.

$$
\begin{aligned}
I(\omega^*_{ijs}) = \\
& cb\frac{\eta^{\sigma_s-1+\beta} - \left(\omega^*_{ijs}\right)^{\sigma_s-1+\beta}}{\sigma_s - 1 + \beta} \\
& + be^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[\Phi\left(\frac{ln\tau - ((\sigma_s-1)Var^2+\mu)}{Var}\right) - \Phi\left(\frac{ln\eta - ((\sigma_s-1)Var^2+\mu)}{Var}\right)\right] \\
& + ab\frac{-\left(\tau\right)^{\sigma_s-1-\zeta}}{\sigma_s-1-\zeta} \qquad \text{if } \zeta > \sigma_s - 1.
\end{aligned}
$$

If $\eta \leq \omega^*_{ijs} \leq \tau$, we get

$$
\begin{aligned}
I(\omega^*_{ijs}) = \\
& be^{\frac{(\sigma_s-1)\left((\sigma_s-1)Var^2+2\mu\right)}{2}}\left[\Phi\left(\frac{ln\tau - ((\sigma_s-1)Var^2+\mu)}{Var}\right) - \Phi\left(\frac{ln\omega^*_{ijs} - ((\sigma_s-1)Var^2+\mu)}{Var}\right)\right] \\
& + ab\frac{-\left(\tau\right)^{\sigma_s-1-\zeta}}{\sigma_s-1-\zeta} \qquad \text{if } \zeta > \sigma_s - 1,
\end{aligned}
$$

while if $\omega^*_{ijs} > \tau$, we have

$$
I(\omega^*_{ijs}) = ab\frac{-\left(\omega^*_{ijs}\right)^{\sigma_s-1-\zeta}}{\sigma_s-1-\zeta} \qquad \text{if } \zeta > \sigma_s - 1.
$$