Welfare and Trade Without Pareto

By Keith Head, Thierry Mayer and Mathias Thoenig *

Heterogeneous firm papers that need parametric distributions—most of the literature following Melitz (2003)—use the Pareto distribution. The use of this distribution allows a large set of heterogeneous firms models to deliver the simple gains from trade (GFT) formula developed by Arkolakis, Costinot and Rodriguez-Clare (2012) (hereafter, ACR). This implication is closely tied to the fact that Pareto allows for a constant elasticity of substitution import system.

Three important criteria have motivated researchers to select the Pareto distribution for heterogeneity. The first is tractability. Assuming Pareto makes it relatively easy to derive aggregate properties in an analytical model. Users of the Pareto distribution also justify it on empirical and theoretical grounds. For example, ACR argue that the Pareto provides “a reasonable approximation for the right tail of the observed distribution of firm sizes” and is “consistent with simple stochastic processes for firm-level growth, entry, and exit...”

This paper investigates the consequences of replacing the assumption of Pareto heterogeneity with log-normal heterogeneity. This case is interesting because it (a) maintains some desirable analytic features of Pareto, (b) fits the complete distribution of firm sales rather than just approximating the right tail, and (c) can be generated under equally plausible processes (see online appendix). The log-normal is reasonably tractable but its use sacrifices some “scale-free” properties conveyed by the Pareto distribution. Aspects of the calibration that do not matter under Pareto lead to important differences in the gains from trade under log-normal.

I. Welfare Theory

We assume CES monopolistic competition with a representative worker of country \( i \) endowed with \( L_i \) efficiency units, paid wages \( w_i \), and facing price index \( P_i \). As shown in the appendix, welfare (defined by real income) is given by

\[
W_i \equiv \frac{w_iL_i}{P_i} = \left( \frac{L}{\sigma f_{ii}^{1/\sigma}} \right)^{\frac{1}{\sigma - 1}} \left( \frac{\sigma - 1}{\tau_{ii} \alpha_{ii}^*} \right)^{\frac{1}{\sigma - 1}},
\]

where \( \alpha_{ii}^* \), \( \tau_{ii} \) and \( f_{ii} \) denote the internal zero-profit cost, trade cost, and fixed production cost.

Following a change in international trade costs, welfare varies according to changes in the only endogenous variable in (1), \( \alpha_{ii}^* \):

\[
\frac{dW_i}{W_i} = -\frac{d\alpha_{ii}^*}{\alpha_{ii}^*} = \frac{1}{\epsilon_{ii}} \left( \frac{d\pi_{ii}}{\pi_{ii}} - \frac{dM_i^e}{M_i^e} \right).
\]

Changes in welfare depend on changes in the domestic trade share, \( \pi_{ii} \), and in the mass of domestic entrants, \( M_i^e \). Both effects are stronger when the partial trade elasticity, \( \epsilon_{ii} \), that affects internal trade is small.

The result in (2) that marginal changes in welfare mirror changes in the domestic cost cutoff focuses our attention on the role of selection. Assuming that successful entry in the domestic market is prevalent, it is the left tail of the distribution that is crucial for welfare. This is the part

---

* Head: University of British Columbia, Sauder School of Business, 2053 Main Mall, Vancouver, B.C., V6T 1Z2, Canada, and CEPR, (email: keith.head@sauder.ubc.ca). Mayer: Sciences-Po, 28 rue des Saints-Pères, 75007 Paris, France, CEPII, and CEPR, (email: thierry.mayer@sciencespo.fr). Thoenig: Faculty of Business and Economics, University of Lausanne, Bâtiment Extranef, 1015 Lausanne, Switzerland, and CEPR, (email: mathias.thoenig@unil.ch). This research has received funding from the European Research Council (FP7/2007-2013) Grant Agreement No. 313522. We thank Maria Bas and Céline Poilly for their help with data, Jonathan Eaton, Andres Rodrigues-Clare, and Arnaud Costinot for valuable insights, Marc Melitz and Stephen Redding for sharing code, and the Douanes Françaises for data.

1Two papers remove the long fat tail of the standard Pareto by bounding productivity from above. The first, Helpman, Melitz and Rubinstein (2008), shows that this leads to variable trade elasticities. The more recent, Feenstra (2013), shows how double truncated Pareto changes the analysis of pro-competitive effects of trade.

2By “partial” we mean that incomes and price indices are held constant as in a gravity equation estimated with origin and destination fixed effects.
of the distribution where Pareto and log-normal differ most strikingly.

Shifting to the last equality in (2), welfare falls with the domestic market share since \( \epsilon_{ii} < 0 \) but it is increasing in the mass of entrants. Under Pareto, \( \epsilon_{ni} = \epsilon \), a constant across country pairs, which implies \( dM \sim 0 \). This means we can integrate marginal changes to obtain the simple welfare formula of ACR, where \( \bar{\Omega} \equiv \frac{\alpha_{d}^{1/\epsilon}}{\tau} \), where “hats” denote total changes. The log-mean parameter \( \bar{\nu} \) is defined as

\[
\bar{\nu} = -\frac{1}{\alpha} \ln \frac{1}{\alpha}.
\]

where we use \( \Phi \) to denote the CDF of the standard normal. The equations needed for the quantification of the gains from trade are therefore (3) and (4), which provide \( \alpha_{d}^{*} \) conditional on \( G(\alpha_{d}^{*}) \), itself defined by (5).

**A. The 4 key moments**

There are four moments that are crucial in order to calibrate the unknown parameters of the two-country model.

**M1:** The share of firms that pay the sunk cost and successfully enter, \( G(\alpha_{d}^{*}) \) in the model. Since the number of firms that pay the entry cost but exit immediately is not observable, **M1** is a challenge to calibrate. We show in the appendix that under Pareto, the GFT calculation is invariant to **M1**. Unfortunately, **M1** matters under log-normal, so our sensitivity analysis considers a range of values.

**M2:** The share of firms that are successful exporters, \( G(\alpha_{e}^{*})/G(\alpha_{d}^{*}) \) in the model. The target value for **M2** is 0.18, based on export rates of US firms reported by [Melitz and Redding (2013)]#cite:melitz2013.

**M3** is the data moment used to calibrate the firm’s heterogeneity parameter: \( \theta \) in Pareto and \( \nu \) in log-normal. There are two alternative moments that the model links closely to the heterogeneity parameters. The first, which we refer to as **M3**, is an estimate derived from the distribution of firm-level sales (exports) in some market: the micro-data approach, on which we concentrate in the main text. The second, which we call **M3’** is the trade elasticity \( \epsilon_{e}^{*} \): the macro-data approach, covered in the appendix.

**M4:** The share of export value in the total sales of exporters. Using CES and symmetry, **M4** sets

\[
\text{The distribution where Pareto and log-normal differ most strikingly.}
\]

Shifting to the last equality in (2), welfare falls with the domestic market share since \( \epsilon_{ii} < 0 \) but it is increasing in the mass of entrants. Under Pareto, \( \epsilon_{ni} = \epsilon \), a constant across country pairs, which implies \( dM \sim 0 \). This means we can integrate marginal changes to obtain the simple welfare formula of ACR, where \( \bar{\Omega} \equiv \frac{\alpha_{d}^{1/\epsilon}}{\tau} \), where “hats” denote total changes. The log-mean parameter \( \bar{\nu} \) is defined as

\[
\bar{\nu} = -\frac{1}{\alpha} \ln \frac{1}{\alpha}.
\]

where we use \( \Phi \) to denote the CDF of the standard normal. The equations needed for the quantification of the gains from trade are therefore (3) and (4), which provide \( \alpha_{d}^{*} \) conditional on \( G(\alpha_{d}^{*}) \), itself defined by (5).

**A. The 4 key moments**

There are four moments that are crucial in order to calibrate the unknown parameters of the two-country model.

**M1:** The share of firms that pay the sunk cost and successfully enter, \( G(\alpha_{d}^{*}) \) in the model. Since the number of firms that pay the entry cost but exit immediately is not observable, **M1** is a challenge to calibrate. We show in the appendix that under Pareto, the GFT calculation is invariant to **M1**. Unfortunately, **M1** matters under log-normal, so our sensitivity analysis considers a range of values.

**M2:** The share of firms that are successful exporters, \( G(\alpha_{e}^{*})/G(\alpha_{d}^{*}) \) in the model. The target value for **M2** is 0.18, based on export rates of US firms reported by [Melitz and Redding (2013)]#cite:melitz2013.

**M3** is the data moment used to calibrate the firm’s heterogeneity parameter: \( \theta \) in Pareto and \( \nu \) in log-normal. There are two alternative moments that the model links closely to the heterogeneity parameters. The first, which we refer to as **M3**, is an estimate derived from the distribution of firm-level sales (exports) in some market: the micro-data approach, on which we concentrate in the main text. The second, which we call **M3’** is the trade elasticity \( \epsilon_{e}^{*} \): the macro-data approach, covered in the appendix.

**M4:** The share of export value in the total sales of exporters. Using CES and symmetry, **M4** sets
the benchmark trade cost $\tau_0$. Indeed, $M4 = \frac{1 - \tau_0}{1 + \tau_0}$, which Melitz and Redding (2013) take as 0.14 from US exporter data. Setting $\sigma = 4$, we have $\tau_0 = \left((1 - M4)/M4\right)^{1/3} = 1.83$.

Two parameters still need to be set: the CES $\sigma$, and the domestic fixed cost, $f_d$. We follow Melitz and Redding (2013) in setting $\sigma = 4$.

Since equations (3) and (4) imply that only relative $f_x/f_d$ matters for equilibrium cutoffs, we set $f_d = 1$.

**B. QQ estimators of shape parameters**

Each of the two primitive distributions is characterized by a location parameter ($\alpha \equiv 1/\varphi$ in Pareto or $\mu$ in log-normal) and a shape parameter ($\theta$ or $\nu$) governing heterogeneity. For the trade elasticities and GFT, location parameters do not matter whereas heterogeneity (falling with $\theta$ and rising with $\nu$) is crucial.

As comprehensive and reliable data on firm-level productivity are difficult to obtain, we instead obtain $M3$ from data on the size distribution of exports for firms from a given origin in a given destination. In so doing, we rely on the CES monopolistic competition assumption, which implies that sales of an exporter from $i$ to $n$, with cost $\alpha$ can be expressed as $x_{ni}(\alpha) = K_{ni}\alpha^{1-\sigma}$. The $K_{ni}$ factor combines all the terms that depend on origin and destination but not on the identity of the firm.

Pareto and log-normal variables share the feature that raising them to a power retains the original distribution, except for simple transformations of the parameters. Therefore, CES-MC combined with productivity distributed Pareto($\varphi, \theta$) implies that the sales of firms in any given market will be distributed Pareto($\bar{\varphi}, \bar{\theta}$), where $\bar{\theta} = \frac{\theta}{\sigma - 1}$. If $\varphi$ is log-N($\mu, \nu$) then $\varphi^{\sigma - 1}$ is log-N($\bar{\mu}, \bar{\nu}$), with $\bar{\nu} = (\sigma - 1)\nu$. Estimating $\bar{\theta}$ and $\bar{\nu}$, and postulating a value for $\sigma$, we can back out estimates of $\theta$ and $\nu$.

We estimate $1/\bar{\theta}$ and $\bar{\nu}$ by taking advantage of a linear relationship between empirical quantiles and theoretical quantiles of log sales data. Originally used for data visualization, the asymptotic properties of this method are analyzed by Kratz and Resnick (1996), who call it a QQ estimator. Dropping country subscripts for clarity, we denote sales as $x$, where $i$ now indexes firms ascending order of individual sales. Thus, $i = 1$ is the minimum sales and $i = n$ is the maximum. The empirical quantiles of the sorted log sales data are $Q_i^E = \ln x_i$ and the empirical CDF is $\hat{F}_i = (i - 0.3)/(n + 0.4)$.

The distribution of $\ln x_i$ takes an exponential form if $x_i$ is Pareto:

$$F_p(\ln x) = 1 - \exp[-\bar{\theta}(\ln x - \ln \bar{x})],$$

whereas the corresponding CDF of $\ln x_i$ under log-normal $x_i$ is normal:

$$F_{LN}(\ln x) = \Phi((\ln x - \bar{\mu})/\bar{\nu}).$$

The QQ estimator minimizes the sum of the squared errors between the theoretical and empirical quantiles. The theoretical quantiles implied by each distribution are obtained by applying the respective formulas for the inverse CDFs to the empirical CDF:

$$Q_i^p = F_p^{-1}(\hat{F}_i) = \ln \bar{x} - \frac{1}{\bar{\theta}} \ln(1 - \hat{F}_i),$$

$$Q_i^{LN} = F_{LN}^{-1}(\hat{F}_i) = \bar{\mu}/\bar{\nu} + \bar{\nu}\Phi^{-1}(\hat{F}_i).$$

The QQ estimator regresses the empirical quantile, $Q_i^p$, on the theoretical quantiles, $Q_i^p$ or $Q_i^{LN}$. Thus, the heterogeneity parameter $\bar{\nu}$ of the log-normal distribution can be recovered as the coefficient on $\Phi^{-1}(\hat{F}_i)$. The primitive productivity parameter $\nu$ is given by $\bar{\nu}/(\sigma - 1)$. In the case of Pareto, the right hand side variable is $-\ln(1 - \hat{F}_i)$. The coefficient on $-\ln(1 - \hat{F}_i)$ gives us $1/\bar{\theta}$ from which we can back out the primitive parameter $\theta = (\sigma - 1)\hat{\theta}$. We provide more information on the QQ estimator and compare it to the more familiar rank-size regression in the appendix.

One advantage of the QQ estimator is that the linearity of the relationship between the theoretical and empirical quantiles means that the same estimate of the slope should be obtained even when the data are truncated. If the assumed distribution (Pareto or log-normal) fits the data well, we should recover the same slope estimate even when estimating on truncated subsamples.

We implement the QQ estimators on firm-level exports for the year 2000, using two sources, one for French exporters, and the other one for Chinese exporters. For both set of ex-
Table 1—Pareto vs Log-normal: QQ regressions (French exports to Belgium in 2000).

<table>
<thead>
<tr>
<th>Sample:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-normal:</td>
<td>ν</td>
<td>2.392</td>
<td>2.344</td>
<td>2.409</td>
<td>2.468</td>
<td>2.450</td>
<td>2.447</td>
<td>2.486</td>
</tr>
<tr>
<td>R²</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>0.999</td>
<td>0.998</td>
<td>0.998</td>
<td>0.996</td>
<td>0.992</td>
</tr>
<tr>
<td>ν</td>
<td>0.797</td>
<td>0.781</td>
<td>0.803</td>
<td>0.823</td>
<td>0.817</td>
<td>0.816</td>
<td>0.819</td>
<td>0.829</td>
</tr>
<tr>
<td>Pareto:</td>
<td>1/θ</td>
<td>2.146</td>
<td>1.390</td>
<td>1.174</td>
<td>0.915</td>
<td>0.884</td>
<td>0.855</td>
<td>0.822</td>
</tr>
<tr>
<td>R²</td>
<td>0.804</td>
<td>0.966</td>
<td>0.981</td>
<td>0.990</td>
<td>0.992</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
</tr>
</tbody>
</table>

The dependent variable is the log exports of French firms to Belgium in 2000. The RHS is $Φ^{-1}(\hat{F}_i)$ for log-normal and $\ln(1 - \hat{F}_i)$ for Pareto. $\nu$ and $\theta$ are calculated using $\sigma = 4$.

Figure 1. QQ graphs

(a) French firms → Belgium

(b) Chinese firms → Japan

Porters we use a leading destination: Belgium for French firms and Japan for Chinese ones. The precise mapping between productivity and sales distributions only holds for individual destination markets. Nevertheless, we also show in the appendix that the total sales distribution for French and Spanish firms follow distributions that resemble the log-normal more than the Pareto. As the theory fits better for producing firms, we show in results available upon request that the sample excluding intermediary firms continues to exhibit log-normality.

Table 1 reports results of QQ regressions for log-normal (top panel) and Pareto (bottom panel) assumptions for the theoretical quantiles. The first column retains all French exporters to Belgium in 2000, whereas the other columns successively increase the amount of truncation. The log-normal quantiles can explain 99.9% of the variation in the untruncated empirical quantiles, compared to 80% for Pareto. In the log-normal case the slope coefficient remains stable even as increasingly high shares of small exporters are removed. This what one would expect if the assumed distribution is correct. On the other hand, truncation dramatically changes the slope for the Pareto quantiles. This echoes results obtained by Eeckhout (2004) for city size distributions.

When running the same regressions on Chinese exports to Japan (the corresponding table can be found in the appendix), the same pattern emerges: log-normal seems to be a much better description of the data. The easiest way to see this is graphically. Figure 1 plots for both the French and the Chinese samples the relationship between the theoretical and empirical quantiles (top) and the histograms (bottom).
III. Micro-data simulations

Here we take as a benchmark M3 the values of θ obtained from truncated sample columns of Table II. While this does not matter much for log-normal (for which we take the un-truncated estimates), it is compulsory for Pareto, since the model needs θ > σ − 1 > 3 for that case. With the value of θ = 4.25 used by Melitz and Redding (2013), in mind, we choose the top 1% estimates as our benchmark: that is θ = 3.849 and ν = 0.797 for the French exporters case, and θ = 4.854 and ν = 0.853 for China.

We present results in a set of figures that show the GFT for both the Pareto and the log-normal cases, for values of τ0/2 < τ < 2τ0, with τ0, our benchmark level of trade costs. An advantage of that focus is that it keeps us within the range of parameters where α∗D < α∗A, ensuring that exporters are partitioned (in terms of productivity) from firms that serve the domestic market only.

As stated above, the share of firms that enter successfully (M1) affects gains from trade in the log-normal case, but not in the Pareto one. Figure 2 investigates the sensitivity of results when entry rates goes from tiny values (0.0055 as in Melitz and Redding (2013)), to very large ones (up to 0.75). The appendix shows that the impact of a rise in M1 on GFT is in general ambiguous, depending on relative rates of changes in α∗ under autarky and trading situations. A unique feature of Pareto is that those rates of change are exactly the same. Under log-normal, α∗D rises faster than α∗A. Intuitively, this is due to an additional detrimental effect on purely local firms under trade. In that situation, exporters at home exert a pressure on inputs, and exporters from the foreign country increase competition on the domestic market, such that the change in expected profits (determining the domestic cutoff) is lower under trade than under autarky, and gains from trade increase with M1. This reinforces the point following from equation 1 that it is not only the behavior in the right tail of the productivity distribution that matters for welfare. When M1 increases, cutoffs lie in regions where the two distributions diverge, and that affects relative welfare in a quantitatively relevant way. This raises the question of the appropriate value of M1. The fact that we do observe in the French, Chinese and Spanish domestic sales data a bell-shaped PDF suggests that more than half the potential entrants are choosing to operate (otherwise we would face a strictly declining PDF). As a conservative estimate, we therefore set M1=0.5 as our benchmark.

The second simulation, depicted in Figure 3 looks at the influence of truncation for combinations of parameters of the distributions. We keep ν at its benchmark level. Now it is the Pareto case that varies according to the different values of θ chosen (which depends on truncation). It is interesting to note that in both cases a larger variance in the productivity of firms (low θ or high ν) increases welfare: heterogeneity matters. Hence truncating the data, which results in larger values of τ—needed for the integrals to be bounded in this model—has an important effect on the size of gains from trade obtained: it lowers them.

IV. Discussion

In alternative simulations (in the appendix), we calibrate heterogeneity parameters on the macro-data trade elasticity, and find slight differences in GFT between the Pareto and log-normal assumptions. Hence, the precise method of calibration matters a great deal when trying to assess the importance of the distributional assumption. The micro-data method points to large GFT differences when the macro-data method points to very similar welfare outcomes.

Which calibration should be preferred? ACR make a compelling case for the macro-data calibration. However, we have several concerns. First, it seems more natural to actually use firm-level data to recover firms’ heterogeneity parameters. More crucially, a gravity equation with a constant trade elasticity is mis-specified under any distribution other than Pareto. That is, the empirical prediction that ϵni is constant across pairs of countries is unique to the Pareto distribution. The two papers we know of that test for non-constant trade elasticities (Helpman, Melitz and Rubinstein 2008 and Novy 2013) find distance elasticities to be indeed non-constant. Our ongoing work investigates the diversity of those reactions to trade costs in a more appropriate way, also departing from the massive simplification of the case of two symmetric countries.
Figure 2. Welfare gains, sensitivity to $M_1$ (entry rate)

(a) French firms → Belgium
(b) Chinese firms → Japan

Figure 3. Welfare gains, sensitivity to $M_3$ (truncation)

(a) French firms → Belgium
(b) Chinese firms → Japan

REFERENCES


Feenstra, Robert C. 2013. “Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity.” UC Davis Mimeo.


