Welfare and Trade Without Pareto*

Keith Head[†] Thierry Mayer[‡] Mathias Thoenig[§]

December 29, 2013

All the heterogeneous firm papers that use parametric distributions—that is, 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 very simple gains from trade (GFT) formula developed by Arkolakis et al. (2012) (hereafter, ACR). This implication is closely tied to fact that Pareto allows for a constant elasticity of substitution import system.[1] This paper investigates trade elasticities and welfare effects of trade cost reductions under a realistic alternative to the Pareto distribution.

As described thoroughly in Melitz and Redding (2014), much progress has been made in investigating the properties of heterogeneous firms models under general distributions. However, quantification of trade and welfare effects still requires parametric assumptions. Three important criteria have been invoked in deciding the appropriate distribution to use for heterogeneity. The first is tractability. The Pareto distribution makes it relatively easy to derive aggregate properties in an analytical model. However, uniform distributions and degenerate spikes are also tractable. Hence, users of the Pareto distribution also justify it on empirical and theoretical grounds. For example, ACR defend the Pareto assumptions with the arguments that it 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..."

We explore the consequences of replacing the assumption of Pareto heterogeneity with log-normal heterogeneity. This case is interesting because it (a) maintains some

*Paper to be presented at AEA, January 2014. This research has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 313522. We thank Maria Bas and Céline Poilly for their help.

[†]Sauder School of Business, University of British Columbia and CEPR, keith.head@sauder.ubc.ca

[‡]Sciences Po, CEPR, and CEPII.

[§]University of Lausanne and CEPR

[1]Two papers remove the long fat tail of the standard Pareto by bounding productivity from above. The first, Helpman et al. (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.

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.[2] The log-normal is reasonably tractable but its use sacrifices some "scale-free" properties conveyed by the Pareto distribution. The consequence is that gains from trade depend on the method of calibration. Under a calibration using macro-data—specifically the elasticity of aggregate bilateral trade with respect to trade costs—approximately the same gains from trade can be obtained. However, calibrating based on micro data—the size distribution of firm sales in a given market—yields very different gains from trade. In the symmetric two-country parameterized model considered by Melitz and Redding (2013), gains from trade can be twice as high under log-normal.

1 Welfare Theory

The heterogeneous firms version of monopolistic competition model with the CES (σ) demand has been stated fully in several places so we include only the most important equations here.[3] Note that we work with firms indexed by α, the unitary cost parameter. Consider a country with representative worker endowed with L efficiency units, paid wages w, and facing price index P. Welfare in i is given by real income:

$$\mathbb{W}_i \equiv \frac{w_i L_i}{P_i} = \left(\frac{L}{\sigma}\right)^{\sigma/(\sigma-1)} \frac{\sigma-1}{\tau_{ii} f_{ii}^{1/(\sigma-1)}} \frac{1}{\alpha_{ii}^*}, \tag{1}$$

where α^* denote the cutoff cost such that profits are zero (τ is the iceberg trade costs and f is the fixed production cost parameter).

As detailed in the online appendix, changes in welfare depend on changes in the price index which in turn depends on changes in the domestic cut-off which can be decomposed as

$$\frac{d\mathbb{W}_i}{\mathbb{W}_i} = -\frac{d\alpha_{ii}^*}{\alpha_{ii}^*} = \frac{1}{\epsilon_{ii}} \left(\frac{d\pi_{ii}}{\pi_{ii}} - \frac{dM_i^e}{M_i^e}\right). \tag{2}$$

Welfare changes depend on the change in the domestic trade share, π_{ii} and in the mass of domestic entrants, M_i^e. Both effects are stronger when the partial trade elasticity,[4] ϵ_{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. It is also important since it shows that one should be careful in arguing for Pareto based on its performance in the

[2] We compare generative processes for the two distributions in the online appendix.

[3] A full treatment using our notation is available in the appendix; in most respects, it follows the exposition in Melitz and Redding (2014).

[4] The "partial" conveys the point that this elasticity holds incomes and price indices constant. Thus, it corresponds to the elasticity obtained from a gravity equation estimated with origin and destination fixed effects. The online appendix contains a formal definition.

right tail of the distribution. This right tail may well be important for exporting (to difficult markets) but, assuming that survival into exporting to the domestic market is prevalent, it is the left part of the tail that is crucial for welfare. As pointed out by Eeckhout (2004) in the context of the city size distribution, it is in this left part of the distribution that Pareto and log-normal differ most dramatically.

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, ϵ_{ni} is constant which implies $dM_i^e = 0$.[5] This means we can integrate marginal changes to obtain the simple welfare formula of ACR, where $\widehat{\mathbb{W}}_i = \widehat{\pi}_{ii}^{1/\epsilon}$, where "hats" denote total changes. The log-normal case is much more complex and requires knowledge of the whole distribution of bilateral cutoffs. Unlike the Pareto case, one can no longer use "exact hat algebra" to solve the system as a function of observable trade shares, the trade elasticity and GDP. While we believe the multi-country log-normal model can be calibrated using additional data, here we want to build intuition on when and why departing from Pareto matters. Hence, we rely on a simpler approach of exploring the symmetric case described by Melitz and Redding (2013).

2 Calibration of the symmetric model

To consider the case of 2 symmetric countries, set $\tau_{ni} = \tau_{in} = \tau$, $\tau_{ii} = 1$, $f_{ii} = f_d$, $f_{ni} = f_{in} = f_x$ and $L_i = L$. We know from (1) that we need to investigate the behavior of the domestic cutoff, $\alpha_{ii}^* = \alpha_d^*$ which will entirely determine welfare. In this model, the cutoff equation is derived from the zero profit condition, one for the domestic and one for the export market in the trading equilibrium. Under symmetry, the ratio of export to domestic cutoffs depends only on a combination of parameters:

$$\frac{\alpha_x^*}{\alpha_d^*} = \frac{1}{\tau}\left(\frac{f_d}{f_x}\right)^{1/(\sigma-1)}, \tag{3}$$

Equilibrium also features the free-entry condition such that expected profits are equal to sunk costs:

$$\begin{aligned} &f_d \times G(\alpha_d^*)\left[\mathcal{H}(\alpha_d^*) - 1\right] \\ &+ f_x \times G(\alpha_x^*)\left[\mathcal{H}(\alpha_x^*) - 1\right] = f^E, \end{aligned} \tag{4}$$

where the $\mathcal{H}$ function is defined as $\mathcal{H}(\alpha^*) \equiv \frac{1}{\alpha^{*1-\sigma}} \int_0^{\alpha^*} \alpha^{1-\sigma} \frac{g(\alpha)}{G(\alpha^*)} d\alpha$, a monotonic, invertible function. Equations (3) and (4) characterize the equilibrium domestic cutoff α_d^*. Once the values for L, τ, f, f^E, f_x, σ have been set, and the functional form for $G()$ has been chosen, one can calculate welfare. The gains from trade are then the ratio of domestic cutoffs, autarkic over openness cases: $\mathbb{T}_i = \alpha_{dA}^*/\alpha_d^*$. The domestic

[5] See the working paper version of Arkolakis et al. (2012) for the proof.

cutoff in autarky is very simply obtained by restating the free entry condition as $f_d \times G(\alpha^*_{dA})\,[\mathcal{H}(\alpha^*_{dA}) - 1] = f^E$.

The last step is therefore to specify $G(\alpha)$. Pareto-distributed productivity implies a power law CDF for α:

$$G(\alpha) = \left(\frac{\alpha}{\bar{\alpha}}\right)^{\theta} \tag{5}$$

The Lognormal distribution of α retains the log-normality of productivity but with a change in the log-mean parameter from μ to $-\mu$, implying a CDF of

$$G(\alpha) = \Phi\left(\frac{\ln\alpha + \mu}{\nu}\right), \tag{6}$$

where for the rest of the paper we use Φ to denote the CDF of the standard normal. Parameter μ is the location parameter for the productivity distribution and ν is the dispersion parameter.

The equations needed for the quantification of the gains from trade are therefore (3) and (4), that provide α^*_d conditional on $G(\alpha^*_d)$, itself defined by (5) under Pareto and (6) under log-normal.

2.1 The 4 key moments to calibrate

There are four moments that are crucial 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, $G(\alpha^*_x)/G(\alpha^*_d)$ in the model. The target value for **M2** is 0.18, based on export rates of US firms reported by Melitz and Redding (2013).

M3 is the data moment used to calibrate the firm's heterogeneity parameter, denoted θ in Pareto and ν in Log-Normal. This is the most important moment for determining welfare effects and thus the method for calibrating it deserves the most attention. As noted by Arkolakis et al. (2012), there are two alternative moments that the model links closely to the heterogeneity parameters. The first, which we refer to simply as **M3**, is an estimate derived from the distribution of firm-level sales (exports) in some market. The second, which we call **M3′** is the trade elasticity ϵ_x. Under the Pareto distribution $\epsilon_x = \epsilon_d = -\theta$. Thus, we calibrate the Pareto heterogeneity parameter as $\theta = -\mathbf{M3'}$. Under log-normal

$$\mathbf{M3'} = 1 - \sigma - \frac{1}{\nu} h\left(\frac{\ln\alpha^*_x + \mu}{\nu} + (\sigma - 1)\nu\right),$$

where $h(x) \equiv \phi(x)/\Phi(x)$, the ratio of the PDF to the CDF of the standard normal.

M4: The share of export value in the total sales of exporters. Using the CES demand structure and country symmetry, this last moment is the one that will set the benchmark trade cost: $\hat{\tau}$. Indeed one can write $\mathbf{M4} = \frac{\hat{\tau}^{1-\sigma}}{1+\hat{\tau}^{1-\sigma}}$, which they take as 0.14 from US exporter data. Setting $\sigma = 4$, we can solve for $\hat{\tau} = ([(1 - \mathbf{M4})/\mathbf{M4}])^{1/3} = 1.83$. Note that this moment only relies on the CES demand structure, and therefore is not sensitive to the distribution assumption under symmetry.

Our simulations keep as a benchmark the target values for **M1**, **M2** and **M4** from Melitz and Redding (2013). The **M3** approach is described in section 2.2, with simulations implemented in section 3. The **M3′** approach is explained and implemented in section 4. Note finally that there are two remaining parameters to be determined: the CES preference parameter, σ, and the domestic fixed cost, f_d. We maintain $\sigma = 4$ and, since equations (3) and (4) imply that only relative f_x/f_d matters for equilibrium cutoffs, we set $f_d = 1$.

2.2 QQ estimators of shape parameters

This section develops the method of estimating the heterogeneity parameters for our two distributions of productivity which are needed for the micro data calibration of **M3**. Each of the two primitive distributions is characterized by a location parameter ($\bar{\alpha} \equiv \underline{\varphi}$ in Pareto or μ in log-normal) and a shape parameter (θ or ν). For most of our analysis, in particular for welfare and trade elasticities, the crucial parameter is the one governing shape, which is also directly related to the degree of heterogeneity of the distribution (falling with θ and rising with ν).

One important advantage of Pareto, pointed out by Redding (2011) is that if φ is Pareto then φ^r is Pareto also. The shape parameter becomes θ/r, the location becomes $\underline{\varphi}^r$. This advantage is shared by the log-normal. If φ is log-$\mathcal{N}(\mu, \nu)$, where μ and ν are, respectively the mean and standard deviation of log productivity, then φ^r is log-$\mathcal{N}(r\mu, r\nu)$.

These properties of the Pareto and log-normal distributions imply that we can estimate the shape parameter of productivity with data on the distribution of sales in some market. This is because the CES monopolistic competition assumption (CES-MC) implies that sales of an exporter from i to n, with efficiency α can be expressed as $x_{ni}(\alpha) = K_{ni}\alpha^{1-\sigma}$ where K_{ni} combines all the terms that depend on origin and destination but not on the identity of the firm. Therefore, CES-MC combined with productivity distributed Pareto($\underline{\varphi}, \theta$) implies that the sales of firms in any given market will be distributed Pareto($\underline{\tilde{\varphi}}, \tilde{\theta}$), where $\tilde{\theta} = \frac{\theta}{\sigma-1}$. If φ is log-$\mathcal{N}(\mu, \nu)$ then $\varphi^{\sigma-1}$ is log-$\mathcal{N}(\tilde{\mu}, \tilde{\nu})$, with $\tilde{\nu} = (\sigma - 1)\nu$.[6] Therefore, provided that we obtain estimates of $\tilde{\theta}$ and $\tilde{\nu}$ and that we postulate a value for σ, we can obtain estimates of θ and ν.

[6] The location parameters for sales are $\underline{\tilde{\varphi}} = K_{ni}\underline{\varphi}^{\sigma-1}$ and $\tilde{\mu} = (\sigma - 1)\mu + \ln K_{ni}$ for Pareto and log-normal respectively.

We apply the method of Kratz and Resnick (1996) to estimate parameters of truncated sales data that may come from either a Pareto or a log-normal distribution. They call the method a QQ estimator because it regresses empirical quantiles on corresponding theoretical quantiles. The original idea was developed for exploratory data visualization.

Dropping country subscripts for clarity, we denote sales as x_i where i now indexes firms ascending order of individual sales. Thus, $i = 1$ is the minimum sales and $i = n$ is the maximum. Our QQ estimators work with logged sales data. The empirical quantiles of the sorted data are $\mathcal{Q}_i^E = \ln x_i$.

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

$$F_{\mathrm{P}}(\ln x) = 1 - \exp[-\tilde{\theta}(\ln x - \ln \underline{x})], \tag{7}$$

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

$$F_{\mathrm{LN}}(\ln x) = \Phi((\ln x - \tilde{\mu})/\tilde{\nu}). \tag{8}$$

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 the empirical CDF:

$$\mathcal{Q}_i^{\mathrm{P}} = F_{\mathrm{P}}^{-1}(\hat{F}_i) = \ln \underline{x} - \frac{1}{\tilde{\theta}} \ln(1 - \hat{F}_i), \tag{9}$$

$$\mathcal{Q}_i^{\mathrm{LN}} = F_{\mathrm{LN}}^{-1}(\hat{F}_i) = \tilde{\mu}/\tilde{\nu} + \tilde{\nu}\Phi^{-1}(\hat{F}_i). \tag{10}$$

Bury (1999) recommends $\hat{F}_i = (i - 0.3)/(n + 0.4)$, as the empirical estimate of the CDF. The QQ estimator regresses the empirical quantile, $\mathcal{Q}_i^E$, on the theoretical quantiles, $\mathcal{Q}_i^{\mathrm{P}}$ or $\mathcal{Q}_i^{\mathrm{LN}}$. Thus, the heterogeneity parameter $\tilde{\nu}$ of the log-normal distribution can be recovered as the coefficient on $\Phi^{-1}(\hat{F}_i)$, and the primitive productivity parameter $\nu = \tilde{\nu}/(\sigma - 1)$.

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 $1/\tilde{\theta}$ from which we can back out the primitive parameter $\theta = (\sigma - 1)\tilde{\theta}$.[7]

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. Both

[7]There is a close relationship between the QQ estimator for the Pareto and the familiar log rank-size regressions examined by Gabaix and Ioannides (2004) since both rank, $1 + (n - i)$, and one minus the empirical CDF are linear in i.

Table 1: Pareto vs log-normal: QQ regressions (French exports to Belgium in 2000).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	all	top 50%	top 25%	top 5%	top 4%	top 3%	top 2%	top 1%
Obs:	34751	17376	8688	1737	1390	1042	695	347
	Log-normal: RHS = $\Phi^{-1}(\hat{F}_i)$, coeff = $\tilde{\nu}$							
$\Phi^{-1}(\hat{F}_i)$	2.392a	2.344a	2.409a	2.468a	2.450a	2.447a	2.457a	2.486a
R^2	0.999	0.999	1.000	0.999	0.998	0.998	0.996	0.992
ν	0.797	0.781	0.803	0.823	0.817	0.816	0.819	0.829
	Pareto: RHS = $-\ln(1-\hat{F}_i)$, coeff = $1/\tilde{\theta}$							
$-\ln(1-\hat{F}_i)$	2.146a	1.390a	1.174a	0.915a	0.884a	0.855a	0.822a	0.779a
R^2	0.804	0.966	0.981	0.990	0.992	0.994	0.994	0.994
θ	1.398	2.158	2.555	3.278	3.392	3.511	3.650	3.849

Notes: the dependent variable is the log exports of French firms to Belgium in 2000. The standard deviation of log exports in this sample is 2.393, which should be equal to $\tilde{\nu}$ if x is log-normally distributed and to $1/\tilde{\theta}$ if distribution if Pareto. ν and θ are calculated using $\sigma = 4$. Standard errors still have to be corrected.

datasets have been used in several recent papers (Eaton et al. (2011) is an example for French data, Manova and Zhang (2012) is one for Chinese exports). We view confrontation of evidence from different sources as important when assessing which distribution is preferred by the data. We apply two restrictions to the data. One follows theory in that the precise mapping between productivity and sales distributions is only valid when considering each destination market separately. For both set of exporters we use a leading destination: Belgium for French firms and Japan for Chinese ones.[8]

Table 1 reports results of QQ regressions for log-normal (top panel) and Pareto (bottom panel) assumptions, running QQ regressions for different truncations of the data. The first column keeps all French exporters to Belgium in 2000, the second one only the top 50%, etc. First, note in column (1) that the log-normal quantiles can explain 99.9% of the variation in the empirical quantiles, compared to 80% for Pareto. A second striking feature of the data is that the coefficients in the log-normal case are much less sensitive to truncation. This what one would expect if the correct distributional assumption was made. On the other hand, truncation dramatically changes the slope for the Pareto quantiles.

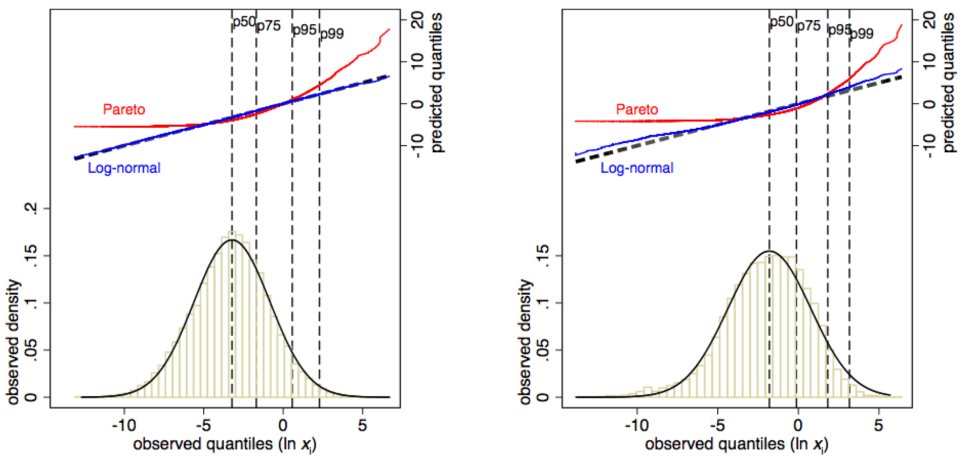
When running the same regressions on Chinese exports to Japan (the corresponding table can be found in the appendix), the same pattern emerges: log-normal seems

[8]The mapping is less clear when considering overall exports, or even total sales of firms (including domestic sales). Chaney (2013) and Di Giovanni et al. (2011) are examples using (truncated) total exports and sales respectively for French firms. The appendix corroborates the evidence in favor of log-normality of total sales of French and Spanish firms. We also show that the superior performance of log-normal is not driven by exports of intermediaries. For both the French and Chinese export samples, restricting to non-intermediaries yields similar results.

to be a much better description of the data. The easiest way to see this is graphically. Figure 1, inspired by the work of Battistin et al. (2009), plots for both the French and the Chinese samples the relationship between the theoretical and empirical quantiles (top) and the histograms (bottom).

Figure 1: QQ graphs

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



3 Micro-data simulations

In these simulation runs, we take as benchmark **M3** the values of θ obtained from *truncated* sample columns of Table 1. 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 $\theta > \sigma - 1 > 3$ for that case. In choosing where to truncate our goal was to obtain an estimate of θ that was close to the 4.25 used by Melitz and Redding (2013). We choose the top 1% estimates as our benchmark: that is $\theta = 3.849$ and $\nu = 0.797$ for the French exporters case, $\theta = 4.854$ and $\nu = 0.853$ for China.

Our graphs show the GFT for both the Pareto and the log-normal cases, for values of $1/2 \times \hat{\tau} < \tau < 2 \times \hat{\tau}$, with $\hat{\tau}$, our benchmark level of trade costs. An advantage of that focus is that it keeps us within the range of parameters where $\alpha_x^* < \alpha_d^*$, ensuring that exporters are partitioned (in terms of productivity) from firms that serve the domestic market only. Results for the whole range of τ values are provided in the online appendix.

A first important difference between the log-normal and Pareto versions of the model is that the share of firms that enter successfully (the value of **M1**) affects gains from trade in the former, but not in the latter case (see the online appendix).

Figure 2: Welfare gains, sensitivity to **M1**

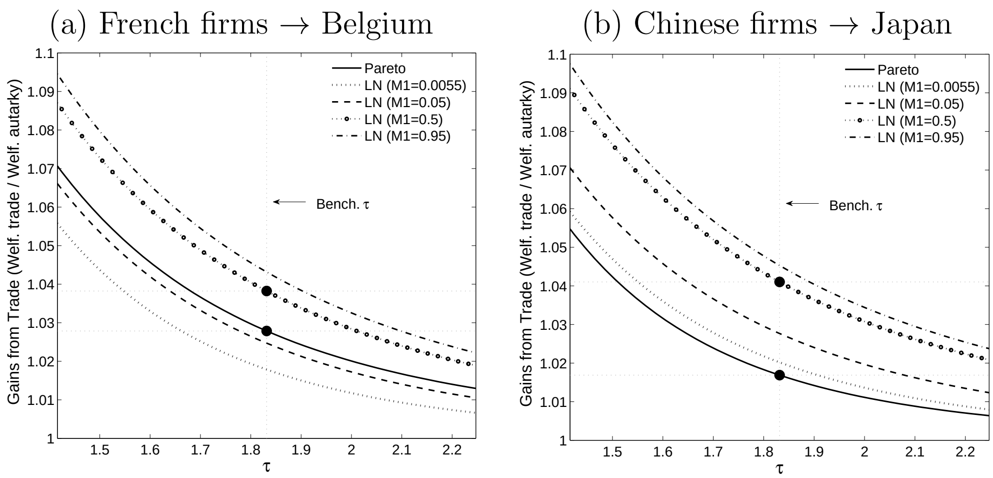


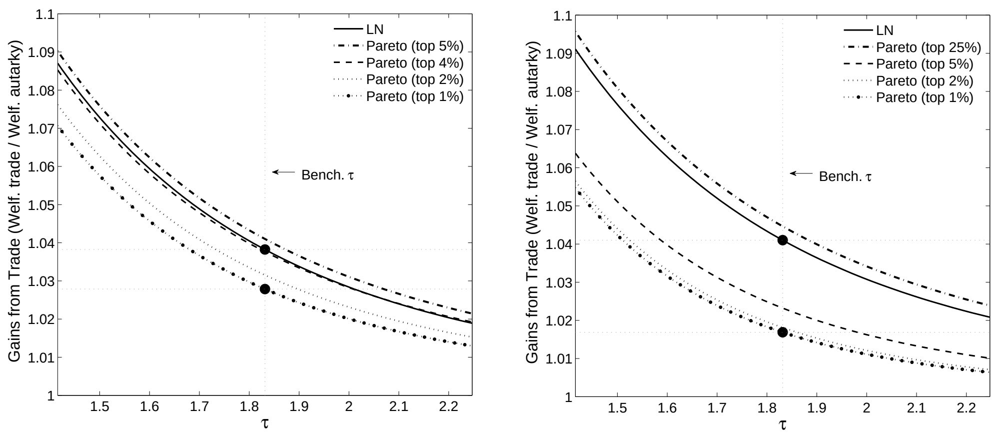
Figure 2 investigates how large a difference does it make to go from extremely small entry rates (0.0055 as in Melitz and Redding (2013)), to very large ones (0.75). What is in general the impact of **M1** on welfare in this model? In terms of calibrated parameters , in order to have a larger proportion of firms enter successfully, the free entry equation (4) suggests that the sunk entry costs f^E needs to rise. This is because a rise in the sunk costs reduces the total mass of entrants and therefore competition on both input and goods markets, which makes entry easier. Therefore α^* and hence $\mathbf{M1} = G(\alpha^*)$ rise (a proof is provided in the online appendix). This rise in f^E leads to a reduction in welfare in both autarky and trading equilibria. Therefore the impact on gains from trade is in general ambiguous, depending on relative rates of changes in α^*. A unique feature of Pareto is that those rates of change are exactly the same. Under log-normal, α^*_{dA} rises faster than α^*_d. Intuitively, this is because the rise in sunk costs, f^E, has 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**. Results in Figure 2 show that the welfare in log-normal is highly dependent on the value of **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** increase, 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**, which is not directly observable. 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 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 which will vary according to the different values of θ chosen (which depend on how much you truncate the data). 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 θ—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.

Figure 3: Welfare gains, sensitivity to truncation

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

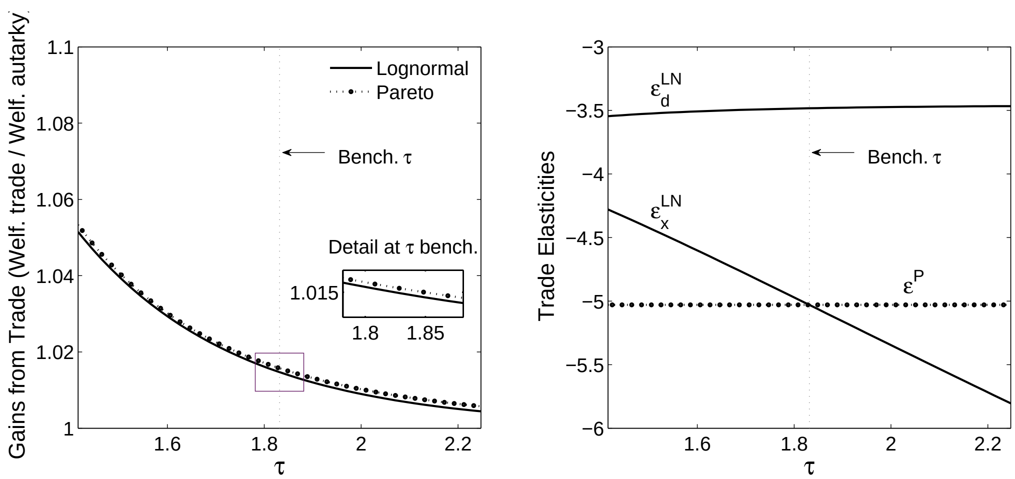


4 Macro-data simulations

In this section, we adopt the **M3′** approach where the underlying micro parameters ν and θ are calibrated to match a trade elasticity. The most obvious strategy is to borrow estimates from the gravity literature regressing trade flows on bilateral applied tariffs. Head and Mayer (2014) survey this literature and report a median estimate of -5.03, which we take as our target for both Pareto and log-normal. The left panel of figure 4 plots the GFT as in previous figures, and the right panel graphs the three relevant trade elasticities: ϵ^{P} for Pareto, constant at -5.03, ϵ_x^{LN} and ϵ_d^{LN}, the international and domestic elasticities for the log-normal case. By construction, ϵ_x^{LN} coincides with Pareto at the benchmark trade cost ($\tau = 1.83$). As τ declines, the elasticity falls in absolute value. The domestic elasticity, ϵ_d^{LN}, is uniformly smaller in absolute value than ϵ_x^{LN}. And it rises with increases in τ because higher international trade costs make the domestic market easier in relative terms.

Despite this large heterogeneity in trade elasticities between Pareto and log-normal, gains from trade happen to be very proximate in this the symmetric country calibration. While the GFT are very similar for this set of parameters, they are not identical, as the zoomed-in box reveals. Second, they can be much more different when one changes other parameter targets, in particular the share of exporters (see online appendix). Third, this calibration sets the heterogeneity parameter ν in order to fit a unique trade elasticity (the international one), while the LN version of the model features *two* elasticities that depend crucially on ν. Calibrating the model to fit an average of the two trade elasticities(see the appendix), the Pareto and log-normal GFT again diverge from each other.

Figure 4: Welfare gains calibrated on trade elasticity



5 Discussion

Our two calibration exercises yield quite different results, the micro-data one pointing to large differences between Pareto and log-normal, the macro-data one pointing to more similar outcomes in terms of welfare. 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 ϵ_{ni} 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 et al. (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.

References

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

Battistin, E., R. Blundell, and A. Lewbel (2009). *Journal of Political Economy 117*(6), 1140–1154.

Bury, K. (1999). *Statistical distributions in engineering*. Cambridge University Press.

Chaney, T. (2013). The gravity equation in international trade: An explanation. Working Paper 19285, National Bureau of Economic Research.

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

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from French firms. *Econometrica 79*(5), 1453–1498.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 1429–1451.

Feenstra, R. C. (2013). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. UC Davis Mimeo.

Gabaix, X. and Y. M. Ioannides (2004). The evolution of city size distributions. *Handbook of regional and urban economics 4*, 2341–2378.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In E. Helpman, G. Gopinath, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4. Elsevier.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics 123*(2), 441–487.

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

Manova, K. and Z. Zhang (2012). Export prices across firms and destinations. *The Quarterly Journal of Economics 127*(1), 379–436.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and S. J. Redding (2013). Firm heterogeneity and aggregate welfare. Working Paper 18919, National Bureau of Economic Research.

Melitz, M. J. and S. J. Redding (2014). Heterogeneous firms and trade. In E. Helpman, G. Gopinath, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4. Elsevier.

Novy, D. (2013). International trade without CES: Estimating translog gravity. *Journal of International Economics 89*(2), 271–282.

Redding, S. J. (2011). Theories of heterogeneous firms and trade. *Annual Review of Economics 3*, 77–105.

Appendix

A.1 Welfare and the share of domestic trade

Here we derive equation (2), showing welfare changes as a function of changes in the domestic share and the mass of domestic entrants. This equation resembles an un-numbered equation in Arkolakis et al. (2012), p. 111. However, it reduces the determinants of welfare to just changes in own trade and changes in the mass of entrants. Along the way, we set up the model in general terms: C asymmetric countries, and general distribution functions, which provides equation (2) and other useful results fo the calibration.

Bilateral trade can be expressed as the product of M^e_{ni}, the mass of entrants from i into destination n, and the mean export revenues of exporters from i serving market n.

$$X_{ni} = G(\alpha^*_{ni}) M^e_i \int_0^{\alpha^*_{ni}} \frac{x_{ni}(\alpha) g(\alpha) d\alpha}{G(\alpha^*_{ni})}, \tag{A.1}$$

where α^*_{ni} is the cutoff cost over which firms in i would make a loss in market n.

With demand being CES (denoted σ), equilibrium markups ($\bar{m} = \sigma/(\sigma-1)$) being constant, and trade costs (τ_{ni}) being iceberg, the export value of an individual firm with productivity $1/\alpha$ is given by

$$x_{ni}(\alpha) = (\bar{m}\alpha w_i \tau_{ni})^{1-\sigma} P_n^{\sigma-1} Y_n, \tag{A.2}$$

with Y_n denoting total expenditure and P_n the price index of the CES composite.

Following Helpman et al. (2008), it is useful to define

$$V_{ni} = \int_0^{\alpha^*_{ni}} \alpha^{1-\sigma} g_i(\alpha) d\alpha. \tag{A.3}$$

Now we can re-express aggregate exports from i to n as

$$X_{ni} = M_i^e Y_n (\bar{m} w_i \tau_{ni})^{1-\sigma} P_n^{\sigma-1} V_{ni}, \qquad \text{with} \qquad P_n^{1-\sigma} \equiv \sum_\ell M_\ell^e (\bar{m} w_\ell \tau_{n\ell})^{1-\sigma} V_{n\ell}. \tag{A.4}$$

Since market clearing and balanced trade imply $Y_i = w_i L_i$, we can replace w_i with Y_i/L_i. We also divide X_{ni} by Y_n to obtain the expenditure shares, π_{ni} for importer n on exporter i:

$$\pi_{ni} = M_i^e L_i^{\sigma-1} Y_i^{1-\sigma} (\bar{m}\tau_{ni})^{1-\sigma} V_{ni} P_n^{\sigma-1}, \tag{A.5}$$

with

$$P_n^{1-\sigma} = \sum_\ell M_\ell^e L_\ell^{\sigma-1} Y_\ell^{1-\sigma} (\bar{m}\tau_{n\ell})^{1-\sigma} V_{n\ell}. \tag{A.6}$$

Gross profits in the CES model are given by x_{ni}/σ. Hence, assuming that fixed costs are paid using labor of the origin country, the cutoff cost such that profits are zero is determined by $x_{ni}(\alpha^*) = \sigma w_i f_{ni}$. Combined with $w_i = Y_i/L_i$ we obtain:

$$\alpha_{ni}^* = \sigma^{1/(1-\sigma)} \left(\frac{L_i}{Y_i}\right)^{\sigma/(\sigma-1)} \left(\frac{Y_n}{f_{ni}}\right)^{1/(\sigma-1)} \frac{P_n}{\bar{m}\tau_{ni}}. \tag{A.7}$$

Welfare in this model is given by real income. Inverting equation (A.7), welfare can be expressed in terms of the domestic cutoff:

$$\mathbb{W}_i \equiv \frac{Y_i}{P_i} = \left(\frac{L_i}{\sigma}\right)^{\sigma/(\sigma-1)} \frac{\sigma-1}{\tau_{ii} f_{ii}^{1/(\sigma-1)}} \frac{1}{\alpha_{ii}^*}. \tag{A.8}$$

This is equation (1) in the main text. Since α_{ii}^* is the sole endogenous variable, $\frac{d\mathbb{W}_i}{\mathbb{W}_i} = -\frac{d\alpha_{ii}^*}{\alpha_{ii}^*}$. The next step is to relate changes in the cutoff to changes in trade shares. To do this we divide both sides of equation (A.6) by $P_n^{1-\sigma}$, and differentiate, to obtain:

$$\sum_\ell \pi_{n\ell} \left[\frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\frac{dY_\ell}{Y_\ell} + \frac{dV_{n\ell}}{V_{n\ell}} + (\sigma-1)\frac{dP_n}{P_n}\right] = 0 \tag{A.9}$$

Analyzing the dV/V term first, we can see from the definition in equation (A.3) that it is the product of the elasticity of V with respect to the cutoff times the percent change in the cutoff. We follow ACR in denoting the first elasticity as γ; it is given by

$$\gamma_{ni} \equiv \frac{d\ln V_{ni}}{d\ln \alpha_{ni}^*} = \frac{\alpha_{ni}^{*2-\sigma} g(\alpha_{ni}^*)}{\int_0^{\alpha_{ni}^*} \alpha^{1-\sigma} g(\alpha) d\alpha}. \tag{A.10}$$

From the definition of V and equilibrium cutoffs in (A.7), we can write the change in V as

$$\frac{dV_{n\ell}}{V_{n\ell}} = \gamma_{n\ell} \frac{d\alpha_{n\ell}^*}{\alpha_{n\ell}^*} = \gamma_{n\ell} \left[\frac{1}{\sigma-1}\frac{dY_n}{Y_n} - \frac{\sigma}{\sigma-1}\frac{dY_\ell}{Y_\ell} + \frac{dP_n}{P_n} - \frac{d\tau_{n\ell}}{\tau_{n\ell}}\right]. \tag{A.11}$$

Combining (A.9) and (A.11) leads to

$$\sum_{\ell} \pi_{n\ell} \left[\frac{dM_\ell^e}{M_\ell^e} + (1-\sigma-\gamma_{n\ell}) \left(\frac{d\tau_{n\ell}}{\tau_{n\ell}} - \frac{dP_n}{P_n} \right) + \left(1 - \sigma - \frac{\sigma\gamma_{n\ell}}{\sigma-1} \right) \frac{dY_\ell}{Y_\ell} + \frac{\gamma_{n\ell}}{\sigma-1} \frac{dY_n}{Y_n} \right] = 0 \tag{A.12}$$

Differentiating bilateral trade shares in equation (A.5),

$$\frac{d\pi_{n\ell}}{\pi_{n\ell}} = \frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\frac{dY_\ell}{Y_\ell} + \frac{dV_{n\ell}}{V_{n\ell}} + (\sigma-1)\frac{dP_n}{P_n}, \tag{A.13}$$

$$\frac{d\pi_{nn}}{\pi_{nn}} = \frac{dM_n^e}{M_n^e} + (1-\sigma)\frac{dY_n}{Y_n} + \frac{dV_{nn}}{V_{nn}} + (\sigma-1)\frac{dP_n}{P_n}. \tag{A.14}$$

Hence, the difference in those share changes gives

$$\frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} = \frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\left[\frac{dY_\ell}{Y_\ell} - \frac{dY_n}{Y_n} \right] + \frac{dV_{n\ell}}{V_{n\ell}} - \frac{dV_{nn}}{V_{nn}}. \tag{A.15}$$

Let us focus now in the difference in V term. From (A.11), we can write:

$$\begin{aligned} \frac{dV_{n\ell}}{V_{n\ell}} - \frac{dV_{nn}}{V_{nn}} &= \gamma_{n\ell}\frac{d\alpha_{n\ell}^*}{\alpha_{n\ell}^*} - \gamma_{nn}\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} \\ &= \gamma_{n\ell}\left[\frac{1}{\sigma-1}\frac{dY_n}{Y_n} - \frac{\sigma}{\sigma-1}\frac{dY_\ell}{Y_\ell} - \frac{d\tau_{n\ell}}{\tau_{n\ell}} + \frac{dP_n}{P_n} \right] \\ &\qquad - \gamma_{nn}\left[-\frac{dY_n}{Y_n} + \frac{dP_n}{P_n} \right]. \\ &= (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} + \gamma_{n\ell}\left[\frac{\sigma}{\sigma-1}\left(\frac{dY_n}{Y_n} - \frac{dY_\ell}{Y_\ell} \right) - \frac{d\tau_{n\ell}}{\tau_{n\ell}} \right]. \end{aligned} \tag{A.16}$$

We then plug (A.16) into (A.15) to obtain

$$\begin{aligned} \frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} - (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} =& \frac{dM_\ell^e}{M_\ell^e} + (1-\sigma-\gamma_{n\ell})\frac{d\tau_{n\ell}}{\tau_{n\ell}} \\ &+ \left(1 - \sigma - \frac{\sigma\gamma_{n\ell}}{\sigma-1} \right)\left[\frac{dY_\ell}{Y_\ell} - \frac{dY_n}{Y_n} \right]. \end{aligned} \tag{A.17}$$

Therefore the term in square brackets inside (A.12) is equal to

$$\frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} - (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} + (1-\sigma-\gamma_{n\ell})\left[\frac{dY_n}{Y_n} - \frac{dP_n}{P_n} \right]. \tag{A.18}$$

After replacing $\frac{dY_n}{Y_n} - \frac{dP_n}{P_n} = -\frac{d\alpha_{nn}^*}{\alpha_{nn}^*}$, and canceling out the terms involving $\gamma_{n\ell}$, we can substitute the result into (A.12) to obtain

$$\sum_{\ell} \pi_{n\ell}\left[\frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} + (\sigma - 1 + \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} \right] = 0 \tag{A.19}$$

Noting that only $d\pi_{n\ell}/\pi_{n\ell}$ terms depend on ℓ we can re-arrange as

$$-(\sigma-1+\gamma_{nn})\frac{d\alpha^*_{nn}}{\alpha^*_{nn}} = -\frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM^e_n}{M^e_n} + \sum_{\ell}\pi_{n\ell}\frac{d\pi_{n\ell}}{\pi_{n\ell}} \tag{A.20}$$

Using $\sum_\ell \pi_{n\ell}\frac{d\pi_{n\ell}}{\pi_{n\ell}} = 0$, we can finally express the welfare change as

$$\frac{d\mathbb{W}_n}{\mathbb{W}_n} = -\frac{d\alpha^*_{nn}}{\alpha^*_{nn}} = \frac{-d\pi_{nn}/\pi_{nn} + dM^e_n/M^e_n}{(\sigma-1+\gamma_{nn})}, \tag{A.21}$$

which after defining $\epsilon_{nn} = 1-\sigma-\gamma_{nn}$, is equation (2) in the text.

A.2 How M1 (entry share) affects welfare in the symmetric model

Under the trading regime, ur micro-data calibration procedure is characterized by the two equilibrium relationships (3) and (4), the two moment conditions $\mathbf{M1}-G(\alpha^*_d)=0$ and $\mathbf{M2}-G(\alpha^*_x)/G(\alpha^*_d)=0$, and four unknowns $(\alpha^*_d, \alpha^*_x; f^E, f_x)$.

Differentiating the two moment conditions with respect to **M1** we obtain

$$\frac{d\alpha^*_d/\alpha^*_d}{d\mathbf{M1}/\mathbf{M1}} = \frac{G(\alpha^*_d)}{\alpha^*_d G'(\alpha^*_d)} > 0, \tag{A.22}$$

$$\frac{d\alpha^*_x/\alpha^*_x}{d\mathbf{M1}/\mathbf{M1}} = \frac{G(\alpha^*_x)}{\alpha^*_x G'(\alpha^*_x)} > 0, \tag{A.23}$$

Simple manipulations of the differentiated system also yields

$$\frac{df_x}{f_x} = (\sigma-1)\times\left[\frac{G(\alpha^*_d)}{\alpha^*_d G'(\alpha^*_d)} - \frac{G(\alpha^*_x)}{\alpha^*_x G'(\alpha^*_x)}\right]\times\frac{d\mathbf{M1}}{\mathbf{M1}}, \tag{A.24}$$

$$\frac{df^E}{f^E} = A_1^+\frac{d\alpha^*_d}{\alpha^*_d} + A_2^+\frac{d\alpha^*_x}{\alpha^*_x} + A_3^+\frac{df_x}{f_x}, \tag{A.25}$$

where (A_1^+, A_2^+, A_3^+) are positive parameters. The right hand side of (A.24) is zero under Pareto. Looking at definition (5), it is indeed clear that $\frac{G(\alpha^*_d)}{\alpha^*_d G'(\alpha^*_d)} - \frac{G(\alpha^*_x)}{\alpha^*_x G'(\alpha^*_x)} = 0$. Therefore, a change of **M1** is i) not related to changes in f_x, ii) affecting all cutoffs in the same way, leaving export propensity, but also gains from trade unaffected. Under log-normal on the contrary, $\frac{G(\alpha^*_d)}{\alpha^*_d G'(\alpha^*_d)} - \frac{G(\alpha^*_x)}{\alpha^*_x G'(\alpha^*_x)} > 0$ (see (6)). Hence

$$\frac{df_x/f_x}{d\mathbf{M1}/\mathbf{M1}} \geqslant 0 \tag{A.26}$$

Combined with (A.22), (A.23), (A.24) and (A.25) thus imply

$$\frac{df^E/f^E}{d\mathbf{M1}/\mathbf{M1}} > 0 \tag{A.27}$$

.

Let consider now the domestic cutoff in autarky, characterized by $G(\alpha^*_{dA})\left[\mathcal{H}(\alpha^*_{dA}) - 1\right] = f^E$. Differentiating this relationship we get

$$\frac{d\alpha^*_{dA}/\alpha^*_{dA}}{d\mathbf{M1}/\mathbf{M1}} > 0 \tag{A.28}$$

We conclude from the previous computations that an increase in $\mathbf{M1}$ leads to an increase in both α^*_d and α^*_{dA}, namely a less selective domestic market both in autarky and in the trading equilibrium.

The change in trade gains is equal to

$$\frac{d\mathbb{T}}{\mathbb{T}} = \left[\frac{d\alpha^*_{dA}}{\alpha^*_{dA}} - \frac{d\alpha^*_d}{\alpha^*_d}\right] \times \frac{d\mathbf{M1}}{\mathbf{M1}} \tag{A.29}$$

The sign of the previous relationship cannot be characterized algebraically and we consequently rely on our quantitative procedure to show that it is positive under log-normal.

A.3 Distribution parameters for Chinese exports to Japan

Here we replicate Table 1 for the case of Chinese exports to Japan in 2000.

A.4 Distributions of total sales

Some of the prior literature asserting Pareto is based on firm size distribution (Di Giovanni et al. (2011) for instance), rather than looking at the distribution of export sales from one origin in a particular importing country (which is also done in Eaton et al. (2011)).

A.5 Macro-data simulation targeting the average elasticity

A.6 Comparison of QQ estimator to other methods

One alternative to the QQ estimators is to use method of moments. In this case that means inferring the distributional parameters from the means and standard deviations of log sales. We can use equations (7) and (8) to obtain an idea of what those coefficients should be. With log of sales distributed Normal, they have a mean value of $\tilde{\mu}$, and a standard deviation of $\tilde{\nu}$. In the Pareto case, the log of sales have a mean value of $\ln \underline{\tilde{\varphi}} + 1/\tilde{\theta}$, and a standard deviation of $1/\tilde{\theta}$. In this sample, the standard

Table 2: Pareto vs Log-Normal: QQ regressions (Chinese exports to Japan in 2000).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	all	top 50%	top 25%	top 5%	top 4%	top 3%	top 2%	top 1%
Obs:	24832	12416	6208	1241	993	745	496	248
	Log-normal: RHS = $\Phi^{-1}(\hat{F}_i)$, coeff = $\tilde{\nu}$							
$\Phi^{-1}(\hat{F}_i)$	2.558a	2.125a	1.950a	1.936a	1.934a	1.929a	1.910a	1.970a
R^2	0.986	0.995	0.999	0.998	0.998	0.997	0.995	0.992
ν	0.853	0.708	0.650	0.645	0.645	0.643	0.637	0.657
	Pareto: RHS = $-\ln(1-\hat{F}_i)$, coeff = $1/\tilde{\theta}$							
$-\ln(1-\hat{F}_i)$	2.194a	1.239a	0.946a	0.718a	0.698a	0.674a	0.640a	0.618a
R^2	0.725	0.930	0.971	0.990	0.991	0.992	0.995	0.994
θ	1.367	2.422	3.170	4.175	4.296	4.452	4.688	4.854

Notes: the dependent variable is the log exports of Chinese firms to Japan in 2000. The standard deviation of log exports in this sample is 2.576, which should be equal to $\tilde{\nu}$ if x is log-normally distributed and to $1/\tilde{\theta}$ if distribution if Pareto. ν and θ are calculated using $\sigma = 4$. Standard errors still have to be corrected.

Table 3: Pareto vs Log-Normal: QQ regressions (French firms total sales in 2000).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	all	top 50%	top 25%	top 5%	top 4%	top 3%	top 2%	top 1%
Obs:	92988	46494	23247	4649	3719	2789	1860	930
	Log-normal: RHS = $\Phi^{-1}(\hat{F}_i)$, coeff = $\tilde{\nu}$							
$\Phi^{-1}(\hat{F}_i)$	1.790a	2.076a	2.330a	2.579a	2.586a	2.603a	2.610a	2.586a
R^2	0.984	0.990	0.996	0.999	0.998	0.998	0.997	0.992
ν	0.597	0.692	0.777	0.860	0.862	0.868	0.870	0.862
	Pareto: RHS = $-\ln(1-\hat{F}_i)$, coeff = $1/\tilde{\theta}$							
$-\ln(1-\hat{F}_i)$	1.658a	1.251a	1.143a	0.955a	0.932a	0.906a	0.869a	0.806a
R^2	0.844	0.988	0.991	0.991	0.991	0.990	0.990	0.989
θ	1.809	2.398	2.624	3.140	3.220	3.312	3.452	3.723

Notes: the dependent variable is the log exports of French total sales in 2000. The standard deviation of log exports in this sample is 1.805, which should be equal to $\tilde{\nu}$ if x is log-normally distributed and to $1/\tilde{\theta}$ if distribution if Pareto. ν and θ are calculated using $\sigma = 4$. Standard errors still have to be corrected.

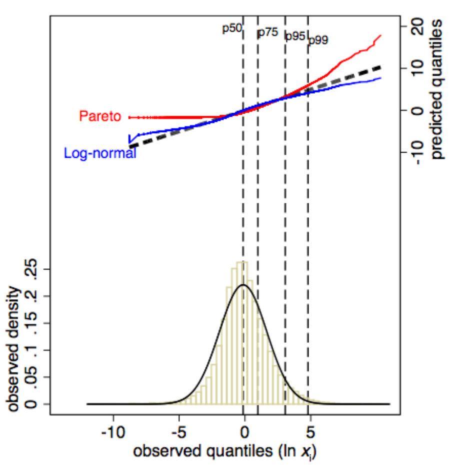
Table 4: Pareto vs Log-Normal: QQ regressions (Spanish firms total sales in 2000).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	all	top 50%	top 25%	top 5%	top 4%	top 3%	top 2%	top 1%
Obs:	87998	43999	21999	4400	3520	2640	1760	880
	Log-normal: RHS = $\Phi^{-1}(\hat{F}_i)$, coeff = $\tilde{\nu}$							
$\Phi^{-1}(\hat{F}_i)$	1.588a	1.859a	2.095a	2.419a	2.435a	2.462a	2.510a	2.599a
R^2	0.986	0.988	0.992	0.998	0.997	0.996	0.995	0.991
ν	0.529	0.620	0.698	0.806	0.812	0.821	0.837	0.866
	Pareto: RHS = $-\ln(1-\hat{F}_i)$, coeff = $1/\tilde{\theta}$							
$-\ln(1-\hat{F}_i)$	1.489a	1.122a	1.032a	0.899a	0.880a	0.861a	0.840a	0.814a
R^2	0.866	0.990	0.995	0.995	0.996	0.997	0.997	0.996
θ	2.015	2.674	2.907	3.337	3.409	3.486	3.573	3.687

Notes: the dependent variable is the log exports of Spanish total sales in 2000. The standard deviation of log exports in this sample is 1.599, which should be equal to $\tilde{\nu}$ if x is log-normally distributed and to $1/\tilde{\theta}$ if distribution if Pareto. ν and θ are calculated using $\sigma = 4$. Standard errors still have to be corrected.

Figure 5: QQ graphs on total sales

(a) French firms



(b) Spanish firms

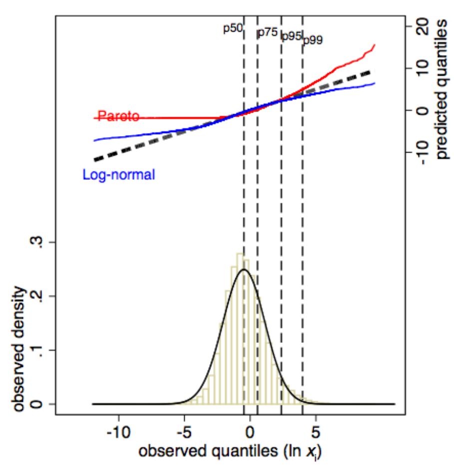
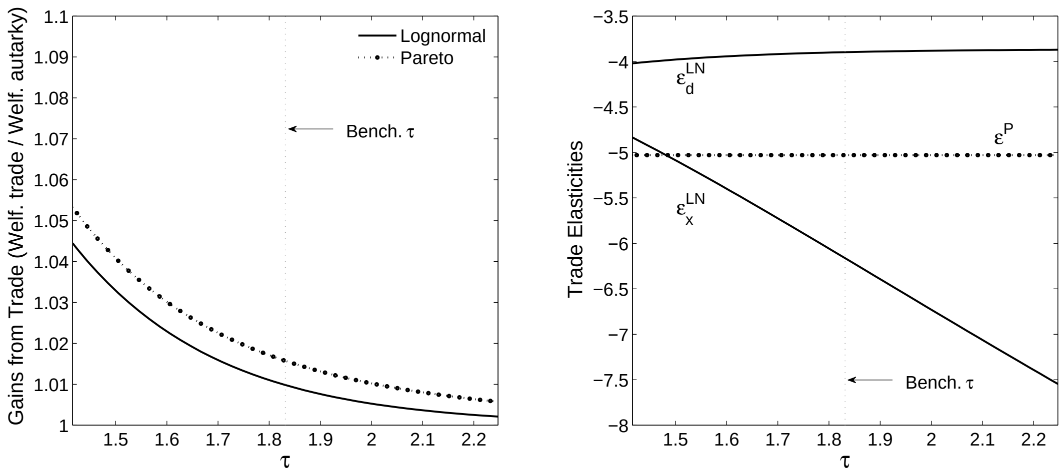


Figure 6: Welfare gains calibrated on average trade elasticity



deviation of log sales is 2.389, hence predicted coefficients in Table 1 are 2.389 for Log-Normal and Pareto independently of truncation. The un-truncated sample estimate almost exactly matches that prediction for the log-normal case, when most estimates of Pareto case are quite far off.

A frequently used estimate of the Pareto parameter derives from a regression of logged rank on log firm size. The coefficient on log sales is $-\tilde{\theta} = -\frac{\theta}{\sigma-1}$. Eaton et al. (2011), Di Giovanni et al. (2011) are recent examples that pursue this approach and it is also referred to by Melitz and Redding (2013) in their parameterization of **M3**.

A.7 Generative processes for log-normal and Pareto

Because the Pareto distribution has been thought to characterize a large set of phenomena in both natural and social sciences, much effort has gone into developing generative models that predict the Pareto as a limiting distribution. The building block emphasized in the literature, see especially Gabaix (1999), is Gibrat's law of proportional growth. Applied to sales of an individual firm i in period t, Gibrat's Law states that $X_{i,t+1} = \Gamma_{it} X_{it}$. The key point is that the growth rate from period to period, $\Gamma_{it} - 1$ is independent of size. A confusion has arisen because it is straightforward to show that the law of proportional growth delivers a log-normal distribution. In period T size is given by

$$X_{iT} = \exp(\ln X_{i0} + \sum_{t=1}^{T} \ln \Gamma_{it})$$

The central limit theorem implies for large T,

$$\sqrt{T}\left(\frac{\sum_t \ln \Gamma_{it}}{T} - \mathbb{E}[\ln \Gamma_{it}]\right) \sim \mathcal{N}(0, \mathbb{V}[\ln \Gamma_{it}]),$$

where $\mathbb{E}$ and $\mathbb{V}$ are the expectation and variance operators. Rearranging and, for convenience only, initializing sizes at $X_{i0} = 1$, $\ln X_{it}$ is normally distributed with expectation $T\mathbb{E}[\ln \Gamma_{it}]$ and variance $T\mathbb{V}[\ln \Gamma_{it}]$. This implies X_{iT} is log-normal with log-mean parameter $\tilde{\mu} = T\mathbb{E}[\ln \Gamma_{it}]$ and log-SD parameter $\tilde{\nu} = \sqrt{T\mathbb{V}[\ln \Gamma_{it}]}$.

This demonstration that Gibrat's Law implies a limiting distribution that is log-normal echoes similar arguments by Sutton (1997) for firms and Eeckhout (2004) for cities. The problem with this formulation is that it is only valid for large T and yet as T grows large, the distribution exhibits some perverse behavior. Assume that sizes are not growing on average, i.e. $\mathbb{E}[\Gamma_{it}] = 1$. By Jensen's Inequality, $\mathbb{E}[\ln \Gamma_{it}] < \ln(\mathbb{E}[\Gamma_{it}]) = 0$. Since the median of X_{iT} is $\exp(\tilde{\mu}) = \exp(T\mathbb{E}[\ln \Gamma_{it}])$, the median should decline exponentially with time. The mode, $\exp(\tilde{\mu} - \tilde{\nu}^2) = \exp[T(\mathbb{E}[\ln \Gamma_{it}] - \mathbb{V}[\ln \Gamma_{it}])]$ should decline even more rapidly with time. Thus, as T becomes large, Gibrat's law with $\mathbb{E}[\Gamma_{it}] = 1$ implies a distribution with a mode going to zero while the variance is becoming infinite. Evidently something must be done to rescue Gibrat's law from generating degeneracy.

A variety of modifications to Gibrat's Law have been investigated. Kalecki (1945) specifies growth shocks that are negatively correlated with the level. This allows for a log-normal with stable variance to emerge. Gabaix (1999) shows in an appendix that a simple change to the growth process, $X_{i,t+1} = \Gamma_{it} X_{it} + \varepsilon$ with $\varepsilon > 0$ (the Kesten process) is enough to solve the problem of degeneracy. But the resulting stable distribution is Pareto, not log-normal. Reed (2001) instead assumes finite-lived agents with exponential life expectancies. This leads to a double-Pareto distribution.

Appendix References

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

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management 17*(1), 99–120.

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

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from French firms. *Econometrica 79*(5), 1453–1498.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 1429–1451.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics 114*(3), 739–767.

Guiso, L., P. Sapienza, and L. Zingales (2013). The value of corporate culture. Discussion Paper 9716, CEPR.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics 123*(2), 441–487.

Kalecki, M. (1945). On the Gibrat distribution. *Econometrica: Journal of the Econometric Society*, 161–170.

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

Lee, S. and Q. Li (2013, November). Uneven landscapes and the city size distribution. *Journal of Urban Economics 78*, 19–29.

Reed, W. J. (2001). The Pareto, Zipf and other power laws. *Economics Letters 74*(1), 15–19.

Sutton, J. (1997). Gibrat's legacy. *Journal of economic Literature 35*(1), 40–59.