2015-19

Working paper

POWERLESS: GAINS FROM TRADE WHEN FIRM PRODUCTIVITY IS NOT PARETO DISTRIBUTED

Marco Bee
University of Trento

Stefano Schiavo
OFCE – SciencesPo Paris and University of Trento

September 2015

Powerless: Gains from trade when firm productivity is not Pareto distributed

Marco Bee* Stefano Schiavo†

September 7, 2015

Abstract

Most trade models featuring heterogeneous firms assume a Pareto productivity distribution, on the basis that it provides a reasonable representation of the data and because of its analytical tractability. However, recent work shows that the characteristics of the productivity distribution crucially affect the estimated gains from trade. This paper thoroughly compares the gains from trade obtained under different productivity distributions: we find that both the magnitude of the welfare gains and the relative importance of the fixed versus variable trade costs change significantly. Relying blindly on a single distribution is therefore dangerous when performing welfare analysis.

Keywords: lognormal, Pareto, Weibull, international trade, welfare, firm heterogeneity

JEL Codes: F10, F12

1 Introduction

This paper investigates what happens to the estimated gains from trade when one departs from the standard assumption of a Pareto productivity

The authors blame each other for any remaining mistake. They nevertheless agree to thank participants at the ETSG 2014 in Munich and the 2014 ISGEP meeting in Stockholm, as well as Mauro Caselli and Thierry Mayer for insightful comments on an earlier draft of the paper.

*Department of Economics and Management, University of Trento; via Inama 5 – 38122 Trento (Italy); e-mail: marco.bee@unitn.it

†School of International Studies and Department of Economics and Management, University of Trento and OFCE-DRIC; via T. Gar 14 – 38122 Trento (Italy); e-mail: stefano.schiavo@unitn.it

distribution, which characterizes most of the existing literature.

The current vintage of international trade models pioneered by Melitz (2003) and Bernard et al. (2003) puts the behavior of firms, rather than countries or industries, at center stage. By recognizing the importance of within-industry heterogeneity among firms, the *new-new* trade theory revolution completes the shift of focus away from nations and toward the firm initiated by Krugman (1979, 1980), who had incorporated differentiated products and increasing returns to scale into trade models. This long trajectory has added a number of useful insights, narrowed the gap between economic models and reality, and moved trade theory closer to business and policymakers by providing microfoundations to aggregate gains from trade (Greenaway and Kneller, 2007; Cernat, 2014). However, the welfare implications of the new-new trade theory have not been much explored until recently, when the issue has started to capture a lot of attention (see, for instance, Arkolakis et al., 2012; di Giovanni and Levchenko, 2013; Head et al., 2014; Melitz and Redding, 2015). The debate has to do both with the *additional* gains associated with heterogeneity (Arkolakis et al., 2012; Melitz and Redding, 2015) with respect to the previous models featuring homogeneous firms, and with the sensitivity of welfare gains to the *degree* (and the shape) of the heterogeneity (di Giovanni and Levchenko, 2013; Head et al., 2014).

The trigger can be traced to a contribution by Arkolakis et al. (2012), who aim at assessing the additional gains from trade (GFT) associated with firm heterogeneity. On top of the standard welfare benefits already present in the new-trade literature (Krugman, 1980), heterogeneity adds an additional source of gains in the form of within-industry reallocation of market shares, forcing low productivity firms to exit and letting more efficient ones expand, thus raising aggregate productivity. Arkolakis et al. (2012) conclude that, in fact, this new insight does not much alter the evaluation of the benefits of trade, as the additional gains brought about by heterogeneity are small. Melitz and Redding (2015) investigate the issue further, and show that the results of Arkolakis et al. (2012) depend on some important restrictions, and are very sensitive to even small departures from the original framework. In particular, the setup of Arkolakis et al. requires a constant trade elasticity: this feature is not robust to, say, moving from an unbounded to a truncated Pareto distribution for productivity, and may thus introduce a potential bias in the computation of the GFT. Using a different approach, Melitz and Redding (2015) find that the differences in aggregate welfare between the two families of models are quantitatively important.[1]

Another interesting take at the issue of the welfare gains from trade is provided by di Giovanni and Levchenko (2013), who investigate the welfare impact of a series of reductions in fixed and variable trade costs under different scenarios. In particular, they show that the degree of firm heterogeneity —defined as the shape parameter of the underlying Pareto distribution of firm productivity— significantly affects both the magnitude and the composition of the gains from trade, as well as the benefits accruing from a reduction in fixed versus variable trade costs. The reason for this is that productivity dispersion influences export participation and thus the importance of changes at the extensive margin of trade. A reduction in trade costs shifts the export productivity threshold and the number of firms that can successfully export: when the upper tail of the productivity distribution is heavy, marginal new exporters are much smaller than firms already exporting and thus have a very limited impact on welfare. Indeed, when comparing a model featuring a Zipf productivity distribution with an economy with lower productivity dispersion (they increase the shape parameter from 1 to 2), di Giovanni and Levchenko (2013) find that the gains from a reduction in fixed costs are an order of magnitude larger, while the impact of a reduction in variable costs is an order of magnitude smaller (compared to Zipf's world).

The question then arises as to whether choosing a different type of productivity distribution might push the argument by di Giovanni and Levchenko (2013) even further. Rephrasing this, we are interested in understanding whether, when it comes to quantifying the welfare gains and potential benefits resulting from a reduction in fixed and/or variable trade

[1] More specifically, the "macro" calibration by Arkolakis et al. (2012) requires the two models to have the same trade elasticity with respect to trade costs, and the same domestic trade share, whereas the "micro" approach taken by Melitz and Redding (2015) only changes the degree of heterogeneity in the models, taking the homogeneous case as a limit (degenerate) case of the more general heterogeneous firm specification. See Melitz and Redding (2015) for more details on the two approaches.

costs, the assumption that productivity follows a Pareto distribution may no longer be an innocuous simplifying assumption, but rather crucially affect the results. The issue is rooted in the old question of the most appropriate distributional assumption to model firm size and productivity.[2]

In the trade literature, most if not all papers assume a Pareto or Zipf productivity distribution. This choice is motivated with two main bases: first, that the Pareto is tractable and very convenient from a modeling point of view, allowing closed-form solutions; second, that it provides a good approximation of the data, at least for US firms. To substantiate this latter claim, reference is made to Axtell (2001), who finds that Zipf's law provides a good representation for the entire distribution of firms in his sample. More recent evidence, however, suggests that the Pareto distribution does not fit very well the whole size/productivity distribution, but only the upper tail (see for instance Combes et al., 2012). And even this is debatable both from a methodological (Virkar and Clauset, 2014) and an empirical point of view (Rossi-Hansberg and Wright, 2007; Bee et al., 2014; Head et al., 2014).[3]

Combes et al. (2012) and Head et al. (2014) provide convincing evidence in favor of the lognormal distribution. Building on this, the latter paper explores what happens to GFT when one abandons the Pareto assumption in favor of lognormality: the authors claim that, depending on the calibration of a few key parameters, the welfare effect can be twice as large as under the Pareto assumption.

In this paper, we contribute to the debate by performing a thorough comparison of the GFT obtained under three alternative productivity distributions: the Pareto, lognormal, and Weibull. The choice of the Weibull distribution is suggested by two main reasons. The first is that the Weibull is a member of the Gamma-type family of size distributions, which is probably, with the Pareto and the lognormal, the most common family of distributions used for size modeling (Kleiber and Kotz, 2003). The second is

[2]In the standard monopolistic competition cum CES preferences that represents the backbone of Melitz-type models firm size and firm productivity are closely related, see Section 2 below.

[3]The main methodological issue has to do with the common practice of binning the data before fitting a distribution. Virkar and Clauset (2014) forcefully show that this is an important source of bias.

that the Weibull shares with the lognormal and Pareto distributions the property of closure under exponentiation described below.

Our findings confirm that the choice of a specific distribution has a strong impact on the estimated welfare effects in terms of the magnitude of GFT, the additional benefits associated with heterogeneity, and the relative importance of reductions in fixed versus variable trade costs. For instance, a 25% reduction in fixed export costs (similar to the projected result following a successful completion of a trade agreement between the US and the EU) would yield negligible GFT under a Pareto distribution, while delivering a significant welfare increase assuming a Weibull distribution. Furthermore, we investigate the sensitivity of the results to key parameters of the model such as the elasticity of substitution, which is notoriously difficult to pin down and plays a crucial role in the estimation of GFT. We find that the welfare effects obtained using a Pareto distribution for firm productivity are especially sensitive to the value of the elasticity of substitution.

We conclude that, when it comes to welfare analysis, the choice of a specific productivity distribution (e.g., the Pareto) cannot be regarded as a mere simplifying assumption, but has important repercussions on the estimated GFT. As a result, we warn against the common practice of relying blindly on a single distribution and urge looking at alternatives to make results more credible and robust. Relative to the existing literature, we not only include an additional distribution, but also perform a detailed sensitivity analysis to understand the degree of variability in the results associated with the choice of a specific distribution over another. Furthermore, we discuss at length the policy dimension of our results, a perspective that the existing literature has not investigated much. In fact, as heterogeneous-firm models are increasingly used to draw policy conclusions, it is important to understand the implications of the key underlying assumptions.

The rest of this paper is organized as follows: the next section presents a brief overview of the theoretical setup we refer to, Section 3 provides detailed information on the calibration of the model parameters, and Section 4 discusses the results from our investigation, comparing GFT under different distributional assumptions. Finally, Section 5 links our findings to the existing literature and concludes.

2 Theoretical background

Melitz and Redding (2015) present a simple two-country model from which they derive a series of general results about the welfare gains of trade liberalization, which are independent of any distributional assumption. The paper assumes two symmetric countries populated by a continuum of heterogeneous firms that incur a sunk entry cost f_e before they can discover their productivity ϕ, which is sampled from a common and invariant distribution $g(\phi)$. Production entails fixed costs (f_d) and a constant marginal cost that depends on productivity. If international trade is allowed, i.e. when countries move from autarky to free trade, then exporting firms face also a fixed export cost (f_x) and an iceberg variable trade cost (τ). All costs are expressed in units of labor, which is the sole factor of production.

Demand is modeled by means of the usual CES preferences giving rise to the standard pricing rule for firms, namely a markup over marginal costs. The zero profit conditions in each market define the productivity cutoffs for serving domestic and foreign consumers: operating profits must cover fixed costs.

Melitz and Redding (2015) compute the gains from trade (GFT) by comparing the welfare in autarky W^A with that under free trade W^T, and show that *irrespective of any assumption* on the productivity distribution, the GFT can be computed as the ratio between the productivity cutoffs for serving the domestic market under the different regimes:

$$\frac{W^T}{W^A} = \frac{\phi_d^T}{\phi_d^A},$$

where ϕ_d^T (ϕ_d^A) is the minimum productivity level required to successfully serve the domestic market in a free trade (autarky) equilibrium.

We exploit this result to investigate what happens to GFT when one assumes different productivity distributions $g(\phi)$, and different parameter values within the same distribution. In particular, we focus on comparing the results obtained assuming a Pareto versus both a lognormal and a Weibull distribution for firm productivity. From a computational point of view, the main difference is that while the Pareto case allows one to get closed-form solutions for the threshold productivity levels, this is not possible in the lognormal and Weibull cases, so that we have to revert to numerical methods; this issue is made clearer in Appendix A, which provides a full derivation of the models in the lognormal and Weibull cases.

As for the homogeneous case, Melitz and Redding (2015) treat it as a special (degenerate) case of the more general model featuring heterogeneous productivity. In particular, they assume that after paying the sunk entry cost f_e firms draw a binary productivity that is either zero or positive ($\bar{\phi}_d$) with exogenous probabilities. The calibration of the homogeneous productivity model is such that the autarky equilibria in the two models are equivalent. This means simultaneously equating the probability of successful entry in the homogeneous and heterogeneous cases ($1-\bar{G}_d = 1-G(\phi_d^A)$), and the average productivity levels ($\bar{\phi}_d = \tilde{\phi}_d^A$, where $\tilde{\phi}_d^A$ is a weighted average of firm productivity in the heterogeneous case).

3 Calibration for quantitative analysis

3.1 Key parameters of the productivity distributions

To evaluate the GFT we need to make assumptions about the distribution of productivity and its key parameters as well as about the parameters that determine the equilibrium, namely the sunk entry costs (f_e), the fixed production and export costs (f_d, f_x), the variable trade costs (τ), and the elasticity of substitution (ε). We employ three different distributions: a Pareto with shape parameter α and scale parameter ϕ_{min}, a lognormal with parameters μ and σ, and a Weibull with scale parameter λ and shape parameter k. For the sake of comparability with the existing literature, we first use values equal or similar to those adopted by Melitz and Redding (2015) and Head et al. (2014), and subsequently investigate the robustness of the results to different parameter values.

The first thing to note is that the standard combination of CES preferences and monopolistic competition yields a closed-form relationship between firm size (as measured by sales) and productivity. In fact, the sales of a firm with productivity ϕ can be expressed as

$$s(\phi) = RP^{\varepsilon-1}p(\phi)^{1-\varepsilon},$$

where R is the total expenditure, P the ideal price index, $p(\phi)$ is the equilibrium price set by a firm of productivity ϕ, and ε is the elasticity of substitution. Hence, sales are a power function of productivity. As noted by Head et al. (2014), this means that, for the Pareto and lognormal distributions, sales follow the same distribution as productivity with appropriate changes to the parameters. The same can be shown to hold for the Weibull distribution. This relationship is exploited in the literature because data on firm sales are more reliable than data on productivity: hence, one can estimate the parameters using sales data, and then derive the relevant parameters for the productivity distributions using the following simple relationships:

$$\alpha = \alpha^{sales} \cdot (\varepsilon - 1) \qquad \text{(Pareto)} \tag{1}$$

$$\sigma = \sigma^{sales}/(\varepsilon - 1) \qquad \text{(lognormal)} \tag{2}$$

$$k = k^{sales} \cdot (\varepsilon - 1) \qquad \text{(Weibull).} \tag{3}$$

Melitz and Redding (2015) set $\alpha = 4.25$ and $\varepsilon = 4$, implying a shape parameter for the distribution of firm size $\alpha^{sales} = 1.42$ roughly halfway from the two cases investigated by di Giovanni and Levchenko (2013), who compare GFT when $\alpha^{sales} = 1$ vs. $\alpha^{sales} = 2$. Other papers, namely di Giovanni and Levchenko (2013) and Head et al. (2014), take the parameter of the Pareto distribution of firm sales as given, setting $\alpha^{sales} = 1.06$ as in Axtell (2001).[4] Head et al. (2014) provide convincing evidence that the lognormal provides a very good fit to their data on export sales by French manufacturing firms (similar results are reported also by Bee et al., 2014): they estimate σ^{sales} by means of a QQ estimator (Kratz and Resnick, 1996) that minimizes the sum of squared distances between the theoretical (i.e., implied by the a given distribution) and empirical quantiles. Looking at the sales of French firms exporting to Belgium (the most popular destination for French exporters), Head et al. (2014) find $\sigma^{sales} = 2.39$. We adopt the same methodology to estimate the shape parameter of a Weibull distribution, obtaining $k^{sales} = 0.554$.[5] Table B1 in Appendix B presents

[4]The value found by Axtell (2001) is by far the most popular point estimate in the literature that assumes a Pareto distribution (which in turn represents the overwhelming majority of the studies).

[5]The parameter μ, as well as the scale parameters of the Weibull and the Pareto distributions, do not enter into the computations of the gains from trade. We follow

the detailed results from the QQ estimation of the shape parameters of the three distributions based on sales of French exporters to Belgium in 2000. We carry out the estimation both on the entire sample and on the subsets corresponding to different upper quantiles of the distribution. The Weibull provides a good fit to the data: the R^2 ranges between 0.945 and 0.995, and is only marginally lower than the corresponding values for the lognormal distribution. Only for the top quantiles (top 1–2%) does the Pareto perform a better job.

From the estimated parameters of the distribution of firm sales one derives the parameters of the productivity distribution conditional on the value of the elasticity of substitution, which therefore plays a very important role. Both Melitz and Redding (2015) and Head et al. (2014) set $\varepsilon = 4$: we also use this value as our benchmark, although the sensitivity of the results to this parameter is one of the main points we address below.

In Table 1 we compare the degree of heterogeneity in labor productivity (sales per employee and value added per employee) observed in the data to the one implied by the different distributions analyzed in the paper. More specifically, based on the estimated (or assumed, as for the Pareto distribution) parameters for the lognormal, Weibull and Pareto distributions, we compute three measures of dispersions: the standard deviation, the log difference between the 75th and the 25th percentiles and that between the 90th and the 10th percentiles.[6] These theoretical benchmarks are then compared to the same measures computed on data for French exporters to Belgium in the year 2000. We see that in all cases (whole sample vs. manufacturing firms only; using sales or value added per employee) the dispersion implied by the Pareto distribution with the parameters normally

Melitz and Redding (2015) and set $\phi_{min} = 1$ for the Pareto, while μ and λ are set at zero and 0.5116 respectively (these two values are consistent with each other).

[6]As pointed out by Malevergne et al. (2011), even though the tail behaviors of the Pareto and lognormal distributions are qualitatively different (with the Pareto being heavier: the Pareto belongs to the Frećhet Maximum Domain of Attraction, while the lognormal is in the Gumbel Maximum Domain of Attraction), when σ^2 is large the lognormal tail becomes essentially indistinguishable from the Pareto one. Thus, when σ^2 increases, both the dispersion and the tail heaviness increase, and the tail becomes more and more similar to the Pareto one. As for the Weibull, its tail gets heavy as k decreases; in particular, it is commonly considered an heavy-tailed distribution when $k < 1$ (Embrechts et al., 1997). The variance gets large as well when k decreases, so that a smaller k implies both a larger dispersion and a fatter tail.

used in the literature is far below the actual dispersion in the data. The measures of productivity heterogeneity computed from sales per employee are well approximated by both the lognormal and especially the Weibull distributions, with the two representing some sort of lower and upper bound respectively. When we focus on value added per employee, the Pareto distribution keeps underestimating significantly the degree of heterogeneity in the data, while the lognormal and the Weibull now overestimate it. In this case the lognormal performs better.

Table 1: Productivity dispersion: actual vs. implied by the Pareto, lognormal and Weibull distributions.

	measure of dispersion		
	st. dev (1)	75–25 (2)	90–10 (3)
empirical evidence: *all firms*			
log(sales per empl)	0.938	1.149	2.219
log(value added per empl)	0.657	0.686	1.412
empirical evidence: *manufacturing only*			
log(sales per empl)	0.743	0.878	1.735
log(value added per empl)	0.585	0.607	1.259
theoretical values implied by distribution			
Pareto ($\alpha = 3.18$)	0.315	0.345	0.691
Pareto ($\alpha = 4.25$)	0.235	0.259	0.517
lognormal	0.797	1.075	2.043
Weibull	1.018	1.248	2.447

Empirical evidence based on French exporters to Belgium in 2000. Columns (2)-(3) report the log difference between the 75th and 25th (90th–10th) percentiles.

Theoretical values are computed based on the appropriate distribution for the logs of a Pareto, lognormal and Weibull random variable, namely an exponential, normal and extreme values distribution.

It is worth noting that such a large degree of productivity heterogeneity does not appear as a peculiar feature of our sample: indeed, similar values are found throughout the literature for a number of different advanced and emerging economies. Syverson (2004) reports an average within-industry inter-quartile range of around 0.66 among US manufacturing firms with an average 90th–10th percentile ratio of 4 to 1 for labor productivity and a smaller ratio of 1.92 to 1 for TFP. Hsieh and Klenow (2009) find much larger gaps, around 5 to 1, when comparing the 90th and the 10th percentiles of the distribution in emerging economies such as India and China. Bartelsman et al. (2013) use harmonized data for five industrial countries and report a weighted average within-industry standard deviation for labor productivity that ranges between 0.53 (France) and 0.71 (Germany). Similar degrees of heterogeneity are reported by Bellone et al. (2014), who compare French and Japanese firms, and by Söllner (2010) for German manufacturers.

The tension between the different distributions, on the one hand, and the empirical evidence, on the other, is confirmed when we look at the degree of tail heaviness in the data. This feature can be measured by the *Obesity index* proposed by Cooke et al. (2014) and again compared to simulated values from the distributions under investigation.[7] The Obesity index approaches 1 from below as the tail gets heavier. With respect to the previous evidence, the Pareto distribution does a better job when it comes to capturing the tail behavior of the data: using the whole sample and measuring productivity as sales per employee we get an almost perfect match between the actual and the simulated values of the obesity index. However, changing the sample or the measure of productivity does affect the result and it is not easy to determine the best performing distribution according to the obesity index.

3.2 Other parameters

For what concerns the remaining parameters, we follow Melitz and Redding (2015) and Head et al. (2014) and calibrate them in order to match some stylized facts about the intensive and the extensive margin of trade in the US economy reported by Bernard et al. (2007). In particular, the fixed costs associated with entry (f_e) and exporting (f_x) are chosen to yield an export participation rate equal to 18%, while the baseline value of variable trade costs (τ) allows them to match the average fraction of exports in firm sales (14%). The latter calibration exploits the relationship $\tau^{1-\varepsilon}/(1+\tau^{1-\varepsilon})$ = export intensity. The wage rate is normalized to 1, as in

[7] Given an ordered sample $X_1 \leq X_2 \leq X_3 \leq X_4$ from some distribution, the obesity index exploits the fact that the probability that $X_4 + X_1 > X_2 + X_3$ is larger for heavy-tailed distributions than for thin-tailed ones (see Cooke et al., 2014).

Table 2: Obesity index for the Pareto, lognormal and Weibull distributions.

Pareto	($\alpha = 3.18$)	0.800
Pareto	($\alpha = 4.25$)	0.789
lognormal		0.745
Weibull		0.634
empirical evidence: *all firms*		
sales per empl		0.802
value added per empl		0.653
empirical evidence: *manufacturing only*		
sales per empl		0.756
value added per empl		0.650

empirical evidence based on French exporters to Belgium in 2000.

Melitz and Redding (2015), while there is no need to specify a value for the domestic and foreign labor force, which is treated as exogenous and does not enter the computation of *relative* welfare. Similarly, we do not need to pick values for the overall price index or total revenues spent.

Head et al. (2014) show that, under lognormality, the magnitude of the GFT is not independent of the value of the sunk entry cost f_e, which in turn determines the share of firms that survive upon entry. This is also true under the assumption of a Weibull productivity distribution or, rather, the independence of GFT with respect to f_e (that is mentioned by Melitz and Redding, 2015 when discussing their choice of $f_e = 1$) is a very peculiar feature of the Pareto distribution and is not generally valid. As a consequence, we calibrate f_e to deliver a very small rate of successful entry (0.0055), as in the working paper version of Melitz and Redding (2013), as well as the preferred value by Head et al. (2014), namely 0.5. We will see below that this change generates very large differences in the estimated GFT. Arguably, the dependence on the entry rate, whose precise value is difficult to calibrate, represents one of the main drawbacks associated with the departure from the Pareto distribution.

4 Gains from trade

This section represents the core of the paper and illustrates the results of a series of quantitative evaluations of GFT under the three distributions under analysis: Pareto, lognormal and Weibull. The section is divided in four parts: in Section 4.1 we look at GFT associated with the move from autarky to free trade, adopting a strategy similar to the one used by Melitz and Redding (2015), while in Section 4.2 we perform some robustness checks. Next, in Section 4.3 we evaluate the impact of a reduction in fixed or variable trade costs from current values to those implied by a successful completion of a EU-US trade deal. Last, in Section 4.4 we quantify the additional benefits associated with firm heterogeneity.

4.1 Main results

We start our analysis by estimating the GFT resulting from a trade liberalization that moves two symmetric countries from autarky to a free trade equilibrium, replicating the same export participation and export intensity seen among US firms (namely 18% and 14% respectively). Table 3 presents the relevant results for different distributions and, in the lognormal and Weibull cases, also for different entry rates.

There are substantial differences between the various distributions or even for different parameterizations of each distribution. For instance, column 2 shows that increasing the value of the shape parameter for the Pareto distribution from 3.18 to 4.25 halves the estimated welfare effect of a trade liberalization.[8] This is consistent with one of the theoretical results in Melitz and Redding (2015), whose Proposition 4 tells us that a smaller shape parameter for the Pareto distribution entails larger welfare gains from opening to trade. This insight appears to carry over to different distributions as well: the lognormal case —characterized by a larger obesity index— always yields higher GFT than the Weibull distribution, although the ranking relative to the Pareto case depends on the rate of firm entry assumed in the calibration. We observe that choosing a different distri-

[8] We use two values for the Pareto shape parameter: $\alpha_1 = 4.25$ (as in Melitz and Redding, 2015), and $\alpha_2 = 3.18$ (resulting from assuming $\alpha^{sales} = 1.06$, i.e., the value suggested by Axtell, 2001, and $\varepsilon = 4$).

Table 3: Gains from trade (in percentage points, relative to autarky) generated by a reduction of variable trade costs: Comparison between heterogeneous and homogeneous case and between different degrees of heterogeneity.

		autarky to $\tau = 1.83$			$GFT(\tau)$: heterog.	
		homog. (1)	hererog. (2)	Δ (3)	range (4)	st.dev (5)
Pareto	($\alpha = 3.18$)	3.62	4.46	0.84	24.0	6.34
Pareto	($\alpha = 4.25$)	0.10	2.25	2.15	20.7	5.53
lognormal	(entry: 0.55%)	0	1.80	1.80	19.8	5.31
lognormal	(entry: 50%)	1.68	3.85	2.17	23.2	5.36
Weibull	(entry: 0.55%)	0	0.34	0.34	18.9	6.09
Weibull	(entry: 50%)	0	1.46	1.46	18.5	5.04

Columns 1–2 report welfare gains of a move from autarky to a trading equlibrium characterized by $\tau = 1.83$ for the homogeneous and heterogeneous models. Values are in percentage points. The additional effect due to heterogeneity is given in Column 3. Column 4 gives the range of welfare gains achieved for values of τ ranging from 1 to 2.5, while column 5 reports the associated standard deviation.

bution for firm productivity (i.e., abandoning the power-law assumption) can have even larger effects on the estimated GFT, with results depending heavily on the assumptions about the entry rate and the associated magnitude of sunk entry costs. Moreover, in all cases, heterogeneity brings about important additional benefits: we will discuss the comparison between the homogeneous and the heterogeneous models in more detail in Section 4.4 below. Note that the different welfare impacts across distributions are not driven by export participation, as the fixed costs of production and export are calibrated in such a way that they yield the same share of exporting firms, namely 18%.

We now replicate the exercise by letting the value of τ vary: in other words, we compare trade liberalizations that move the system form autarky to trade equilibria characterized by different levels of variable trade costs and, as a consequence, by different levels of export participation and export intensity. Figure 1 displays the results for τ ranging from 1 (complete liberalization) to 2.5. This is one of the key quantitative exercises discussed in Head et al. (2014) and Melitz and Redding (2013), and therefore provides us with an interesting starting point.

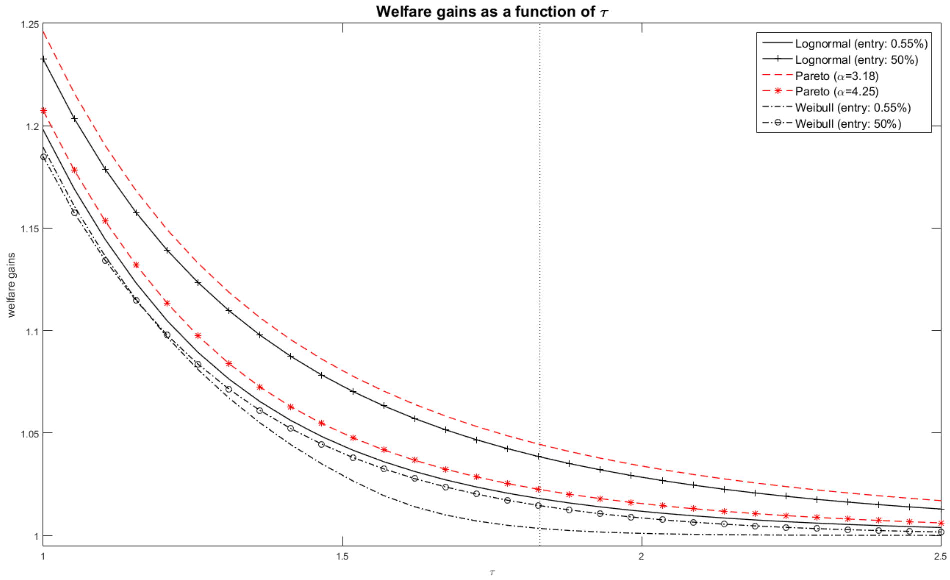

Figure 1: Welfare gains of a trade liberalization as a function of τ: Pareto, lognormal and Weibull cases.

It is worth noting that in the theoretical setting described in Section 2 above, the average export intensity only depends on the elasticity of substitution and the variable trade costs, and is therefore common to the three distributions. A complete elimination of variable trade costs ($\tau = 1$) implies an export intensity of 50%, as the domestic and the foreign markets are identical and there is no longer any wedge between the price paid by domestic and foreign consumers. At the opposite end of the spectrum, for $\tau = 2.5$, firms export (on average) a meager 6% of their production. On the other hand, the distributional assumption has an important effect on the degree of export participation, with the Weibull featuring a particular sensitivity to changes in τ. For instance, with an entry rate of 0.55%, values of $\tau \leq 1.5$ are associated with an export participation of 100%.

Overall, Figure 1 shows both that τ plays an important role in determining the welfare effect of trade, and that the estimated gains from trade are substantially different with different distributions. Hence, the specific distributional assumption underlying the analysis has a strong bearing on the results and sticking to a single productivity distribution may be dangerous in terms of policy implications.

Having considered the impact of a reduction in variable trade costs, we now turn to analyze the effect of a liberalization that affects *fixed* export costs, while keeping the other cost parameters (f_e and τ) fixed at their benchmark values.[9] We hypothesize five different levels of f_x: the benchmark is the value associated with column (3) of Table 4, where cost parameters yield values of GFT equal to those discussed above, associated with an export participation of 18% (and an average export intensity of 14%, which does not vary as τ is held constant). We then look at GFT in case the trade equilibrium features fixed costs that are 25% or 50% lower, 25% higher, or double the benchmark.[10] For each of the productivity distributions and for each level of fixed export costs, we determine the export participation rate, the welfare effect of the homogeneous and heterogeneous models, and the difference from the benchmark case.

As opposed to Figure 1 above, in the case of changes in f_x thicker tails in the distribution of productivity are associated with a lower sensitivity of GFT. Also, fatter tails imply larger GFT from small reductions in f_x (i.e., trading equilibria characterized by high fixed costs, as in columns 4–5). This can be understood by comparing the rates of export participation associated with different levels of fixed export costs under the three distributional assumptions. For instance, we see that the Weibull distribution makes export participation very sensitive to changes in f_x: halving this parameter from its baseline value moves the fraction of exporting firms from 18% to 100%, increasing the welfare effect from a meager 0.34% to 2.01%. On the other hand, for a Pareto distribution with a shape parameter $\alpha = 3.18$, reducing the level of fixed export costs that characterizes the trading equilibrium by one-half moves export participation up to 37.76% and increases the GFT by 0.15% only.

Similar effects (in the opposite direction) are associated with values of f_x higher than the baseline case (see columns 4–5). In fact, when more probability is concentrated in the upper tail of the productivity (and size) distribution, even a small reduction in the threshold of the productivity required to become an exporter affects many firms: in fact, export partic-

[9]The values in Table 4 represent the welfare effects of a move from autarky to a free trade equilibrium characterized by the same level of τ and f_e and various levels of f_x.

[10]For the sake of simplicity Table 4 only reports results obtained using assuming an entry rate equal to 0.55%. Results for entry equal to 50% are available upon request.

Table 4: Welfare gains of trade liberalization as a function of f_x.

			$1/2 f_x$ (1)	$3/4 f_x$ (2)	f_x (3)	$5/4 f_x$ (4)	$2 f_x$ (5)
Pareto	$\alpha = 3.18$	*export participation*	*37.76*	*24.57*	*18.11*	*14.30*	*8.69*
		heterogeneous	4.61	4.50	4.46	4.37	4.26
			[0.15]	[0.06]		[-0.09]	[-0.20]
		homogeneous	4.36	3.97	3.59	3.21	2.1
			[0.77]	[0.38]		[-0.38]	[-1.49]
Pareto	$\alpha = 4.25$	*export participation*	*48.35*	*27.22*	*18.11*	*13.20*	*6.78*
		heterogeneous	2.95	2.51	2.25	2.04	1.69
			[0.70]	[0.26]		[-0.21]	[-0.56]
		homogeneous	2.49	1.26	0.08	0	0
			[2.41]	[1.18]		[-0.08]	[-0.08]
lognormal	$\sigma = 0.797$	*export participation*	*46.13*	*26.93*	*18.11*	*13.19*	*6.60*
		heterogeneous	2.62	2.11	1.80	1.56	1.15
			[0.82]	[0.31]		[-0.24]	[-0.65]
		homogeneous	2.06	0.64	0	0	0
			[2.06]	[0.64]		[–]	[–]
Weibull	$\lambda = 1.662$	*export participation*	*100.00*	*49.86*	*18.11*	*7.35*	*0.73*
		heterogeneous	2.01	0.81	0.34	0.15	0.02
			[1.67]	[0.47]		[-0.19]	[-0.32]
		homogeneous	1.96	0.50	0	0	0
			[1.96]	[0.50]		[–]	[–]

f_x is the calibrated value of fixed export costs that guarantees an export participation rate of 18.11%. f_e and τ do not vary relative to their benchmark values. f_e calibrated to yield an entry rate of 0.55%. Values in square brackets indicate the difference with respect to the benchmark case of column (3).

ipation in column 5 is largest in the case of a Pareto with $k = 3.18$, the distribution with the heaviest tail. On the opposite end of the spectrum, a large reduction in f_x is associated with a significantly larger increase in export participation for distributions with thinner tails, namely the Weibull and the lognormal, where more firms are located in the body of the distribution rather than in the upper tail.

The overall lesson we draw from this first set of results is that there are important differences in the size of GFT among the three distributions we analyze. The ranking in terms of the welfare effect of liberalization does not depend on the distribution, and is related both to the tail heaviness and to the entry rate one uses (apart from the case of a Pareto distribution, where GFT are independent of entry).

4.2 Robustness

The results discussed so far show that the magnitude of the estimated GFT depend significantly on the underlying distributional assumptions concerning firm size and productivity. Among others, there are at least two questions that follow from the above analysis and should be addressed: how sensitive are the results to the specific parameters that govern the tail heaviness (and the productivity dispersion) for each distribution? How sensitive are the results to the value of the elasticity of substitution, which, in addition to entering into the formula for welfare, also determines the relationship between the distribution of size (sales) and productivity illustrated in equations (1)–(3) above? We tackle these issues in Sections 4.2.1 and 4.2.2, which follow.

4.2.1 Sensitivity to tail heaviness

Melitz and Redding (2015) show that, under a Pareto productivity distribution, it is possible to establish a closed-form relationship between the welfare effect of trade liberalization and the degree of size heterogeneity among firms, summarized by the shape parameter. Here we focus on the lognormal and Weibull distributions and investigate what happens when we vary the parameter governing the productivity dispersion (respectively σ and k), adjusting the values of f_x and f_e to maintain constant the export participation at the 18% level.[11]

The results are summarized in Figure 2. The computations were performed keeping all the other parameters at their benchmark values ($\tau = 1.83, f_e = 0.0145, f_x = 0.545, \varepsilon = 4$). We let σ and k vary between 0.2 and 2.5; it is worth remembering that while for the lognormal distribution an increase in σ implies more dispersion, the opposite happens for the Weibull (as for the Pareto). The outcome presented by Melitz and Redding (2015) for the Pareto distribution appears to carry over to the other

[11] We focus on the lognormal and Weibull distributions since we have already experimented with two Pareto distributions featuring different shape parameters in Section 4.1 above.

two distributions under analysis: we see that the estimated GFT increase with the tail thickness while tending to zero when dispersion is low. Under our parametrization, GFT peak at around 5% for both distributions, and display a similar degree of sensitivity to changes in σ and k. If anything, the Weibull distribution appears to respond more quickly to changes in k, whereas the GFT under a lognormal distribution have a horizontal asymptote at (roughly) 5%, corresponding to values of $\sigma \geq 2$.

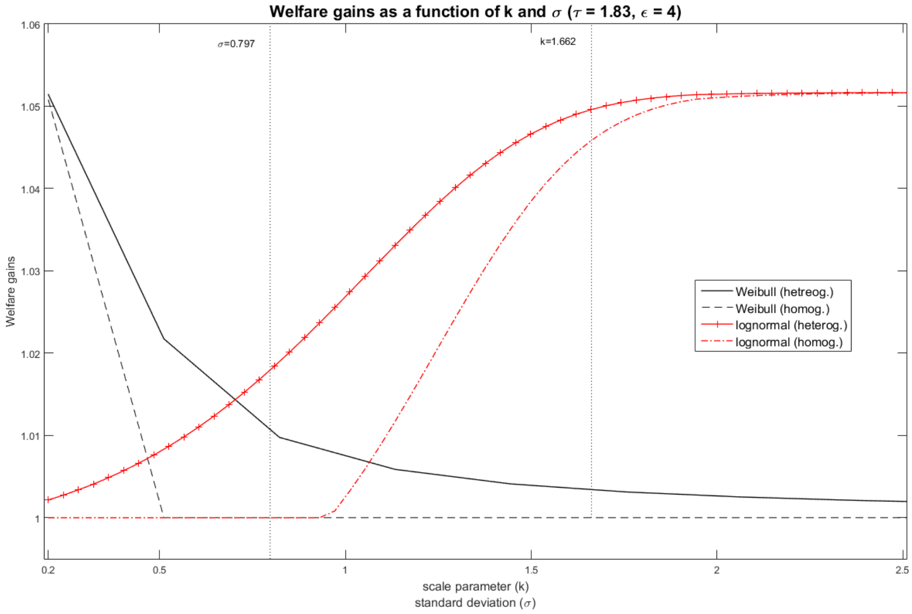

Figure 2: Welfare gains as a function of the degree of tail heaviness in the productivity distributions (σ and k). Lognormal and Weibull cases, homogeneous and heterogeneous models. The elasticity of substitution ϵ is equal to 4. The trade and production costs τ, f_x, f_e are calibrated to yield the baseline values of export intensity (14%) and participation (18%).

4.2.2 Sensitivity to variations in the elasticity of substitution

Another important issue has to do with the calibration of the elasticity of substitution (ε), something which affects the parameters governing the degree of heterogeneity (α, σ and k), as it links the distributions of sales and productivity. Given that ε is notoriously difficult to pin down (e.g. Behrens et al., 2012), the question arises as to how sensitive are GFT to the numerical value of this parameter.[12]

di Giovanni and Levchenko (2013) touch upon this problem by using three different values for ε: 4, 6 (their benchmark), and 8. They report larger welfare gains under a lower ε, which induces thicker tails in the productivity distribution, although they are mainly concerned with the fact that their results are unaffected from a qualitative point of view. Figure 3 compares GFT under the different distributions when we let ε vary and, with it, we modify the degree of firm heterogeneity in the model. The exercise is performed by fixing the relevant parameter of the distribution of firm sales and exploiting the relationships (1)–(3) that link them to the associated productivity distribution. Fixed (f_e and f_x) and variable (τ) costs change as well in order to keep both export participation and average export intensity at their calibrated values.[13] This is done in order to allow a comparison between different trading equilibria characterized by the same outcome in terms of extensive and intensive margins. What changes is the share of exports carried out by each firm, or, in other words, export concentration.[14] It is also interesting to note that in the Pareto case, f_x does not vary with the elasticity of substitution ε, although it does when the productivity distribution follows either a lognormal or a Weibull.

We also find that a lower elasticity of substitution is associated with higher GFT (as it implies thicker tails in the productivity distribution) and welfare is rather sensitive to the choice of ε. GFT range between 2% and 14% when the elasticity of substitution moves from 8 to 2 in the case of a Pareto productivity distribution, which appears the most sensitive to the specific value of ε chosen. This sensitivity represents the main limitation of choosing the Pareto distribution to describe firm size and productivity,

[12]In fact, Behrens et al. (2012, footnote 8) note that "estimation results for ε depend both on the level of aggregation and the estimation method, and vary widely. For example, Hanson (2005) using aggregate U.S. data, obtains about 7 with non-linear least squares and about 2 with GMM. Estimates in Hummels (1999) vary from 2 to 5.26. Using extremely disaggregated data, Broda and Weinstein (2006) estimate several thousand elasticities of substitution, which range, depending on the industry and the level of aggregation, from 1.3 (telecommunication equipments) to 22.1 (crude oil)."

[13]Remember that variable export costs are calibrated exploiting the relationship $\tau^{1-\varepsilon}/(1+\tau^{1-\varepsilon}) = 0.14$, where the right-hand-side of the equation equals the average share of exports in total sales.

[14]The only cost parameter that does not change is the cost of serving the domestic market $f_d = 1$, which acts as a reference point throughout the paper and does not affect GFT.

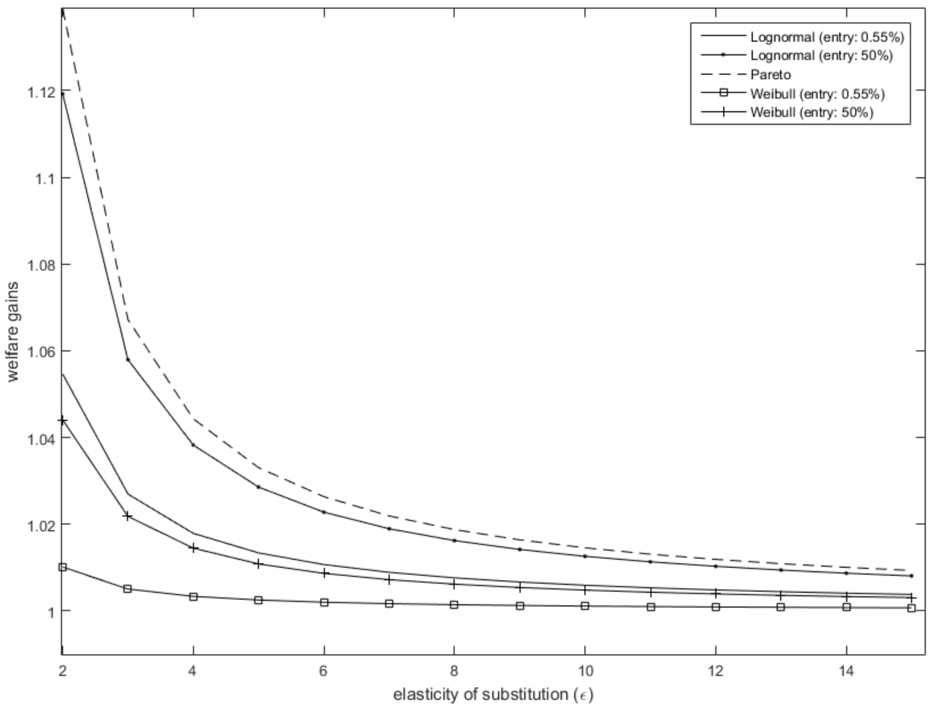

Figure 3: Welfare gains of trade liberalization as a function of ε: α, k and σ vary with ε; moreover, we also adjust τ and f_x in order to maintain constant the average export intensity and export participation.

since there is a great deal of uncertainty about the actual value of ε. On the other hand, the Weibull distribution is not very affected by changes in the elasticity, and even when we use the high entry rate of 50%, the variation in GFT is half as large as the one displayed under a Pareto, with the Weibull falling somewhat in between these two extremes.

The robustness checks performed in this section convey two main messages. First, the degree of tail heaviness of the productivity distribution has a strong impact on the magnitude of the GFT, and this is true irrespective of the specific distributional assumption one makes. Second, the sensitivity of the GFT to the elasticity of substitution, whose precise value is difficult to determine, increases with the degree of tail heaviness. Hence, heavier tails imply not only larger GFT but also more uncertainty in their actual magnitude.

4.3 Policy experiment: Reduction in variable and fixed trade costs

So far we have focused on the welfare effect of a trade liberalization that moves an economy from autarky to a free trade equilibrium. While this is the traditional way welfare effects are measured in the international trade literature, autarky is not a very relevant benchmark, as it is very seldom observed in reality. If, as we have argued in the Introduction, an important feature of the new-new trade revolution is to make economic theory closer to business and policy by providing a firm-level microfoundation to aggregate gains from trade, then it is worth looking at a more meaningful policy experiment as well. This is what we do in this section, where we focus on the effects of reducing either variable or fixed trade costs. In so doing, we compare welfares of two different trading equilibria, rather than looking at the effect of switching from autarky to free trade. This exercise is similar to the analysis performed in Melitz and Redding (2015) and answers a more immediate policy question, since the estimated welfare impact of a trade agreement often captures a lot of attention.[15] A case in point is the Transatlantic Trade and Investment Partnership (TTIP) currently under negotiation between the US and the EU, whose potential benefits have been subject to numerous studies (see, among others, Bertelsmann, 2013; CEPII, 2013; CEPR, 2013; Felbermayr and Larch, 2013).

We start by looking at the welfare effect of a 10% reduction in iceberg transport costs: we do this for various levels of ε since, as shown above, GFT are rather sensitive to this parameter and little consensus exists on its actual value. The results are presented in Figure 4. The GFT under Pareto are always larger than those obtained with alternative distributions,

[15]A recent contribution by Breinlich and Cuñat (2015) shows that a workhorse heterogeneous-firm model à la Melitz (2003) severely underestimates the gains from NAFTA, unless it is extended to allow for within-firm productivity increases. We are aware that the quantitative evaluation performed in this section therefore represents a rough approximation. On the other hand, as long as within-firm productivity improvements do not affect the shape of the distribution, the extension advocated by Breinlich and Cuñat should not impact the comparison of the model results under different distributional assumptions, which is the focus of our work.

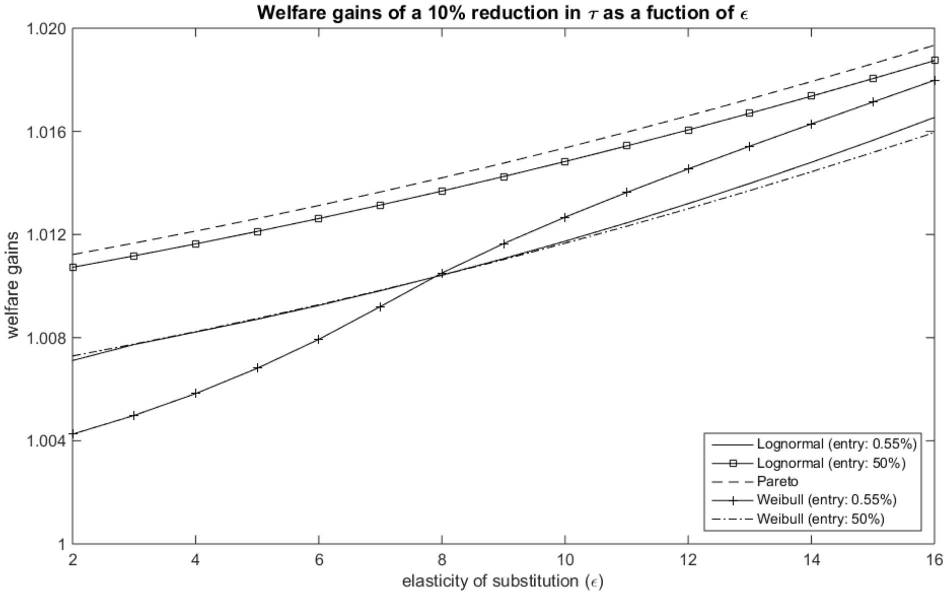

Figure 4: Welfare effect of a 10% reduction in iceberg transport costs (τ) in the lognormal and Pareto cases.

although (as already seen in Figure 1) when the entry rate is large the lognormal distribution yields a very similar welfare effect for each level of ε. The system displays a rather peculiar behavior when the productivity distribution follows a Weibull distribution and the entry rate is low (0.55%). For low values of the elasticity of substitution ($2 < \varepsilon < 8$), the GFT are low, but tend to increase quickly as ε increases: for $\varepsilon > 8$, the GFT associated with the Weibull distribution are larger than those estimated using a lognormal one. This nonlinear behavior is driven by the fact that in the Weibull case (with an entry rate of 0.55% and $\varepsilon > 8$) the new trading equilibrium after the reduction in τ features an export participation of 100%. Hence, there is no more selection into exporting, and welfare is much larger than before. Overall, the welfare impact of a 10% reduction in variable trade costs is limited and ranges between 0.2% (Weibull, entry rate 0.55%, $\varepsilon = 2$) and 1.9% (Pareto, $\varepsilon = 16$). Once again we observe substantial variation depending on the distribution and depending on the values of the elasticity of substitution.

What about a reduction in fixed export costs? This question is particularly interesting as the most recent trade negotiations, such as the TTIP between the EU and the US or the Trans-Pacific Partnership (TPP) involving 12 Pacific countries, are especially focused on eliminating the myriad non-tariff barriers that impose additional fixed costs on cross-border trade. In the case of the TTIP, most investigations have estimated that a successful deal would cut non-tariff barriers by something between 10% and 25%. We therefore look at the impact of a 25% reduction in f_x relative to the benchmark value that yields an export participation of 18%, and evaluate this impact for various levels of the elasticity of substitution.[16]

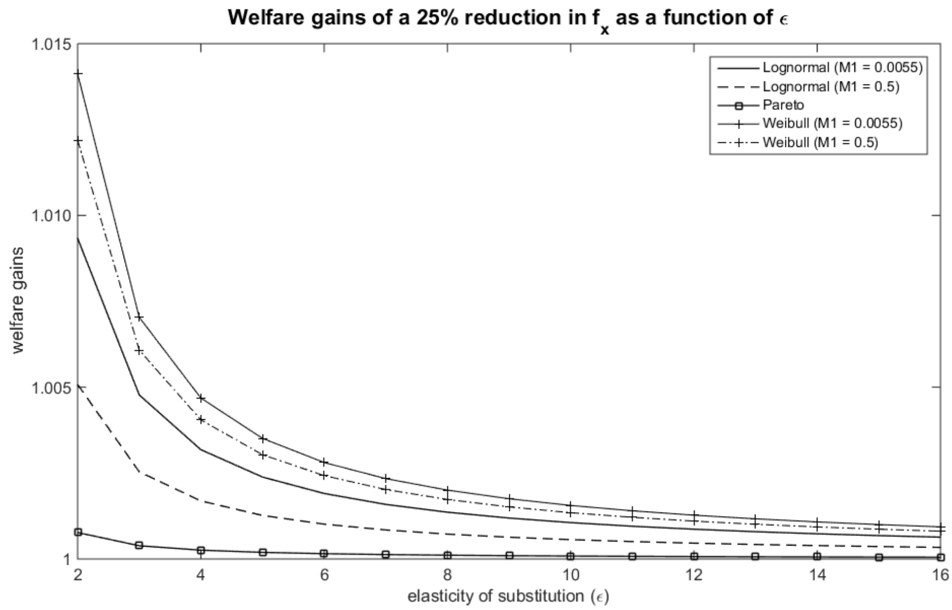

Figure 5: Welfare effect of a 25% reduction in fixed export costs (f_X).

The results are summarized in Figure 5. In this case, higher values of ε —which induce thicker tails in the productivity distribution— are associated with lower GFT. In fact, when more probability is concentrated in the upper tail of the productivity (and size) distribution, any given reduction in fixed export costs affects, *ceteris paribus*, a smaller number of firms and has a smaller impact on the extensive trade margin. In other words, the thicker the tails of the size distribution, the more likely it is that very large

[16]The results for a more modest reduction of 10% in fixed export costs are qualitatively similar and are available upon request.

firms are already exporting even before the reduction in f_x and, thus, the smaller the welfare effect of the liberalization. As noted by di Giovanni and Levchenko (2013), thicker tails imply a bigger difference between the largest exporters and the marginal firm that represents the extensive trade margin of a variation in fixed costs. Hence, in this specific example, countries with a more homogeneous distribution of firm size should reap more benefits from a reduction in fixed trade costs (provided that size reflects productivity and that smaller firms do not face specific constraints to participating in exporting activities such as, for instance, restricted access to financial resources, see for example Bellone et al., 2010; Minetti and Zhu, 2011)

More generally, Figure 5 shows the dramatic difference in GFT across distributions: if one assumes a Pareto distribution, then the results are negligible across the whole range of ε, whereas in the lognormal or Weibull cases, the estimated welfare effects range between 0.2% and 0.5% (assuming $\varepsilon = 4$) depending on the entry rate one decides to use.

The results in this subsection suggest that even when we look at the effect of a reduction in the (variable or fixed) trade costs starting from calibrated values (rather than looking at a putative move from autarky to free trade), we still find important differences in the GFT across distributions. What is more, the relative impact of a change in the fixed rather than the variable trade costs changes dramatically.

4.4 Homogeneous vs. heterogeneous models

We conclude the analysis with a discussion of the additional benefits associated with firm heterogeneity relative to a model where all firms are assumed homogeneous, an issue that is discussed at length in the recent papers by Arkolakis et al. (2012) and Melitz and Redding (2015).

Going back to the comparison between autarky and a free trade equilibrium, Table 3 (column 3) shows that the *additional* welfare effect of heterogeneity can be substantial, but depends on the specific distribution. It ranges from a meager 0.34% for the Weibull with an entry rate of 0.55% to 2.15% in the Pareto case with $\alpha = 4.25$, and 2.17% for a lognormal distribution with entry set at 50%.

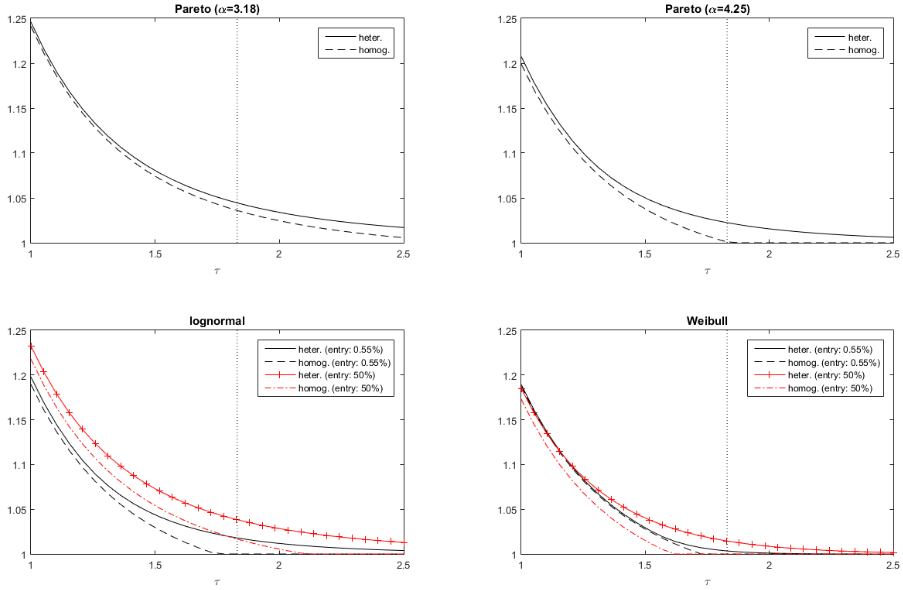

Figure 6: Welfare gains of trade liberalization as a function of τ: Heterogeneous vs. homogeneous models.

Figure 6 compares the GFT of the homogeneous model with those of the heterogeneous model under the different distributional assumptions we have been adopting. Clearly, the closer the trading equilibrium is to free trade ($\tau \rightarrow 1$), the smaller the difference is between the two models, since in both cases almost all firms would export once trade is possible. For moderate values of the variable costs, the additional welfare benefit associated with heterogeneity can be large.

In Table 4, we see that the homogeneous model is often more sensitive to changes in f_x than is the heterogeneous setup, although the overall welfare effect remains lower. This is consistent with the mechanism discussed above, which refers to the impact of liberalization on the extensive margin of trade: after all, the homogeneous model is nothing but a limiting case featuring a degenerate productivity distribution where all firms are alike.

A variation in the degree of firm heterogeneity such as the one discussed in Section 4.2.1 does have an effect also in the homogeneous model, because it alters the average productivity that maps the two models and makes them comparable. Figure 2 shows that for a large range of values of k, the

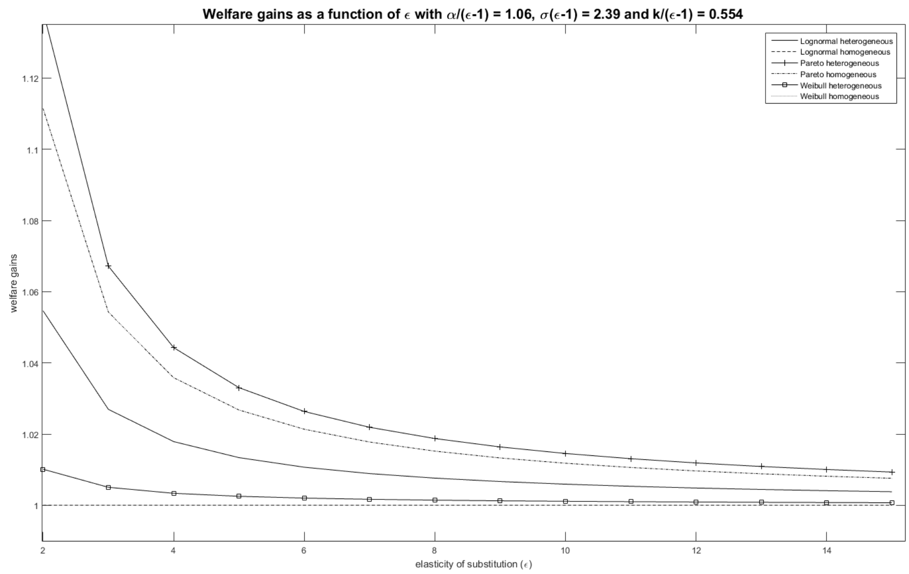

Figure 7: Welfare gains of trade liberalization as a function of ε: Heterogeneous vs. homogeneous models. The parameters α, k and σ vary with ε; moreover, we also adjust τ and f_x in order to keep average export intensity and export participation constant. GFT in the lognormal and Weibull homogeneous cases are constantly equal to 1; hence, they are indistinguishable in the figure.

homogeneous model associated with a Weibull distribution would deliver no gains from a trade liberalization. The same happens with the lognormal distribution, albeit for a smaller range of values of σ. In any case, depending on the specific parametrization, the gap between the estimated GFT under the two models can be important.

Lastly, we compare the heterogeneous and homogeneous models when ε is allowed to vary between 2 and 15. Figure 7 displays a few significant results. For the Pareto case, the difference between the two models is slightly larger for small values of ε, but remains roughly constant across the range of values taken by the elasticity of substitution. On the other hand, the parametrization we have chosen implies that the homogeneous model yields no GFT in either the lognormal or the Weibull case. Since welfare increases as ε goes down in the heterogeneous model, the gap between the two setups becomes larger and larger. This is especially true for the case of a lognormal distribution of firm productivity. Hence, even the additional welfare effect associated with firm heterogeneity is sensitive to the choice of the elasticity of substitution.

5 Conclusion

Large and persistent heterogeneity among firms has become a central tenet of the present-day trade literature, but most existing research has crystallized around a specific shape of firm heterogeneity, postulating that the latter is well described by a Pareto distribution. As more evidence emerges showing how the Pareto distribution does not do a good job of describing the distribution of firm size, this paper has investigated what happens to the magnitude of the gains from trade (GFT) when one departs from the standard Pareto assumption and considers alternative distributions, such as the lognormal and the Weibull.

We took stock of the existing literature showing that the degree of heterogeneity in firm size and productivity matters a lot for both the magnitude and the composition of the welfare effect of trade liberalization (di Giovanni and Levchenko, 2013). The reason for this is the relative importance of marginal firms that represent the extensive margin of trade, versus large infra-marginal enterprises. Instead of simply comparing different parameters for the same distribution, we pushed the argument one step further and evaluated the effect of choosing different productivity distributions. In so doing, we complement recent evidence by Head et al. (2014) by offering a thorough comparison between the GFT obtained under a Pareto, a lognormal, and a Weibull distribution, as well as the sensitivity of the results to a number of key parameters, such as the elasticity of substitution.

We have found that the distributional assumption used in a model featuring heterogeneous firms matters and has a sizable effect on the magnitude of gains from trade. The GFT increase as the upper tail of the productivity distribution gets heavier (Melitz and Redding, 2015), and this result carries over to the lognormal and Weibull cases.

Recent empirical evidence makes it quite clear that the Pareto distribution does not provide a very good fit to firm size and productivity, as it only captures (at best) the upper tail of the distribution. However, it allows closed-form solutions, which is an important advantage, as are the facts that the GFT are independent of the entry rate and the calibration of f_x is not affected by changes in the elasticity of substitution. On the other hand, the estimated GFT under a Pareto distribution appear more sensitive to variations in ε, relative to the other two distributions investigated here.

When we applied the welfare analysis to a policy experiment that mimics the potential effect of a 25% reduction in fixed trade costs (which represents one of the "optimistic" estimated results from successful TTIP negotiations), we have seen that the magnitude of the welfare impact changes dramatically with the underlying productivity distribution.

Overall, we have confirmed our working hypothesis: the choice of the productivity distribution has an important impact on the estimated GFT implied by the standard two-country heterogeneous-firm trade model. The Pareto distribution has a number of very useful analytical properties that make it very well suited for theoretical analysis, but at the same time does not provide a very good fit to the data. As a result, its use for policy analysis may lead to grossly biased estimates of the welfare effects of trade liberalization. We suggest using varied distributional assumptions and experimenting with alternative parameter sets for each of them. Presenting a range of estimated GFT might be less elegant and eye-catching than delivering a single number, but it is surely safer.

References

Arkolakis, C., Costinot, A. and Rodriguez-Clare, A. (2012), 'New trade models, same old gains?', *American Economic Review* **102**(1), 94–130.

Axtell, R. L. (2001), 'Zipf distribution of U.S. firm sizes', *Science* **293**(5536), 1818–1820.

Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2013), 'Cross-Country Differences in Productivity: The Role of Allocation and Selection', *American Economic Review* **103**(1), 305–34.

Bee, M., Riccaboni, M. and Schiavo, S. (2014), Where Gibrat meets Zipf:

Scale and scope of French firms, Discussion Paper Series 2014/3, Department of Economics and Management - University of Trento.

Behrens, K., Ertur, C. and Koch, W. (2012), ‘"Dual" gravity: Using spatial econometrics to control for multilateral resistance’, *Journal of Applied Econometrics* **27**(5), 773–794.

Bellone, F., Kiyota, K., Matsuura, T., Musso, P. and Nesta, L. (2014), ‘International productivity gaps and the export status of firms: Evidence from France and Japan’, *European Economic Review* **70**(C), 56–74.

Bellone, F., Musso, P., Nesta, L. and Schiavo, S. (2010), ‘Financial constraints and firm export behaviour’, *World Economy* **33**(3), 347–373.

Bernard, A. B., Eaton, J., Jensen, J. B. and Kortum, S. (2003), ‘Plants and productivity in international trade’, *American Economic Review* **93**(4), 1268–1290.

Bernard, A., Jensen, B., Redding, S. and Schott, P. (2007), ‘Firms in international trade’, *Journal of Economic Perspectives* **21**(3), 105–130.

Bertelsmann (2013), Transatlantic trade and investment partnership (TTIP). Who benefits from a free trade deal?, Technical report, Bertelsmann Stiftung.

Breinlich, H. and Cuñat, A. (2015), ‘Tariffs, trade and productivity: A quantitative evaluation of heterogeneous firm models’, *The Economic Journal* **in press**.

Broda, C. and Weinstein, D. E. (2006), ‘Globalization and the Gains from Variety’, *The Quarterly Journal of Economics* **121**(2), 541–585.

CEPII (2013), Transatlantic trade: Whither partnership, which economic consequences?, Technical report, Centre d’Études Prospectives et d’Informations Internationales, Paris.

CEPR (2013), Reducing transatlantic barriers to trade and investment, Technical report, Centre for Economic Policy Research, London.

Cernat, L. (2014), Towards "trade policy analysis 2.0": From national comparative advantage to firm-level trade data, Chief Economist Note 4, EU Commission - DG TRADE, Brussels.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012), 'The productivity advantages of large cities: Distinguishing agglomeration from firm selection', *Econometrica* **80**(6), 2543–2594.

Cooke, R. M., Nieboer, D. and Misiewicz, J. (2014), *Fat-Tailed Distributions: Data, Diagnostics and Dependence*, Wiley-ISTE, London, UK.

di Giovanni, J. and Levchenko, A. A. (2013), 'Firm entry, trade, and welfare in Zipf's world', *Journal of International Economics* **89**(2), 283–296.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer.

Felbermayr, G. and Larch, M. (2013), 'The transatlantic trade and investment partnership (TTIP): Potential, problems and perspectives', *CESifo Forum* **14**(2), 49–60.

Greenaway, D. and Kneller, R. (2007), 'Firm heterogeneity, exporting and foreign direct investment', *The Economic Journal* **117**(517), F134–F161.

Hanson, G. (2005), 'Market potential, increasing returns and geographic concentration', *Journal of International Economics* **67**, 1–24.

Head, K., Mayer, T. and Thoenig, M. (2014), 'Welfare and trade without Pareto', *Papers and Proceedings of the American Economic Review* **104**(5), 3010–316.

Hsieh, C.-T. and Klenow, P. J. (2009), 'Misallocation and Manufacturing TFP in China and India', *The Quarterly Journal of Economics* **124**(4), 1403–1448.

Hummels, D. (1999), Toward a geography of trade costs, GTAP Working Papers 1162, Center for Global Trade Analysis, Purdue University.

Kleiber, C. and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley.

Kratz, M. and Resnick, S. (1996), 'The Q-Q estimator and heavy tails', *Stochastic Models* **12**(4), 699–724.

Krugman, P. (1979), 'Increasing returns, monopolistic competition, and international trade', *Journal of International Economics* **9**(4), 469–479.

Krugman, P. (1980), 'Scale Economies, Product Differentiation, and the Pattern of Trade', *American Economic Review* **70**(5), 950–59.

Malevergne, Y., Pisarenko, V. and Sornette, D. (2011), 'Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal', *Physical Review E* **83**, 036111.

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 Redding, S. J. (2013), Firm heterogeneity and aggregate welfare, Working Paper 18919, NBER.

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

Minetti, R. and Zhu, S. C. (2011), 'Credit constraints and firm export: Microeconomic evidence from Italy', *Journal of International Economics* **83**(2), 109–125.

Rinne, H. (2009), *The Weibull Distribution: A Handbook*, Chapman and Hall.

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

Söllner, R. (2010), Product Diversification and Labor Productivity Dispersion in German Manufacturing Industries, Jena Economic Research Papers 2010-028, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics.

Syverson, C. (2004), 'Product Substitutability and Productivity Dispersion', *The Review of Economics and Statistics* **86**(2), 534–550.

Virkar, Y. and Clauset, A. (2014), 'Power-law distributions in binned empirical data', *Annals of Applied Statistics* **8**(1), 89–119.

Appendices

A Model derivation under the lognormal and Weibull assumptions

The threshold productivity levels needed to compute welfare gains are implicitly defined by the free entry condition of the model. This equates the expected value of profits (conditional on surviving) to the sunk entry costs, $\left[1 - G(\phi_d^T)\right] \bar{\pi} = w f_e$, where G is the productivity cumulative distribution function (CDF) and $\left[1 - G(\phi_d^T)\right]$ gives the proportion of firms that successfully enter the market. The zero profit condition implies, both in the domestic and the foreign market, that the marginal firm possessing the threshold value of productivity earns just enough to pay for its fixed costs (of production and, possibly, export). Exploiting the zero profit conditions for the domestic and foreign markets, one can establish a relationship between the domestic and export productivity cutoffs:

$$\phi_x^T = \tau \left(\frac{f_x}{f_d}\right)^{\frac{1}{\varepsilon-1}} \phi_d^T. \tag{A4}$$

Combining these elements, and choosing a specific distribution G for the productivity ϕ, it is then possible to compute the GFT. We next give in detail the solution for the lognormal and Weibull cases.

A.1 Closed economy

We have to solve for ϕ_d^A in the free entry condition, which can be written as

$$\frac{f_e}{f_d} = [1 - G(\phi_d^A)] \left[\left(\frac{\tilde{\phi}_d^A}{\phi_d^A}\right)^{\varepsilon-1} - 1\right], \tag{A5}$$

where $(\tilde{\phi}_d^A)^{\varepsilon-1}$ is given by

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \int_{\phi_d^A}^{\infty} \phi^{\varepsilon-1} \frac{g(\phi)}{1 - G(\phi_d^A)} d\phi \tag{A6}$$

and $g(\cdot)$ is the probability distribution function (PDF) of ϕ. The integral (A6) is

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A). \tag{A7}$$

For integer ε, equation (A7) is the $(\varepsilon-1)$th moment of $\phi|\phi > \phi_d^A$. Hence, the problem consists in solving (A5) for ϕ_d^A with $\tilde{\phi}_d^A$ given by (A7).

A.1.1 The lognormal case

Under the assumption of lognormality, $G(x) \stackrel{\text{def}}{=} G_{\mu,\sigma^2}(x) = \Phi_{\mu,\sigma^2}(\log(x))$ is the Logn(μ, σ^2) CDF, while $(\tilde{\phi}_d^A)^{\varepsilon-1}$ is given by

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \int_{\phi_d^A}^{\infty} \phi^{\varepsilon-1} \frac{g_{\mu,\sigma^2}(\phi)}{1 - G_{\mu,\sigma^2}(\phi_d^A)} d\phi, \tag{A8}$$

with g_{μ,σ^2} being the Logn(μ, σ^2) PDF. By definition, the integral (A8) is equal to (A6) with $\phi \sim$ Logn(μ, σ^2). Under this assumption, (A8) is the $(\varepsilon-1)$th moment of the Logn(μ, σ^2) distribution left-truncated at ϕ_d^A, which can be rewritten as

$$\mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A) = \mathrm{E}(\phi^{\varepsilon-1}) \frac{\Phi((\varepsilon-1)\sigma - a_0)}{\Phi(-a_0)}, \tag{A9}$$

where $a_0 = (\log(\phi_d^A) - \mu)/\sigma$. As $\mathrm{E}(\phi^{\varepsilon-1}) = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}$, (A9) is finally given by

$$\mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A) = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\} \frac{\Phi((\varepsilon-1)\sigma - a_0)}{\Phi(-a_0)}. \tag{A10}$$

From (A8) and (A10) it readily follows that

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\} \frac{\Phi((\varepsilon-1)\sigma - a_0)}{\Phi(-a_0)}. \tag{A11}$$

Hence, the problem consists in solving (A5) for ϕ_d^A with $\tilde{\phi}_d^A$ given by (A11), namely:

$$\frac{f_e}{f_d} = [1 - G_{\mu,\sigma^2}(\phi_d^A)] \left[\frac{\exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}\Phi((\varepsilon-1)\sigma - a_0)/\Phi(-a_0)}{(\phi_d^A)^{\varepsilon-1}} - 1 \right].$$

The solution of this equation must be found numerically.

A.1.2 The Weibull case

Assuming that the productivity ϕ follows a Weibull distribution with scale and shape parameters respectively equal to λ and k, $\phi \sim \text{Weib}(\lambda, k)$, we have that $G(x) \stackrel{\text{def}}{=} G_{\lambda,k}(x)$ is the Weibull CDF, while $(\tilde{\phi}_d^A)^{\varepsilon-1}$ is given by

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \int_{\phi_d^A}^{\infty} \phi^{\varepsilon-1} \frac{g_{\lambda,k}(\phi)}{1 - G_{\lambda,k}(\phi_d^A)} d\phi, \tag{A12}$$

with $g_{\lambda,k}(\cdot)$ being the $\text{Weib}(\lambda, k)$ probability distribution function (PDF). Equation (A12) is the $(\varepsilon-1)$th moment of the $\text{Weib}(\lambda, k)$ distribution left-truncated at ϕ_d^A. This conditional expectation is given by (Rinne, 2009, equation 3.49a):

$$\begin{aligned}
&(\tilde{\phi}_d^A)^{\varepsilon-1} = \text{E}(\phi^{\varepsilon-1} | \phi > \phi_d^A) = \\
&= \exp\left\{ \left(\frac{\phi_d^A}{\lambda} \right)^k \right\} \lambda \left(\Gamma\left(\frac{1}{k} + 1, \infty \right) - \Gamma\left(\frac{1}{k} + 1, \left(\frac{\phi_d^A}{\lambda} \right)^k \right) \right),
\end{aligned} \tag{A13}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete Gamma function. Thus, one has to solve (A5) for ϕ_d^A with $\tilde{\phi}_d^A$ given by (A13), namely:

$$\frac{f_e}{f_d} = [1 - G_{\lambda,k}(\phi_d^A)] \left[\frac{\exp\left\{ \left(\frac{\phi_d^A}{\lambda} \right)^k \right\} \lambda \left(\Gamma\left(\frac{1}{k} + 1, \infty \right) - \Gamma\left(\frac{1}{k} + 1, \left(\frac{\phi_d^A}{\lambda} \right)^k \right) \right)}{(\phi_d^A)^{\varepsilon-1}} - 1 \right].$$

As in the lognormal case, the solution of this equation must be found numerically.

A.2 Open economy

Similar results hold for an open economy. If $\tau(f_x/f_d)^{1/(\varepsilon-1)} > 1$, only the most productive firms export. Given the relationship between the domestic and export productivity thresholds defined by equation (A4) above, we also have that $\phi_x^T > \phi_d^T$.[17] The free entry condition to be solved for ϕ_d^T can be

[17] On the other hand, if $\tau(f_x/f_d)^{1/(\varepsilon-1)} \leq 1$, then all firms export and $\phi_x^T = \phi_d^T$.

written as

$$f_e = f_d[1-G(\phi_d^T)]\left[\left(\frac{\tilde{\phi}_d^T}{\phi_d^T}\right)^{\varepsilon-1}-1\right]+f_x[1-G(\phi_x^T)]\left[\left(\frac{\tilde{\phi}_x^T}{\phi_x^T}\right)^{\varepsilon-1}-1\right], \tag{A14}$$

where $\tilde{\phi}_x^T$ is the average productivity in the export market, and

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \int_{\phi_x^T}^{\infty} \frac{\phi^{\varepsilon-1}g(\phi)}{1-G(\phi_x^T)}d\phi. \tag{A15}$$

As in the closed-economy, lognormality implies that equation (A15) is given by

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \exp\{(\varepsilon-1)\mu+(\varepsilon-1)^2\sigma^2/2\}\frac{\Phi((\varepsilon-1)\sigma-a_0)}{\Phi(-a_1)}, \tag{A16}$$

where $a_1 = (\log(\phi_x^T)-\mu)/\sigma$.

Analogously, if $\phi \sim \text{Weib}(\lambda,k)$, (A15) can be rewritten as

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \exp\left\{\left(\frac{\phi_d^A}{\lambda}\right)^k\right\}\lambda\left(\Gamma\left(\frac{1}{k}+1,\infty\right)-\Gamma\left(\frac{1}{k}+1,\left(\frac{\phi_d^A}{\lambda}\right)^k\right)\right), \tag{A17}$$

For both distributions, the functional form of the ratio W^T/W^A is given by:[18]

$$\frac{W^T}{W^A} = \begin{cases} \frac{\phi_d^T}{\phi_d^A} & \text{if } \tau\left(\frac{f_x}{f_d}\right)^{1/(\varepsilon-1)} > 1; \\ \left(\frac{1+\tau^{1-\sigma}f_d}{f_d+f_x}\right)^{1/(\sigma-1)}\frac{\phi_d^T}{\phi_d^A} & \text{if } \tau\left(\frac{f_x}{f_d}\right)^{1/(\varepsilon-1)} \leq 1. \end{cases} \tag{A18}$$

The GFT for both distributions are computed by means of (A18). The solution corresponding to the hypothesis of lognormal productivity is obtained by plugging into (A18) the solution of (A14) with $\tilde{\phi}_x^T$ given by (A16). Similarly, the GFT corresponding to the Weibull assumption are given by (A18) with ϕ_d^T replaced by the solution of (A14) with $\tilde{\phi}_x^T$ given by (A17).

In the Pareto case, $1-G(\phi) = (\phi_{min}/\phi)^k$ and $(\tilde{\phi}_d^A/\phi_d^A)^{\epsilon-1} = (\tilde{\phi}_d^T/\phi_d^T)^{\epsilon-1} = \alpha/(\alpha-\epsilon+1)$: this greatly simplifies the computations in that (A5) and (A14) can be solved explicitly, so that the productivity cutoffs are com-

[18] See equations (11) and (27) in the Web Appendix of Melitz and Redding (2015).

puted in closed form.[19] On the other hand, in the lognormal case neither the CDF nor the average productivity can be computed explicitly, whereas for the Weibull only the CDF is available in closed form. Thus, in both cases, (A5) and (A14) can only be solved numerically.

[19]We refer the interested reader to the excellent Web Appendix of the paper by Melitz and Redding (2015).

B Distribution Fitting

Table B1: QQ estimation of the shape parameter for the Pareto, Lognormal and Weibull distributions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sample:	all	top 50%	top 25%	top 5%	top 4%	top 3%	top 2%	top 1%
obs.:	36881	18441	9220	1844	1475	1106	737	369
Pareto:								
α^{sales}	2.143*	1.400*	1.182*	0.922*	0.891*	0.859*	0.826*	0.781*
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
R^2	0.809	0.966	0.981	0.989	0.992	0.994	0.994	0.995
α	1.400	2.143	2.538	3.254	3.366	3.492	3.633	3.839
Lognormal:								
σ^{sales}	2.380*	2.360*	2.425*	2.487*	2.471*	2.461*	2.470*	2.493*
	(0.000)	(0.000)	(0.001)	(0.002)	(0.003)	(0.003)	(0.005)	(0.011)
R^2	1.000	0.999	0.999	0.999	0.998	0.998	0.997	0.993
σ	0.793	0.787	0.808	0.829	0.824	0.820	0.823	0.831
Weibull:								
$1/k^{sales}$	1.805*	2.818*	3.362*	4.247*	4.309*	4.404*	4.573*	4.857*
	(0.002)	(0.003)	(0.004)	(0.007)	(0.009)	(0.012)	(0.016)	(0.031)
R^2	0.945	0.982	0.988	0.995	0.994	0.993	0.991	0.985
k	1.662	1.065	0.892	0.706	0.696	0.681	0.656	0.618

Standard errors in parentheses. * indicates significance at the 1% level
α^{sales}, σ^{sales} and k^{sales} represent estimates from the distribution of export sales to Belgium; the corresponding parameters for the distribution of productivity (α, σ and k) are obtained assuming an elasticity of substitution $\epsilon = 4$.