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

Marco Bee · Stefano Schiavo

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 three different productivity distributions (Pareto, lognormal, and Weibull) and investigates their policy implications. We find that both the magnitude of the welfare gains and the relative importance of the fixed versus variable trade costs change significantly. Hence, relying blindly on a single distribution is dangerous when performing trade policy analysis.

Keywords lognormal · Pareto · Weibull · international trade · welfare · firm heterogeneity

PACS 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 distribution, which characterizes most of the recent 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. This shift adds a number of useful insights and moves trade theory closer to business and policymakers by providing microfoundations to aggregate gains from trade

Marco Bee · Stefano Schiavo
Department of Economics and Management, University of Trento, via Inama 5, 38122 Trento (Italy)
E-mail: marco.bee@unitn.it; stefano.schiavo@unitn.it

Stefano Schiavo
School of International Studies, University of Trento

Stefano Schiavo
DRIC, OFCE-Sciences Po

(Greenaway and Kneller, 2007; Cernat, 2014). However, the welfare implications of the new-new trade theory have not been much explored until recently (see, for instance, Arkolakis et al., 2012; di Giovanni and Levchenko, 2013; Head et al., 2014; Behrens et al., 2014; Melitz and Redding, 2015; Bas et al., 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).[1]

The starting point of this recent stream of research can be traced to the contribution by Arkolakis et al. (2012), who aim at assessing the additional gains from trade (GFT) associated with firm heterogeneity. On top of the welfare benefits already present in the new-trade literature (Krugman, 1980), heterogeneity adds a further 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 heterogeneity does not add much insight to what is already known about the size and composition of the gains from trade. Melitz and Redding (2015) 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 any departure from the assumption of a Pareto distribution for firm productivity (Bas et al., 2015), 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.[2] Bas et al. (2015) explore in further detail the relationship between trade elasticities and welfare: they find evidence that the hypothesis of a constant trade elasticity across destinations, implied by the assumption of a Pareto distribution, is not supported by the data.

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 is that productivity

---

[1] A few papers have explored the impact of modifying the demand (preference) rather than the supply (productivity distribution) side of recent trade models. This is done either by means of a generalization of the standard CES demand system (as in Behrens et al., 2014, who nest the baseline CES as a special case) or by an outright departure from CES (as in Mrázová and Neary, 2014). The bottom line is that GFT are much harder to pin down. Although we find this line of research extremely interesting, we do not pursue this approach in the present paper.

[2] 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.

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, di Giovanni and Levchenko (2013) find that if the shape parameter of the productivity distribution increases from 1 to 2 (thus lowering dispersion), 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.

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 trade 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.[3]

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 often 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 a poor job when applied to the whole size/productivity distribution, and fits reasonably well only to the upper tail (see for instance Combes et al., 2012; Freund and Pierola, 2015). And even this milder result 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).[4]

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 extending the analysis to a third distribution, namely the Weibull, by thoroughly exploring the policy implications of the results, and by performing a series of robustness checks. The choice of the Weibull distribution is suggested by several reasons. First, the Weibull is a member of the Gamma-type family of size distributions, which is probably, with the Pareto and the

---

[3] 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.

[4] 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.

lognormal, the most common family of distributions used for size modeling (Kleiber and Kotz, 2003). Second, the Weibull shares with the lognormal and Pareto distributions the property of closure under exponentiation, which is practical as it allows one to easily link the parameters of the distributions of firm size and productivity. Third, similarly to the Pareto and the lognormal, also the Weibull distribution can be derived from economic theory.[5] Last, even though the tail behaviors of the Pareto and lognormal distributions are qualitatively different, the two can become almost indistinguishable Malevergne et al. (2011). Comparing them with a third distribution can therefore help to shed light on the effect that different distributional assumptions have on GFT.

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 relative importance of reductions in fixed versus variable trade costs, and the additional benefits associated with heterogeneity. For instance, a 25% reduction in fixed export costs (one of the projected results from a possible EU-US trade agreement) 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. We find that under the Pareto distribution welfare effects are especially sensitive to the value of this parameter.

We conclude that, when it comes to trade policy analysis, the choice of a specific productivity distribution 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 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; Section 4 discusses the results from our investigation, comparing GFT under different distributional assumptions; Section 5 evaluates the impact of a reduction in variable and fixed trade costs. Finally, Section 6 links our findings to the existing literature and concludes.

---

[5] Gabaix (2009) provides a comprehensive review of models leading to a Pareto distribution; the lognormal is generated under a process of proportionate growth à la Gibrat (1931); finally, Growiec (2013, p. 2337) shows that, if production consists of many complementary stages, "the Weibull should approximate the true productivity distribution better than anything else".

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 that 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 possible, 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 trade W^T, and show that in case only some firms self-select into export markets, *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 an equilibrium featuring international trade (autarky).

The threshold productivity levels ϕ_d^A and ϕ_d^T 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 $\bar{\pi}$ are the average firm profit, f_e are the sunk entry costs, 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:

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

where ε is the elasticity of substitution. Combining these elements, and choosing a specific distribution G for the productivity ϕ, it is then possible to compute the GFT.

We next give the solution for the Pareto, lognormal and Weibull cases. From a computational point of view, the main difference is that while the Pareto case yields

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.

2.1 Autarky

We have to solve for ϕ_d^A in the free entry condition, which (following Melitz and Redding, 2015) can be written as

$$\frac{f_e}{f_d} = [1 - G(\phi_d^A)] \left[\left(\frac{\tilde{\phi}_d^A}{\phi_d^A} \right)^{\varepsilon - 1} - 1 \right], \tag{2}$$

where $(\tilde{\phi}_d^A)$ is a weighted average of firm productivity and is given by

$$(\tilde{\phi}_d^A)^{\varepsilon - 1} = \int_{\phi_d^A}^{\infty} \phi^{\varepsilon - 1} \frac{g(\phi)}{1 - G(\phi_d^A)} d\phi \tag{3}$$

and $g(\cdot)$ is the probability distribution function (PDF) of ϕ. The integral (3) is

$$(\tilde{\phi}_d^A)^{\varepsilon - 1} = \mathrm{E}(\phi^{\varepsilon - 1} | \phi > \phi_d^A). \tag{4}$$

For integer ε, equation (4) is the $(\varepsilon - 1)$th moment of $\phi | \phi > \phi_d^A$. Hence, the problem consists in solving equation (2) for ϕ_d^A with $\tilde{\phi}_d^A$ given by (4).

The functional forms and the notation used for the PDF and the CDF of the three distributions are displayed in Table 1. It is worth noting that the lognormal is the only distribution whose CDF is not explicit.

Table 1: PDFs and CDFs of the Pareto, lognormal and Weibull distributions. $\mathbb{I}_A$ denotes the indicator function of the set A.

	PDF	CDF
$\mathrm{Par}(x; \phi_{min}, \alpha)$	$g_{\phi_{min}, \alpha}(x) = \frac{\alpha \phi_{min}^\alpha}{x^{\alpha+1}} \mathbb{I}_{x \geq \phi_{min}}$	$G_{\phi_{min}, \alpha}(x) = 1 - \left(\frac{\phi_{min}}{x} \right)^\alpha \mathbb{I}_{x \geq \phi_{min}}$
$\mathrm{Logn}(x; \mu, \sigma^2)$	$g_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{ -\frac{1}{2} \left(\frac{\log x - \mu}{\sigma} \right)^2 \right\} \mathbb{I}_{x > 0}$	$G_{\mu, \sigma^2}(x) = \int_0^x g_{\mu, \sigma^2}(t) dt$
$\mathrm{Weib}(x; \lambda, k)$	$g_{\lambda, k}(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} \exp\{-(x/\lambda)^k\} \mathbb{I}_{x \geq 0}$	$G_{\lambda, k}(x) = 1 - \exp\{-(x/\lambda)^k\} \mathbb{I}_{x \geq 0}$

When the productivity follows a Pareto distribution, one has $(\tilde{\phi}_d^A / \phi_d^A)^{\varepsilon - 1} = (\tilde{\phi}_d^T / \phi_d^T)^{\varepsilon - 1} = \alpha/(\alpha - \varepsilon + 1)$: this greatly simplifies the computations in that (2) can be solved explicitly, so that the productivity cutoff is computed in closed form (Melitz and Redding, 2015, online appendix):

$$\phi_d^A = \left[\frac{f_d}{f_e} \left(\phi_{min}^\alpha \frac{\varepsilon - 1}{\alpha - \varepsilon + 1} \right) \right]^{1/\alpha}, \qquad \alpha > \varepsilon - 1.$$

Under the assumption of lognormality, $(\tilde{\phi}_d^A)^{\varepsilon-1}$ is the $(\varepsilon-1)$th moment of the Logn(μ, σ^2) distribution left-truncated at ϕ_d^A, which can be rewritten as

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A) = \mathrm{E}(\phi^{\varepsilon-1})\frac{G_{0,1}((\varepsilon-1)\sigma - a_0)}{G_{0,1}(-a_0)}, \tag{5}$$

where $G_{0,1}$ is the cdf of the $N(0,1)$ random variable and $a_0 = (\log(\phi_d^A) - \mu)/\sigma$. As $\mathrm{E}(\phi^{\varepsilon-1}) = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}$, equation (5) is finally given by

$$\mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A) = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}\frac{G_{0,1}((\varepsilon-1)\sigma - a_0)}{G_{0,1}(-a_0)}. \tag{6}$$

From (5) and (6) it readily follows that

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}\frac{G_{0,1}((\varepsilon-1)\sigma - a_0)}{G_{0,1}(-a_0)}. \tag{7}$$

Hence, the zero profit condition (2) becomes:

$$\frac{f_e}{f_d} = [1 - G_{\mu,\sigma^2}(\phi_d^A)]\left[\frac{\exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2\sigma^2/2\}G_{0,1}((\varepsilon-1)\sigma - a_0)/G_{0,1}(-a_0)}{(\phi_d^A)^{\varepsilon-1}} - 1\right]$$

and the solution of this equation with respect to ϕ_d^A must be found numerically.

If the productivity follows a Weibull distribution with scale and shape parameters respectively equal to λ and k, Equation (4) becomes (Rinne, 2009, equation 3.49a):

$$(\tilde{\phi}_d^A)^{\varepsilon-1} = \mathrm{E}(\phi^{\varepsilon-1}|\phi > \phi_d^A) =$$
$$= \exp\left\{\left(\frac{\phi_d^A}{\lambda}\right)^k\right\}\lambda\left(\Gamma\left(\frac{1}{k}+1, \infty\right) - \Gamma\left(\frac{1}{k}+1, \left(\frac{\phi_d^A}{\lambda}\right)^k\right)\right), \tag{8}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete Gamma function. Thus, one has to solve the following zero profit condition for ϕ_d^A:

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

As in the lognormal case, the solution (ϕ_d^A) must be found numerically.

2.2 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 (1) above, we also have that $\phi_x^T > \phi_d^T$.[6] The free entry

---

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

condition to be solved for ϕ_d^T can be written as

$$f_e = f_d[1 - G(\phi_d^T)] \left[\left(\frac{\tilde{\phi}_d^T}{\phi_d^T} \right)^{\varepsilon-1} - 1 \right] + f_x[1 - G(\phi_x^T)] \left[\left(\frac{\tilde{\phi}_x^T}{\phi_x^T} \right)^{\varepsilon-1} - 1 \right], \qquad (9)$$

where $\tilde{\phi}_x^T$ is the average productivity in the export market, and the open economy counterpart of equation (3) is

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \int_{\phi_x^T}^{\infty} \frac{\phi^{\varepsilon-1} g(\phi)}{1 - G(\phi_x^T)} d\phi. \qquad (10)$$

Under the Pareto assumption, (9) can be solved explicitly, so that the productivity cutoff is again computed in closed form (Melitz and Redding, 2015, online appendix):

$$\phi_d^T = \left(\frac{\varepsilon - 1}{\alpha - \varepsilon + 1} \right)^{1/\alpha} \left[\frac{f_d + \left(\frac{1}{\tau(f_x/f_d)^{1/(\varepsilon-1)}} \right)^{\alpha} f_x}{f_e} \right]^{1/\alpha} \phi_{min}, \qquad \alpha > \varepsilon - 1.$$

As in the closed economy, lognormality implies that equation (10) is given by

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \exp\{(\varepsilon-1)\mu + (\varepsilon-1)^2 \sigma^2/2\} \frac{G_{0,1}((\varepsilon-1)\sigma - a_1)}{G_{0,1}(-a_1)}, \qquad (11)$$

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

Analogously, if $\phi \sim \text{Weib}(\lambda, k)$, (10) can be rewritten as

$$(\tilde{\phi}_x^T)^{\varepsilon-1} = \exp\left\{ \left(\frac{\phi_x^T}{\lambda} \right)^k \right\} \lambda \left(\Gamma \left(\frac{1}{k} + 1, \infty \right) - \Gamma \left(\frac{1}{k} + 1, \left(\frac{\phi_x^T}{\lambda} \right)^k \right) \right), \qquad (12)$$

For both distributions, the functional form of the ratio W^T/W^A is given by[7]

$$\frac{W^T}{W^A} = \begin{cases} \frac{\phi_d^T}{\phi_d^A} & \text{if } \tau \left(\frac{f_x}{f_d} \right)^{1/(\varepsilon-1)} > 1, \\ \left(\frac{1 + \tau^{1-\sigma} f_d}{f_d + f_x} \right)^{1/(\sigma-1)} \frac{\phi_d^T}{\phi_d^A} & \text{if } \tau \left(\frac{f_x}{f_d} \right)^{1/(\varepsilon-1)} \leq 1, \end{cases} \qquad (13)$$

and the GFT are computed by means of (13). The solution corresponding to the hypothesis of lognormal and Weibull productivity is obtained by plugging into equation (13) the solution of (9) with $\tilde{\phi}_x^T$ given either by (11) or by (12).

Note that, under both the lognormal and Weibull distributional assumptions, the free entry conditions in closed and open economy, given respectively by (2) and (9), can only be solved numerically.

2.3 The homogeneous case

Melitz and Redding (2015) treat the homogeneous case as a special (degenerate) instance 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

---

[7] See equations (13) and (26) in the Web Appendix of Melitz and Redding (2015).

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 ($\bar{\phi}_d = \tilde{\phi}_d^A$).

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.

To estimate those parameters we use information on the sales of French exporters to Belgium in 2003 and follow the same methodology applied by Head et al. (2014), namely QQ estimation.[8] From the parameters of the distribution of (log) sales we can infer those of the productivity distribution by virtue of the closed-form relationship between sales and productivity that is implied by the standard combination of CES preferences and monopolistic competition. 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 holds for the Weibull distribution (see McCool, 2012, p. 83). This relationship is useful 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:

$$\alpha = \alpha^{sales} \cdot (\varepsilon - 1) \qquad \text{(Pareto)} \tag{14}$$

$$\sigma = \sigma^{sales}/(\varepsilon - 1) \qquad \text{(lognormal)} \tag{15}$$

$$k = k^{sales} \cdot (\varepsilon - 1) \qquad \text{(Weibull).} \tag{16}$$

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

---

[8] The data on firm export are collected by French customs and have been accessed through the secure data access center CASD.

investigated by di Giovanni and Levchenko (2013), who compare GFT when $\alpha^{sales} = 1.06$ vs. $\alpha^{sales} = 2$. 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). Looking at the sales of French firms exporting to Belgium (the most popular destination for French exporters) in 2000, Head et al. (2014) find $\sigma^{sales} = 2.39$.[9]

3.2 QQ estimation

Kratz and Resnick (1996) introduced an estimation method that minimizes the sum of squared distances between the theoretical (i.e., implied by a given distribution) and empirical quantiles. In the present setup, this approach allows one to exploit the linearity of the relationship between the theoretical and the empirical quantiles of log sales data for the three aforementioned distributions.

Recall first that the logarithms of $X_1 \sim \text{Par}(x; \phi_{min}, \alpha)$, $X_2 \sim \text{Logn}(x; \mu, \sigma^2)$ and $X_3 \sim \text{Weib}(x; \lambda, k)$ are respectively distributed as $\log X_1 \sim \text{GExp}(\alpha)$, $\log X_2 \sim N(\mu, \sigma^2)$ and $\log X_3 \sim \text{EV}_{min}(\log \lambda, 1/k)$ (Rinne, 2009, Sect. 3.2.2). Here $\text{GExp}(\alpha)$ is the generalized exponential random variable with density

$$g(x; \phi_{min}, \alpha) = \alpha \exp\{-\alpha(x - \log \phi_{min})\}\mathbb{I}_{x > \log \phi_{min}},$$

and $\text{EV}_{min}(\log \lambda, 1/k)$ is the Extreme Value type-I minimum random variable with density

$$g_{EVmin}(x; \log \lambda, 1/k) = k \exp\{k(x - \log \lambda) - \exp\{k(x - \log \lambda)\}\}, \quad x \in \mathbb{R}.$$

The quantile functions of $\text{GExp}(\alpha)$, $N(\mu, \sigma^2)$ and $\text{EV}_{min}(\log \lambda, 1/k)$ are respectively given by

$$F_{GExp(\phi_{min}, \alpha)}^{-1}(p) = \log \phi_{min} - \frac{1}{\alpha} \log(1 - p),$$
$$F_{N(\mu, \sigma^2)}^{-1}(p) = \frac{\mu}{\sigma} + \sigma \Phi^{-1}(p),$$
$$F_{\text{EV}_{min}(\log \lambda, 1/k)}^{-1}(p) = \log(\lambda) + \frac{1}{k} \log(-\log(1 - p)),$$

with $p \in (0, 1)$. Accordingly, the QQ method regresses the empirical quantiles Q_i^E ($i = 1, \ldots, n$) of the log sales $x_1, \ldots, x_n$ on the theoretical quantiles:

$$Q_i^E = \log \phi_{min} - \frac{1}{\alpha} \log(1 - \hat{G}^E(x_i)),$$
$$Q_i^E = \frac{\mu}{\sigma} + \sigma \Phi^{-1}(\hat{G}^E(x_i)),$$
$$Q_i^E = \log(\lambda) + \frac{1}{k} \log(-\log(1 - \hat{G}^E(x_i))),$$

---

[9] Head et al. (2014) note that the theoretical link between sales and productivity only holds for individual export markets, not total exports (although the lognormal distribution provides a good fit to total sales by firm as well). Belgium is among the most popular destinations for French firms.

where $\hat{G}^E$ is the empirical CDF. By means of these three regressions one estimates respectively $1/\alpha$, σ and $1/k$, which represent the slopes of the corresponding regressions. The three distributions are equally parsimonious in that they are characterized by two parameters: on top of that, we only need to estimate one of them in order to quantify the GFT.[10] In fact, the mean of the lognormal (μ), as well as the scale parameters of the Weibull (λ) and the Pareto distributions (ϕ_{min}), do not enter into the computations of the gains from trade. This is because relative welfare is computed as a ratio of thresholds: hence, what matters is the shape of the distribution, not its location.

Table 2 reports results from a QQ estimation of the shape parameters for the Pareto, lognormal and Weibull distributions based on firm-level export by French firms to Belgium in 2003 (Table 7 in the Appendix performs the same exercise using data for different years and destinations, and find qualitatively similar results). The estimation is performed on the entire sample as well as on the subsets corresponding to different upper quantiles of the distribution. For each regression we report the sample size, the estimated parameter for sales, the corresponding value for the distribution of productivity, and the R^2 that we use to assess the goodness of fit. For the Pareto and lognormal, we find results very close to those reported by Head et al. (2014) both in terms of the parameter values and of the qualitative performance of the distribution. The Weibull provides a good fit to the data: its R^2 ranges between 0.950 and 0.995, and it 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 GFT to this parameter is one of the main points we address below. Based on the QQ-regression results and on the choice of $\varepsilon = 4$ we take as our benchmark value for α the one obtained using the largest 1% of the distribution, which corresponds to the best performance of the Pareto: hence, $\alpha = 3.616$. To enhance the comparability with the previous literature we also experiment using $\alpha = 4.25$ as in Melitz and Redding (2015). On the other hand, for the lognormal and the Weibull we use the estimates resulting from the full sample: in fact, the main advantage of these two distributions lies in their ability to fit the whole size distribution of firms, not only its upper tail as the Pareto. Therefore, we set $\sigma = 0.791$ and $k = 1.663$.[11]

---

[10] 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 Fréchet 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.

[11] The GFT are unaffected by the mean of the lognormal as well as the scale parameters of the Pareto and Weibull. We set $\phi_{min} = 1$ for the Pareto (as Melitz and Redding, 2015), $\mu = 0$ and $\lambda = 0.71$ (these two values are consistent with each other) for the lognormal and the Weibull.

Table 2: QQ estimation of the shape parameter for the Pareto, lognormal and Weibull distributions.

	(1) all	(2) top 50%	(3) top 25%	(4) top 5%	(5) top 4%	(6) top 3%	(7) top 2%	(8) top 1%
sample: obs.:	33120	16560	8280	1656	1325	993	662	331
Pareto:								
$1/\alpha^{sales}$	2.129*	1.382*	1.177*	0.939*	0.917*	0.894*	0.869*	0.830*
R^2	0.804	0.969	0.984	0.993	0.994	0.994	0.994	0.990
α	1.409	2.171	2.548	3.193	3.273	3.354	3.454	3.616
Lognormal:								
σ^{sales}	2.373*	2.326*	2.410*	2.527*	2.534*	2.558*	2.594*	2.643*
R^2	0.999	0.999	0.999	0.997	0.997	0.996	0.993	0.986
σ	0.791	0.775	0.803	0.842	0.845	0.853	0.865	0.881
Weibull:								
$1/k^{sales}$	1.804*	2.775*	3.338*	4.309*	4.414*	4.572*	4.798*	5.145*
R^2	0.950	0.980	0.986	0.990	0.989	0.988	0.986	0.978
k	1.663	1.081	0.899	0.696	0.680	0.656	0.625	0.583

* 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 $\varepsilon = 4$.

An additional way to assess the relative performance of the three distributions is to
compare the degree of heterogeneity in productivity implied by them to that observed
in the data. More specifically, based on the estimated parameters for the lognormal,
Weibull and Pareto distributions, we can compute measures of dispersions such as the
standard deviation, the log difference between the 75th (90th) and the 25th (10th)
percentiles (see Table 3).

Table 3: Productivity dispersion implied by the Pareto, lognormal and Weibull distributions.

	measure of dispersion		
	st. dev (1)	75–25 (2)	90–10 (3)
Pareto ($\alpha = 1.409$)	0.710	0.780	1.559
Pareto ($\alpha = 3.616$)	0.277	0.304	0.608
Pareto ($\alpha = 4.25$)	0.235	0.259	0.517
lognormal	0.791	1.067	2.027
Weibull ($k = 1.662$)	0.772	0.946	1.856

Columns (2)-(3) report the log difference between the 75th
and 25th (90th–10th) percentiles.
Theoretical values are computed based on the appropriate dis-
tribution for the logs of a Pareto, lognormal and Weibull ran-
dom variable, namely an exponential, normal and Gumbel dis-
tribution.

These notional values can then be compared to those found in the empirical liter-
ature. For instance, Syverson (2004) reports an average within-industry inter-quartile
range of around 0.66 among US manufacturing firms with an average 90th–10th per-
centile 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.

It is clear from Table 3 that the lognormal and Weibull distributions yield figures that are close to their empirical counterparts, whereas the Pareto distribution tends to underestimate the degree of productivity dispersion found in the data, especially with the two largest values of the shape parameter. The only case in which the Pareto displays a performance that is comparable to the other two distributions is when we use the value of the shape parameter $\alpha = 1.409$, found when fitting a Pareto to the entire firm size distribution. This however, provides a very poor fit (see column 1 in Table 7). Hence, given that $\alpha = 1.409$ is also very far from the values of the shape parameter normally adopted in the literature, we do not include this value in the subsequent analysis.

3.3 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 as found in the data for French exporters. In particular, the baseline value of variable trade costs (τ) is calibrated to yield the observed export intensity, i.e. the average value of export on sales. More precisely, under the assumptions of CES preferences and two symmetric countries one can derive the following relationship:

$$\tau^{1-\varepsilon}/(1 + \tau^{1-\varepsilon}) = \text{export intensity (export sales / total sales)}.$$

Setting export intensity equal to 21.62% (the rate observed among French firms in 2003) the benchmark value for τ is 1.54.

The fixed costs associated with entry (f_e) and exporting (f_x) are chosen to yield an export participation rate equal to 21.58% (again, the observed rate among French manufacturing firms). 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: in other words, 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. As it will become apparent below, this change generates very large differences in the estimated GFT: we treat these two scenarios (high and low entry rates) as polar cases, and we consider them as providing reasonable upper and lower bounds. Arguably, the dependence on the entry rate —

whose precise value is difficult to calibrate— may be seen either as a major drawback associated with the departure from the Pareto distribution, or as offering an additional degree of flexibility.[12]

Finally, the wage rate is normalized to 1, as in 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. Table 4 summarizes the benchmark values of the key parameters used throughout the paper.

Table 4: Benchmark values for key coefficients used in the simulations

parameter	benchmark	alternative
elasticity of substitution (ε)	4	
fixed production costs* (f_d)	1	
variable trade cost (τ)	1.54	
entry rate	0.55%	50%
Pareto		
scale* (ϕ_{min})	1	
shape (α)	3.616	4.250
sunk entry cost[†] (f_e)	2.952	1.411
fixed export costs[†] (f_x)	0.982	0.813
Lognormal		
mean* (μ)	0	
standard deviation (σ)	0.791	
sunk entry cost[‡] (f_e)	0.014	19.507
fixed export costs[‡] (f_x)	0.894	5.191
Weibull		
scale* (λ)	0.710	
shape (k)	1.663	
sunk entry cost[‡] (f_e)	0.002	2.878
fixed export costs[‡] (f_x)	0.439	2.263

∗ irrelevant for computation of GFT.
† values change with α; ‡ values change with entry rate.

4 Gains from trade

This section illustrates the results of a series of quantitative evaluations of GFT under the three distributions under analysis assuming the system moves from autarky to a trade equilibrium. The section is divided in four parts: in Section 4.1 we quantify the GFT associated with varying degree of variable trade costs, while Section 4.2 focuses on the variation in GFT when fixed costs change. In Section 4.3 we evaluate the robustness of the results to changes in the elasticity of substitution, whose value is difficult to pin down. Last, in Section 4.4 we quantify the additional benefits associated with firm heterogeneity relative to a model where all firms are considered equal.

---

[12] We thank an anonymous referee for pointing this out.

4.1 Impact of a reduction in variable trade costs

We start our analysis by adopting a strategy similar to the one used by Melitz and Redding (2015): we estimate the GFT resulting from a trade liberalization that moves two symmetric countries from autarky to a trade equilibrium that mimics the rates of export participation and export intensity observed among French firms (both equal to 21.6% in 2003). Table 5 presents the relevant results for different distributions and, in the lognormal and Weibull cases, also for different entry rates.

Table 5: 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 $\tau = 1.54$			$GFT(\tau)$: heterog.	
		heterog. (1)	homog. (2)	Δ (3)	range (4)	st.dev (5)
Pareto	($\alpha = 3.616$)	5.47	3.00	2.47	20.19	5.53
Pareto	($\alpha = 4.25$)	3.88	0.97	2.91	18.34	4.95
lognormal	(entry: 0.55%)	3.16	0	3.16	17.30	4.65
lognormal	(entry: 50%)	6.51	3.88	2.63	21.57	5.98
Weibull	(entry: 0.55%)	0.66	0	0.66	15.47	4.18
Weibull	(entry: 50%)	2.76	0	2.76	16.41	4.48

Columns 1–2 report welfare gains of a move from autarky to a trading equilibrium characterized by $\tau = 1.54$ 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 between 1 and 2.5, while column 5 reports the associated standard deviation.

There are substantial differences between the various distributions or even for different parameterizations of each distribution. For instance, column 1 shows that increasing the value of the shape parameter for the Pareto distribution from 3.616 to 4.25 reduces the estimated welfare effect of a trade liberalization by more than 40%. 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 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 distribution for firm productivity (i.e., abandoning the power-law assumption) can yield larger or smaller GFT depending on the entry rate and the associated magnitude of sunk costs. Given the same entry rate, the distance between the lognormal and the Weibull GFT tend to be larger than the difference between Pareto and lognormal, suggesting that the specific distributional assumption one starts with is indeed crucial in the quantification of GFT. 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). 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 always yield the same fraction of exporting firms, namely 21.6%.
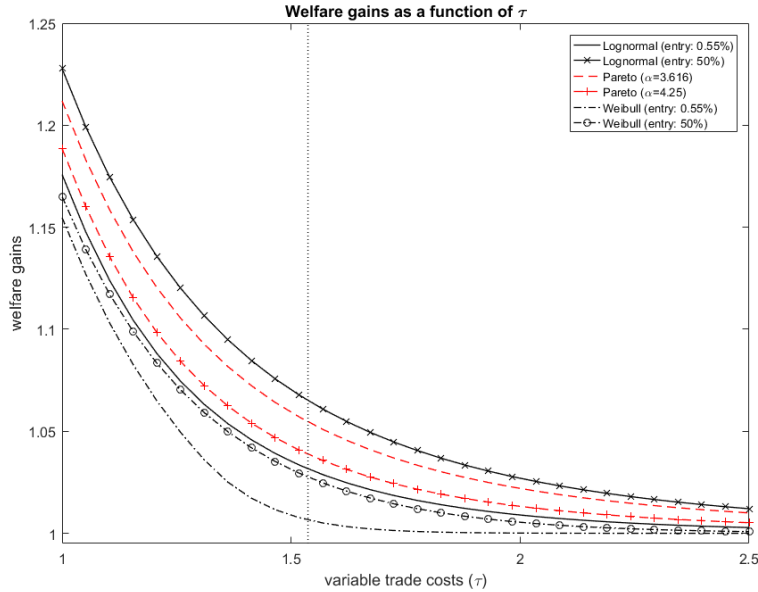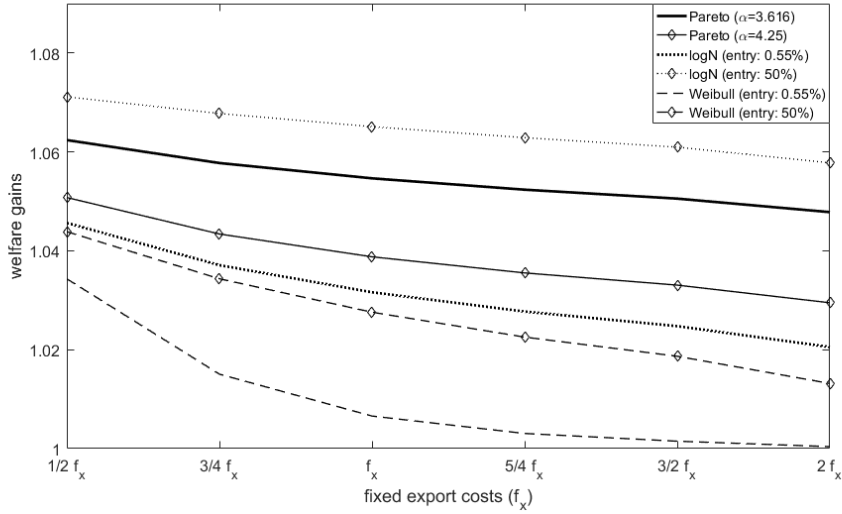
Fig. 1: Welfare gains of a trade liberalization as a function of τ: Pareto, lognormal and Weibull cases.

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 reference point.

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%: in fact, 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.

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 vary substantially across the different distributions. This is particularly true for the Weibull distribution which, although it provides a fit of the entire distribution of firm size that is only marginally worse than the lognormal, delivers much smaller welfare gains. Hence, the specific distributional assumption underlying the analysis has a strong bearing on the results and sticking to a single productivity distribution may lead to fragile conclusions in terms of policy implications.

Fig. 2: Welfare gains of a trade liberalization as a function of f_x: Pareto, lognormal and Weibull cases. Note that the actual levels of f_x changes across distributions (see Table 4 for the benchmark values).

4.2 Impact of a reduction in fixed trade costs

Having considered the impact of a reduction in variable trade costs, we now turn to analyze the effect of a movement from autarky to a series of trade equilibria characterized by different *fixed* export costs. We hypothesize six different levels of f_x, while keeping the other cost parameters (f_e and τ, as well as f_d) fixed at their benchmark values. The baseline for f_x is the value associated with an export participation of 21.6%: in this case, the GFT are the same reported in column (1) of Table 5, as all the parameters are identical. We analyze the welfare effect of departing from this value by looking at the GFT when the trade equilibrium features fixed costs that are 25% or 50% lower, 25%, 50% or 100% higher than the baseline value.

Figure 2 provides a graphical representation of the estimated GFT. As expected, a trade equilibrium characterized by higher fixed export costs brings about smaller welfare gains. Yet, the sensitivity of the GFT to changes in f_x is very different: low for the Pareto and lognormal cases, relatively high for the Weibull, especially in the baseline case of low entry. Table 6 reports the detailed numbers associated with the simulations (panel A).[13] The rows of the table refers to different distributions, while the columns identify the value of fixed export costs in the trade equilibrium: for each case we compute the welfare effect of the notional liberalization, the difference with respect to the benchmark case, and the rate of export participation.[14] So, for instance, if one assumes that productivity follows a Pareto distributions with a shape parameter

---

[13] Panel B of Table 6 presents the results in the case all firms are considered equal. We will discuss this case in Section 4.4 below.

[14] For the sake of simplicity we only discuss results obtained assuming an entry rate equal to 0.55%. Results for entry equal to 50% are available upon request.

$\alpha = 3.616$, the baseline case (column 3) yields a welfare effect of 5.466% and an export participation of 21.6%; if the trade equilibrium were associated with a level of fixed export costs twice as large (as in column 6), then export participation would drop to 9.37% and the associated GFT decrease to 4.783%.

Table 6: 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)	$^{3}/_{2}f_x$ (5)	$2f_x$ (6)
		panel A: heterogeneous model						
Pareto	$k = 3.616$	welfare gains	6.240	5.776	5.466	5.237	5.056	4.783
		diff. wrt col (3)	[0.77]	[0.31]		[-0.23]	[-0.41]	[-0.68]
		export partic.	*49.81*	*30.55*	*21.60*	*16.51*	*13.25*	*9.37*
Pareto	$k = 4.25$	welfare gains	5.078	4.340	3.879	3.553	3.306	2.950
		diff. wrt col (3)	[1.20]	[0.46]		[-0.33]	[-0.57]	[-0.93]
		export partic.	*57.67*	*32.47*	*21.60*	*15.75*	*12.16*	*8.09*
lognormal	$\sigma = 0.791$	welfare gains	4.564	3.710	3.162	2.772	2.477	2.054
		diff. wrt col (3)	[1.40]	[0.55]		[-0.39]	[-0.68]	[-1.11]
		export partic.	*54.57*	*32.01*	*21.60*	*15.79*	*12.15*	*7.95*
Weibull	$\lambda = 1.663$	welfare gains	3.433	1.506	0.658	0.306	0.150	0.040
		diff. wrt col (3)	[2.77]	[0.85]		[-0.35]	[-0.51]	[-0.62]
		export partic.	*100*	*57.82*	*21.60*	*9.03*	*4.07*	*0.97*
		panel B: homogeneous model						
Pareto	$k = 3.616$	welfare gains	5.583	4.258	2.997	1.795	0.648	0
		diff. wrt col (3)	[2.59]	[1.26]		[-1.20]	[-2.35]	[-3.00]
Pareto	$k = 4.25$	welfare gains	4.444	2.649	0.972	0	0	0
		diff. wrt col (3)	[3.47]	[1.68]		[-0.97]	[-0.97]	[-0.97]
lognormal	$\sigma = 0.791$	welfare gains	3.827	1.789	0	0	0	0
		diff. wrt col (3)	[3.83]	[1.79]		[–]	[–]	[–]
Weibull	$\lambda = 1.663$	welfare gains	3.323	1.094	0	0	0	0
		diff. wrt col (3)	[3.32]	[1.09]		[–]	[–]	[–]

f_x is the calibrated value of fixed export costs that guarantees an export participation rate of 21.6%. 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).

Table 6 shows that heavier tails in the distribution of productivity are associated with (*ceteris paribus*) larger welfare effects and with a lower sensitivity of GFT to changes in f_x. Also, fatter tails imply that sizable welfare gains can be obtained already from small reductions in f_x (i.e., moving from autarky to trade equilibria characterized by high fixed costs, as in columns 4–6). 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: reducing this parameter by 25% from its baseline value moves the fraction of exporting firms from 21.6% to almost 60%, more than doubling the resulting welfare effect, which jumps from 0.658% to 1.506%. On the other hand, for a Pareto distribution with a shape parameter $\alpha = 3.616$, the same reduction in fixed export costs (-25%) moves export participation to 30.55% and increases the GFT by 0.31% only.

Similar effects (in the opposite direction) are associated with values of f_x higher than the baseline case (see columns 4–6). 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 participation in column 6 is largest in the case of a Pareto with $k = 3.616$, the distribution with the heaviest tail of all our candidates. 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 intuition behind the results is reminiscent of the argument made by Chaney (2008) to discuss the sensitivity of the extensive margin to trade costs. As trade barriers decrease, some low-productivity firms enter the export market: when the productivity distribution has heavy tails, these new exporters will be small relative to incumbents, so that the extra export (and the additional gains) generated by a reduction in f_x is small. However, heavy tails imply that —starting from autarky— even small reduction in f_x can yield (relatively) large benefits because the few firms that manage to start exporting are very productive/large and there are more of them (compared with a distribution with thinner tails).

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 magnitude of the welfare benefits depends crucially on export participation, and this in turn hinges upon the degree of tail heaviness in the productivity distribution. In this respect, since the tail behavior of the Pareto and the lognormal can be very similar (Perline, 2005; Malevergne et al., 2011), the inclusion of the Weibull as an additional case study makes the relationship between fixed export costs, export participation and GFT much more evident.

4.3 Sensitivity to variations in the elasticity of substitution

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. One question that follows directly from the above results has to do with the calibration of the elasticity of substitution (ε): in fact, its value 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.[15]

---

[15] 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)."
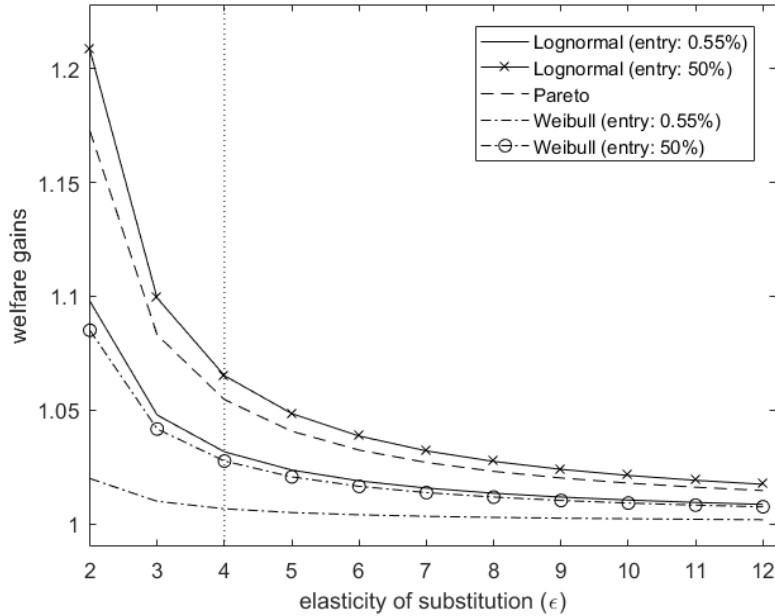
Fig. 3: Welfare gains of trade liberalization as a function of ε: α, k and σ vary with ε; moreover, we also adjust τ and f_x in order to maintain constant the average export intensity and export participation. The vertical dotted line marks the benchmark value $\varepsilon = 4$.

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. Figure 3 compares GFT under the different distributions when we let ε vary between 2 and 12 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 (14)–(16) that link them to the associated productivity distribution. Fixed (f_e and f_x) and variable (τ) costs are adjusted in order to keep both export participation and average export intensity at their empirical value (21.6%). This adjustment is necessary in order to properly compare trade equilibria characterized by the same outcome in terms of extensive and intensive margins. What changes, as we modify ε and the degree of firm heterogeneity, is the share of exports carried out by each firm, or, in other words, export concentration.[16] It is also interesting to note that in the Pareto case, f_x does not vary with the elasticity of substitution ε, while it varies when the productivity distribution follows either a lognormal or a Weibull.

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, especially under the Pareto and lognormal distribution. Indeed, GFT range between 1.5% and 17.3% when the elasticity of substitution moves from 12 to 2 in the Pareto

---

[16] 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.

case, and between 0.9% (1.7) to 9.8% (20.8%) in the lognormal one with low (high) entry rate. The lognormal and the Pareto show similar degrees of sensitivity to ε when one assumes a high entry rate. This sensitivity represents a major limitation, 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.

The analysis performed in this section conveys two main messages. First of all, results confirm that the degree of tail heaviness of the productivity distribution has a strong impact on the magnitude of the GFT. Given that the elasticity of substitution links the parameters of the sales and productivity distributions, a reduction in ε implies heavier tails: as a result, trade liberalization affects more firms and these tend to be larger, giving rise to a larger welfare effect. Second, distributions that feature (on average) heavier tails tend to be more sensitive to changes in the values of the elasticity of substitution, whose precise value is difficult to determine. Hence, heavier tails imply not only larger GFT but also more uncertainty in their actual magnitude.
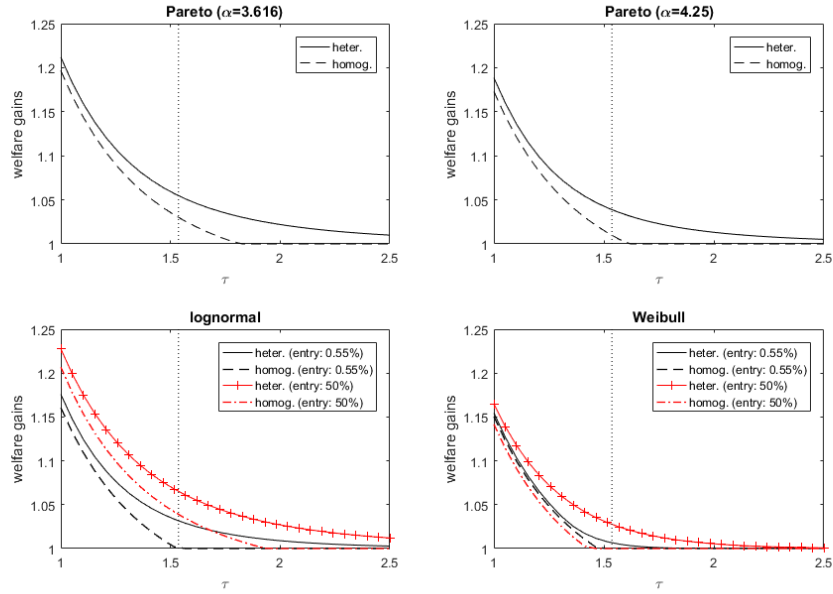
4.4 Homogeneous vs. heterogeneous models

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 by Arkolakis et al. (2012) and Melitz and Redding (2015).

Table 5 (column 3) shows that the *additional* welfare effect of heterogeneity can be substantial, but depends on the specific distribution. It ranges from a meager 0.66% for the Weibull with an entry rate of 0.55% to 2.91% in the Pareto case with $\alpha = 4.25$, and 3.16% for a lognormal distribution.

Figure 4 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 ($\tau \to 1$), 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 panel B of Table 6, we see that the homogeneous model tends to be 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. Therefore, either a liberalization triggers export by all firms, or it fails to generate any export at all.

Lastly, we compare the heterogeneous and homogeneous models when ε is allowed to vary between 2 and 12 (see Figure 5). For the Pareto case, the difference between the two models is slightly larger for small values of ε, but remains roughly constant
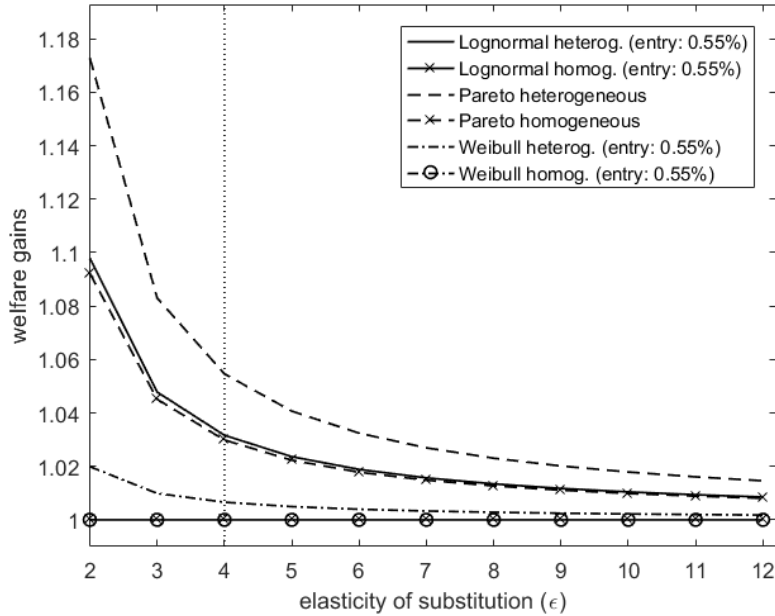
Fig. 4: Welfare gains of trade liberalization as a function of τ: Heterogeneous vs. homogeneous models.

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 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 specific trade equilibrium. While this is the traditional way welfare effects are measured in the literature, autarky is not a very relevant benchmark, as it is seldom observed in reality. If, as we argue in the Introduction, an important feature of new-new trade models 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. 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 welfare in two different trading equilibria, rather than looking at the effect of switching from autarky to trade. This exercise answers a more immediate policy question, since the estimated welfare impact of a trade agreement often captures

Fig. 5: 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. In the lognormal and Weibull homogeneous cases none of the firms exports, the associated welfare gains are constantly equal to 1 and the two lines are indistinguishable in the figure.

a lot of attention.[17] A case in point is a proposed US-EU trade agreement that has been under negotiations for years (before losing momentum in the second half of 2016), and 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 discussed above, GFT are very sensitive to this parameter. The results are presented in Figure 6.

The largest GFT occur under the assumptions of a lognormal productivity distribution with a large entry rate, and of a Pareto distribution, in line with what we have seen in the previous sections. For the other cases, however, the ranking is not stable as ε grows larger. In particular, the system displays a peculiar behavior when the productivity distribution follows a Weibull distribution and the entry rate is low (0.55%). This is because beyond a certain value of the elasticity of substitution, the condition that determines selection into exporting $\left(\tau(f_x/f_d)^{\frac{1}{\varepsilon-1}} > 1\right)$ no longer holds and export

---

[17] 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.
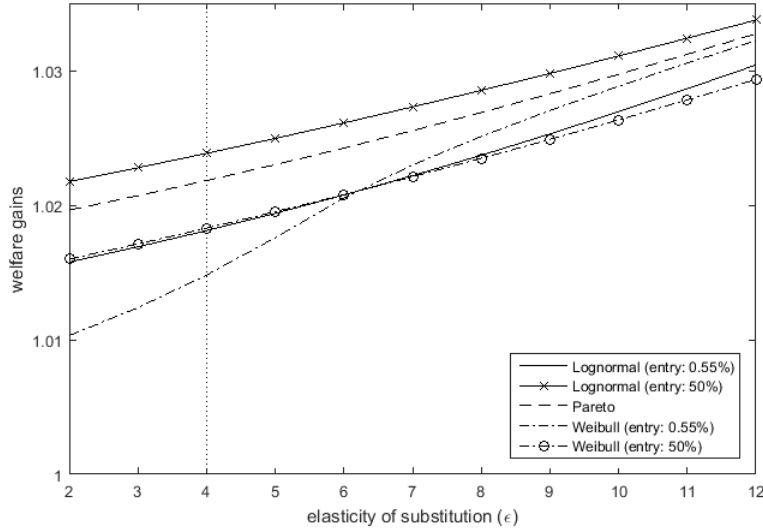
Fig. 6: Welfare effect of a 10% reduction in iceberg transport costs (τ); the starting point is a trade equilibrium featuring an export participation and intensity both equal to 21.6%.

participation jumps to 100%: with the parametrization we use, this happens already for $\varepsilon \geq 6$. In fact, under a Weibull distribution, export participation appears very sensitive to variations in variable trade costs: a 10% reduction yields a large increase in the number of exporting firms for each value of the elasticity of substitution. For instance, when $\varepsilon = 2$ the reduction in τ moves export participation from the baseline 21.6% to 31.4%, the largest increase among the different distributional assumptions we have explored. The share of exporters moves to 63% and 100% if we assume $\varepsilon = 4$ or $\varepsilon = 6$. All in all, we observe once again substantial variation both across different distributions and values of the elasticity of substitution.

As ε grows larger, the new trade equilibrium that is obtained after the reduction in variable trade costs features both higher export participation and higher export intensity. This explains both the positive relationship between welfare gains and the elasticity of substitution, and also the difference with respect to Figure 3, where welfare gains are decreasing in ε. In fact, while in the previous simulations we compare trade equilibria that are characterized by the same degree of export participation and intensity (relative to a notional autarky situation where the specific productivity distribution is irrelevant), in this case the starting point is the same under all assumptions, but the final equilibrium differs across distributions. As mentioned, larger values of ε are associated with more export intensity and export participation, resulting in larger welfare gains of a reduction in variable trade costs.

What about a reduction in fixed export costs? This question is particularly interesting as non-tariff barriers represent the bulk of obstacles still hampering trade, and the focus of most recent trade negotiations. In the EU-US case for instance, most analyses have estimated that a successful deal would cut non-tariff barriers by something be-

tween 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 21.6%, and evaluate this impact for various levels of the elasticity of substitution.
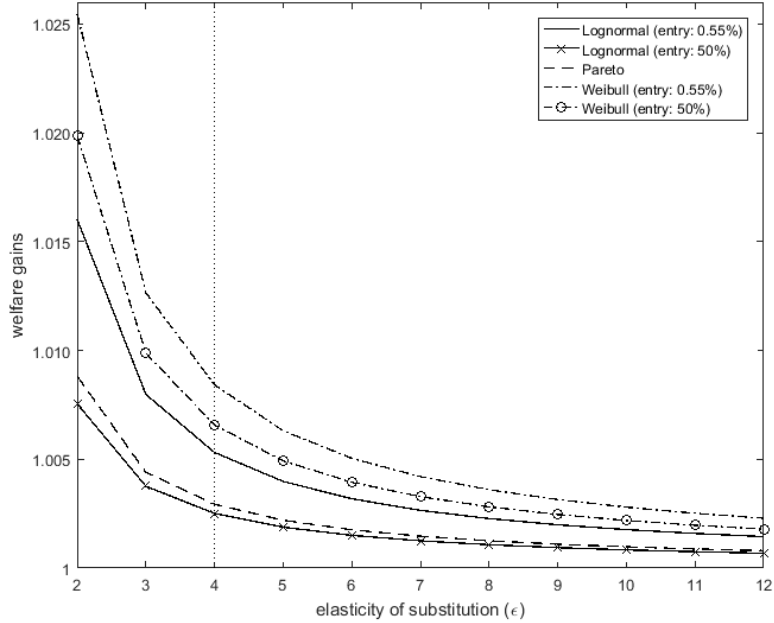
Fig. 7: Welfare effect of a 25% reduction in fixed export costs (f_x); the starting point is a trade equilibrium featuring an export participation and intensity both equal to 21.6%.

The results are summarized in Figure 7. In this case, we find the usual negative correlation between the values of the elasticity of substitution (ε) and the level of the welfare effect. The main mechanism driving this result lies in the higher survival rate among small, inefficient firms that we observe as ε grows larger (so that the productivity distribution features a slimmer right tail). In fact, the larger the number of small firms that are active in the market, the smaller the welfare benefit from an increase in the extensive margin of export (triggered by the reduction in f_x), because many of the new entrants are very small relative to the top exporters. This mechanism also explains why, in the Weibull and lognormal cases, a higher entry rate is associated with lower GFT (everything else equal): since higher entry implies less selection in the domestic market (i.e. a lower domestic productivity cutoff), a decrease in f_x yields lower welfare gains.[18]

---

[18] It is worth noting that the final share of exporting firms (for a given distribution) and the initial calibrated value of f_x are both invariant to ε and therefore these channels are not operating. Indeed, the interplay between τ, ε and the dispersion parameter of each distribution (being it σ, α or k) keeps f_x constant (again, for a given distribution) when ε increases. This behavior is not limited to the Pareto case, where one can derive the result analytically, but holds also for the Weibull and the lognormal distributions, although in these cases we have to rely on numerical results.

This interpretation is consistent with the one provided by di Giovanni and Levchenko (2013), who explain that thicker tails in the productivity distribution imply a bigger difference between the largest exporters and the marginal firm that starts exporting after a reduction in fixed costs. Hence, in this specific example, countries with a more heterogeneous distribution of firm size (a larger share of big firms) 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).

An additional reading of Figure 7 concerns the ranking of welfare effects *across* distributions (for each value of ε). In this case (and opposite to what we have often seen so far), the largest GFT are found under the assumption of a Weibull distribution, which is more sensitive to variations in fixed costs, as already discussed in Section 4.2. In fact, the rate of export participation in the new equilibrium that results from the reduction in f_x is 58% (up from the initial value of 21.6%) in the Weibull case, whereas it stops at 32% and 30.6% respectively under a lognormal or Pareto distribution. The important role of the extensive margin in determining the GFT is in line with the recent results by Bas et al. (2015), who find that trade elasticities and welfare gains are country-pair specific. Moreover, they claim that a trade liberalization with respect to a "more difficult" destination market is likely to yield larger welfare effects: higher trade barriers imply more homogeneity among exporting firms (the productivity threshold for entry into export will be larger) and therefore there will be a larger trade-promoting effect of any reduction in trade costs due to the extensive margin.

More generally, Figure 7 shows the dramatic difference in GFT across distributions: if one assumes a Pareto distribution, then the results are negligible (below 1%, often by a large margin) across the whole range of ε, and the same applies under a lognormal if the entry rate is supposed to be large. On the contrary, in the Weibull case, the estimated gains from a reduction in fixed export costs can be at least twice as large as in the Pareto case, ranging between 0.2% and 2.5%. What is more, the different rate of firm entry does not seem to play a very relevant role.

The results in this section 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 trade), we still find important differences in the GFT across distributions. Furthermore, the relative impact of a change in the fixed rather than the variable trade costs changes dramatically. Once again, the different tail behavior of the Weibull distribution relative to both the Pareto and the lognormal magnifies the difference in the results across the various scenarios and reinforce the key massage of the paper: relying on a single distributional assumption is risky, it may lead to very fragile conclusions, and should be avoided when one aims at deriving policy implications.

6 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 in describing 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; Head et al., 2014; Bas et al., 2015). 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 models 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, 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. Although the welfare gains of new trade models may appear "elusive", as Behrens et al. (2014) have recently put it, we think that models featuring heterogeneous firms at least make explicit the different assumptions that are needed to derive GFT and the associated policy implications. In this respect, a deeper understanding of the role played by each of these hypotheses and the robustness of results to departures from them is both necessary and welcome. In particular,

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.

Bas, M., Mayer, T. and Thoenig, M. (2015), From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity, Discussion Papers 10637, CEPR.

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.

Behrens, K., Kanemoto, Y. and Murata, Y. (2014), New trade models, elusive welfare gains, Discussion Papers 10255, CEPR.

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.

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.

Chaney, T. (2008), 'Distorted gravity: The intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–1721.

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.

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.

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

Gabaix, X. (2009), 'Power laws in economics and finance', *Annual Review of Economics* **1**, 255–93.

Gibrat, R. (1931), *Les Inegalites Economiques*, Sirey, Paris.

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

Growiec, J. (2013), 'A microfoundation for normalized {CES} production functions with factor-augmenting technical change', *Journal of Economic Dynamics and Control* **37**(11), 2336 – 2350.

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. (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.

McCool, J. I. (2012), *Using the Weibull Distribution. Reliability, Modeling, and Inference*, Wiley, New Jersey.

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.

Mrázová, M. and Neary, J. P. (2014), 'Together at Last: Trade Costs, Demand Structure, and Welfare', *American Economic Review* **104**(5), 298–303.

Perline, R. (2005), 'Weak and false inverse power laws', *Statistical Science* **20**, 68–88.

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

Distribution Fitting

Table 7: QQ estimation of the shape parameter for the Pareto, lognormal and Weibull distributions. Alternative years and destinations.

sample:	(1) all	(2) top 50%	(3) top 25%	(4) top 5%	(5) top 4%	(6) top 3%	(7) top 2%	(8) top 1%
panel (a): **Germany 2000**								
Pareto:								
$1/\alpha^{sales}$	2.470*	1.610*	1.305*	0.923*	0.887*	0.848*	0.806*	0.783*
R^2	0.823	0.951	0.968	0.986	0.989	0.992	0.996	0.994
α	1.214	1.863	2.299	3.25	3.383	3.538	3.723	3.832
Lognormal:								
σ^{sales}	2.722*	2.736*	2.694*	2.493*	2.460*	2.429*	2.407*	2.501*
R^2	0.999	1	0.999	0.998	0.998	0.998	0.997	0.996
σ	0.907	0.912	0.898	0.831	0.82	0.81	0.802	0.834
Weibull:								
$1/k^{sales}$	2.049*	3.281*	3.751*	4.262*	4.294*	4.348*	4.453*	4.877*
R^2	0.931	0.99	0.996	0.996	0.995	0.993	0.99	0.991
k	1.464	0.914	0.8	0.704	0.699	0.69	0.674	0.615
obs	30063	15032	7516	1503	1202	902	601	300
panel (b): **ITALY 2003**								
Pareto:								
$1/\alpha^{sales}$	2.401*	1.574*	1.270*	0.904*	0.879*	0.853*	0.819*	0.787*
R^2	0.825	0.95	0.967	0.991	0.993	0.993	0.994	0.991
α	1.249	1.906	2.363	3.317	3.412	3.518	3.661	3.81
Lognormal:								
σ^{sales}	2.641*	2.677*	2.622*	2.435*	2.433*	2.440*	2.447*	2.515*
R^2	0.999	0.999	0.999	0.999	0.999	0.998	0.997	0.995
σ	0.88	0.892	0.874	0.812	0.811	0.813	0.816	0.838
Weibull:								
$1/k^{sales}$	1.987*	3.211*	3.652*	4.156*	4.241*	4.365*	4.527*	4.906*
R^2	0.93	0.991	0.996	0.994	0.993	0.993	0.991	0.992
k	1.51	0.934	0.822	0.722	0.707	0.687	0.663	0.612
obs	23243	11622	5811	1162	930	697	465	232
panel (c): **UK 2006**								
Pareto:								
$1/\alpha^{sales}$	2.419*	1.589*	1.301*	0.910*	0.877*	0.833*	0.779*	0.750*
R^2	0.824	0.954	0.966	0.984	0.985	0.988	0.994	0.993
α	1.24	1.888	2.306	3.298	3.423	3.6	3.852	3.998
Lognormal:								
σ^{sales}	2.663*	2.697*	2.688*	2.456*	2.432*	2.387*	2.322*	2.385*
R^2	1	0.999	0.998	0.997	0.996	0.995	0.994	0.989
σ	0.888	0.899	0.896	0.819	0.811	0.796	0.774	0.795
Weibull:								
$1/k^{sales}$	2.006*	3.230*	3.744*	4.198*	4.246*	4.274*	4.293*	4.636*
R^2	0.932	0.988	0.996	0.996	0.995	0.992	0.987	0.981
k	1.496	0.929	0.801	0.715	0.707	0.702	0.699	0.647
obs	21277	10639	5319	1064	851	638	425	213

* indicates significance at the 1% level

α^{sales}, σ^{sales} and k^{sales} represent estimates from the distribution of export sales to each destination; the corresponding parameters for the distribution of productivity (α, σ and k) are obtained assuming an elasticity of substitution $\varepsilon = 4$.