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.[1]

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 on-

line appendix). The log-normal is reasonably tractable but its use sacrifices some "scale-free" properties conveyed by the Pareto distribution. Aspects of the 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

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

where α_{ii}^*, τ_{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), α_{ii}^*:

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

Changes in welfare depend on changes 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, ϵ_{ii}, that affects internal trade is small.[2]

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 of the distribution where Pareto and log-normal differ most strikingly.

Shifting to the last equality in (2), welfare falls

[1] Two 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.

[2] By "partial" we mean that incomes and price indices are held constant as in a gravity equation estimated with origin and destination fixed effects.

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 $dM_i^e = 0$.[3] 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. To build intuition on when and why departing from Pareto matters, we investigate the simplest possible case, the two-country symmetric version of the model described by Melitz and Redding (2013).

II. Calibration of the symmetric model

To consider the case of two symmetric countries of size L, set $\tau_{ni} = \tau_{in} = \tau$, $\tau_{ii} = 1$, $f_{ii} = f_d$, $f_{ni} = f_{in} = f_x$. We know from (1) that the domestic cutoff, $\alpha_{ii}^* = \alpha_d^*$ is the sole endogenous determinant of 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:

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

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

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

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. Following (1), the GFT simplifies to the ratio of domestic cutoffs, autarkic over openness cases: $\mathbb{T}_i = \alpha_{dA}^*/\alpha_d^*$. The domestic cutoff in autarky is obtained by restating the free entry condition as $f_d \times G(\alpha_{dA}^*)\left[\mathcal{H}(\alpha_{dA}^*) - 1\right] = f^E$.

The last step is therefore to specify $G(\alpha)$.

---

[3]See the working paper version of ACR for the proof.

Pareto-distributed productivity $\varphi \equiv 1/\alpha$ implies a power law CDF for α, with shape parameter θ. A log-normal distribution of α retains the log-normality of productivity (with location parameter μ and dispersion parameter ν) but with a change in the log-mean parameter from μ to $-\mu$. The CDFs for α are therefore given by

$$(5) \qquad G(\alpha) = \begin{cases} \left(\frac{\alpha}{\bar{\alpha}}\right)^\theta & \text{Pareto} \\ \Phi\left(\frac{\ln\alpha+\mu}{\nu}\right) & \text{Log-normal}, \end{cases}$$

where we use Φ 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 α_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, $G(\alpha_x^*)/G(\alpha_d^*)$ 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: θ in Pareto and ν 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 ϵ_x: 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 τ_0. Indeed, **M4** $= \frac{\tau_0^{1-\sigma}}{1+\tau_0^{1-\sigma}}$, which Melitz and Redding (2013) take

as 0.14 from US exporter data. Setting $\sigma = 4$, we have $\tau_0 = ([(1 - \mathbf{M4})/\mathbf{M4}])^{1/3} = 1.83$.

Two parameters still need to be set: the CES σ, 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 ($\bar{\alpha} \equiv 1/\underline{\varphi}$ in Pareto or μ in log-normal) and a shape parameter (θ or ν) governing heterogeneity. For the trade elasticities and GFT, location parameters do not matter whereas heterogeneity (falling with θ and rising with ν) 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 α 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($\underline{\varphi}, \theta$) implies that the sales of firms in any given market will be distributed Pareto($\tilde{\varphi}, \tilde{\theta}$), 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 $\tilde{\nu} = (\sigma - 1)\nu$. Estimating $\tilde{\theta}$ and $\tilde{\nu}$, and postulating a value for σ, we can back out estimates of θ and ν.

We estimate $1/\tilde{\theta}$ and $\tilde{\nu}$ 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_i 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 $\mathbb{Q}_i^E = \ln x_i$ 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:

$$(6) \quad F_{\mathrm{P}}(\ln x) = 1 - \exp[-\tilde{\theta}(\ln x - \ln \underline{x})],$$

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

$$(7) \quad F_{\mathrm{LN}}(\ln x) = \Phi((\ln x - \tilde{\mu})/\tilde{\nu}).$$

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:

$$(8) \quad \mathbb{Q}_i^{\mathrm{P}} = F_{\mathrm{P}}^{-1}(\hat{F}_i) = \ln \underline{x} - \frac{1}{\tilde{\theta}} \ln(1 - \hat{F}_i),$$

$$(9) \quad \mathbb{Q}_i^{\mathrm{LN}} = F_{\mathrm{LN}}^{-1}(\hat{F}_i) = \tilde{\mu}/\tilde{\nu} + \tilde{\nu}\Phi^{-1}(\hat{F}_i).$$

The QQ estimator regresses the empirical quantile, $\mathbb{Q}_i^E$, on the theoretical quantiles, $\mathbb{Q}_i^{\mathrm{P}}$ or $\mathbb{Q}_i^{\mathrm{LN}}$. Thus, the heterogeneity parameter $\tilde{\nu}$ of the log-normal distribution can be recovered as the coefficient on $\Phi^{-1}(\hat{F}_i)$. The primitive productivity parameter ν is given by $\tilde{\nu}/(\sigma - 1)$. 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 $\theta = (\sigma - 1)\tilde{\theta}$. 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 exporters we use a leading destination: Belgium for French firms and Japan for Chinese ones.

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: $\tilde{\nu}$	2.392	2.344	2.409	2.468	2.450	2.447	2.457	2.486
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: $1/\tilde{\theta}$	2.146	1.390	1.174	0.915	0.884	0.855	0.822	0.779
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

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

FIGURE 1. QQ GRAPHS

(a) French firms → Belgium   (b) Chinese firms → Japan
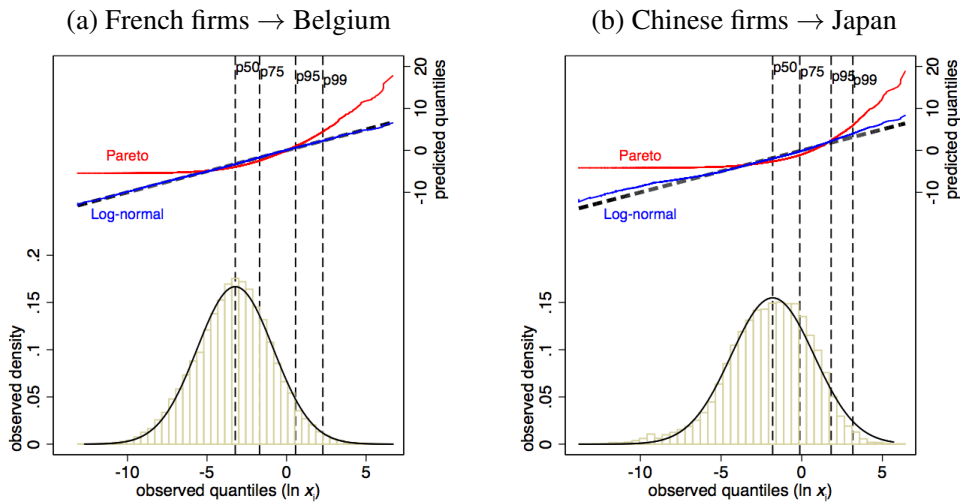
The precise mapping between productivity and sales distributions only holds for individual destination markets. Nevertheless, we also show in the appendix that the total sales distribution for French and Spanish firms follow distributions that resemble the log-normal more than the Pareto. As the theory 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 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. With the value of $\theta = 4.25$ used by Melitz and Redding (2013) in mind, we choose the top 1% estimates as our benchmark: that is $\theta = 3.849$ and $\nu = 0.797$ for the French exporters case, and $\theta = 4.854$ and $\nu = 0.853$ 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 $\tau_0/2 < \tau < 2\tau_0$, with τ_0, 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.

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_{dA}^* rises faster than α_d^*. 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
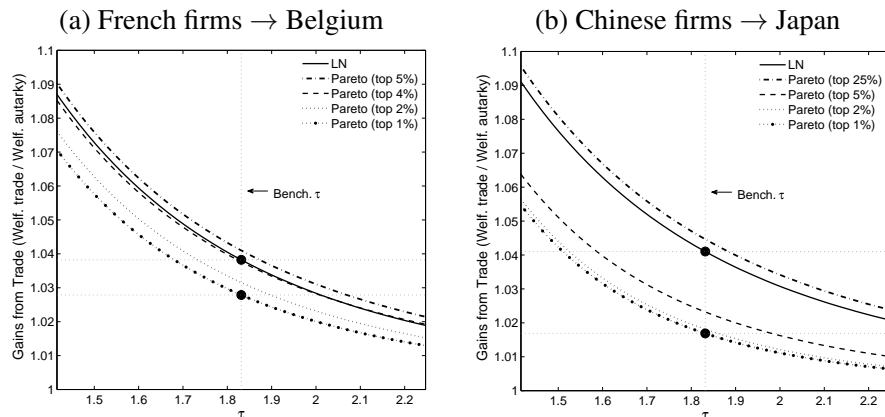
data a bell-shaped PDF suggests that more than half the potential entrants are choosing to operate (otherwise we would face a strictly declining PDF). As a conservative estimate, we therefore set **M1**=0.5 as our benchmark.

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 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.

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 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, 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 **M1** (ENTRY RATE)

(a) French firms → Belgium

(b) Chinese firms → Japan
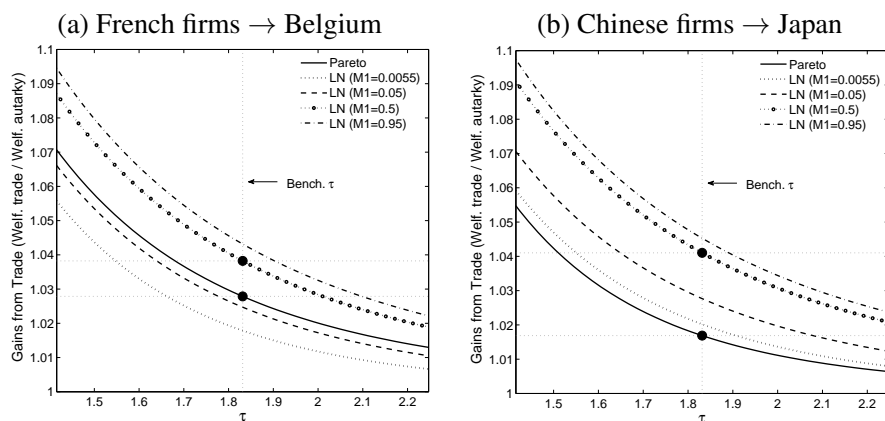
FIGURE 3. WELFARE GAINS, SENSITIVITY TO **M3** (TRUNCATION)

(a) French firms → Belgium

(b) Chinese firms → Japan

REFERENCES

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodriguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

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

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

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quarterly Journal of Economics*, 123(2): 441–487.

Kratz, Marie, and Sidney I Resnick. 1996. "The QQ-estimator and heavy tails." *Stochastic Models*, 12(4): 699–724.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Melitz, Marc J., and Stephen J. Redding. 2013. "Firm Heterogeneity and Aggregate Welfare." National Bureau of Economic Research Working Paper 18919.

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

ONLINE APPENDIX

A1. 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, Costinot and Rodriguez-Clare (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_{ni}^e, the mass of entrants from i into destination n, and the mean export revenues of exporters from i serving market n.

$$(A1) \qquad X_{ni} = G(\alpha_{ni}^*) M_i^e \int_0^{\alpha_{ni}^*} \frac{x_{ni}(\alpha) g(\alpha) d\alpha}{G(\alpha_{ni}^*)},$$

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

$$(A2) \qquad x_{ni}(\alpha) = (\bar{m}\alpha w_i \tau_{ni})^{1-\sigma} P_n^{\sigma-1} Y_n,$$

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

Following Helpman, Melitz and Rubinstein (2008), it is useful to define

$$(A3) \qquad V_{ni} = \int_0^{\alpha_{ni}^*} \alpha^{1-\sigma} g_i(\alpha) d\alpha.$$

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

$$(A4) \qquad X_{ni} = M_i^e Y_n (\bar{m} w_i \tau_{ni})^{1-\sigma} P_n^{\sigma-1} V_{ni}, \qquad \text{with} \qquad P_n^{1-\sigma} \equiv \sum_\ell M_\ell^e (\bar{m} w_\ell \tau_{n\ell})^{1-\sigma} V_{n\ell}.$$

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:

$$(A5) \qquad \pi_{ni} = M_i^e L_i^{\sigma-1} Y_i^{1-\sigma} (\bar{m} \tau_{ni})^{1-\sigma} V_{ni} P_n^{\sigma-1},$$

with

$$(A6) \qquad P_n^{1-\sigma} = \sum_\ell M_\ell^e L_\ell^{\sigma-1} Y_\ell^{1-\sigma} (\bar{m} \tau_{n\ell})^{1-\sigma} V_{n\ell}.$$

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:

$$(A7) \qquad \alpha_{ni}^* = \sigma^{1/(1-\sigma)} \left(\frac{L_i}{Y_i}\right)^{\sigma/(\sigma-1)} \left(\frac{Y_n}{f_{ni}}\right)^{1/(\sigma-1)} \frac{P_n}{\bar{m}\tau_{ni}}.$$

Welfare in this model is given by real income. Inverting equation (A7), welfare can be expressed in

terms of the domestic cutoff:

$$(A8) \qquad \mathbb{W}_i \equiv \frac{Y_i}{P_i} = \left(\frac{L_i}{\sigma}\right)^{\sigma/(\sigma-1)} \frac{\sigma-1}{\tau_{ii} f_{ii}^{1/(\sigma-1)}} \frac{1}{\alpha_{ii}^*}.$$

This is equation (1) in the main text. Since α_{ii}^* is the sole endogenous variable, a change in international trade costs implies that $\frac{d\mathbb{W}_i}{\mathbb{W}_i} = -\frac{d\alpha_{ii}^*}{\alpha_{ii}^*}$. The next step is to relate changes in the cutoff to changes in trade shares. To do this we divide both sides of equation (A6) by $P_n^{1-\sigma}$, and differentiate, to obtain:

$$(A9) \qquad \sum_\ell \pi_{n\ell} \left[\frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\frac{dY_\ell}{Y_\ell} + \frac{dV_{n\ell}}{V_{n\ell}} + (\sigma-1)\frac{dP_n}{P_n}\right] = 0$$

Analyzing the dV/V term first, we can see from the definition in equation (A3) 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

$$(A10) \qquad \gamma_{ni} \equiv \frac{d\ln V_{ni}}{d\ln \alpha_{ni}^*} = \frac{\alpha_{ni}^{*2-\sigma} g(\alpha_{ni}^*)}{\int_0^{\alpha_{ni}^*} \alpha^{1-\sigma} g(\alpha) d\alpha}.$$

From the definition of V and equilibrium cutoffs in (A7), we can write the change in V as

$$(A11) \qquad \frac{dV_{n\ell}}{V_{n\ell}} = \gamma_{n\ell}\frac{d\alpha_{n\ell}^*}{\alpha_{n\ell}^*} = \gamma_{n\ell}\left[\frac{1}{\sigma-1}\frac{dY_n}{Y_n} - \frac{\sigma}{\sigma-1}\frac{dY_\ell}{Y_\ell} + \frac{dP_n}{P_n} - \frac{d\tau_{n\ell}}{\tau_{n\ell}}\right].$$

Combining (A9) and (A11) leads to

$$(A12)$$
$$\sum_\ell \pi_{n\ell}\left[\frac{dM_\ell^e}{M_\ell^e} + (1-\sigma-\gamma_{n\ell})\left(\frac{d\tau_{n\ell}}{\tau_{n\ell}} - \frac{dP_n}{P_n}\right) + \left(1-\sigma-\frac{\sigma\gamma_{n\ell}}{\sigma-1}\right)\frac{dY_\ell}{Y_\ell} + \frac{\gamma_{n\ell}}{\sigma-1}\frac{dY_n}{Y_n}\right] = 0$$

Differentiating bilateral trade shares in equation (A5),

$$(A13) \qquad \frac{d\pi_{n\ell}}{\pi_{n\ell}} = \frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\frac{dY_\ell}{Y_\ell} + \frac{dV_{n\ell}}{V_{n\ell}} + (\sigma-1)\frac{dP_n}{P_n},$$

$$(A14) \qquad \frac{d\pi_{nn}}{\pi_{nn}} = \frac{dM_n^e}{M_n^e} + (1-\sigma)\frac{dY_n}{Y_n} + \frac{dV_{nn}}{V_{nn}} + (\sigma-1)\frac{dP_n}{P_n}.$$

Hence, the difference in those share changes gives

$$(A15) \qquad \frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} = \frac{dM_\ell^e}{M_\ell^e} + (1-\sigma)\frac{d\tau_{n\ell}}{\tau_{n\ell}} + (1-\sigma)\left[\frac{dY_\ell}{Y_\ell} - \frac{dY_n}{Y_n}\right] + \frac{dV_{n\ell}}{V_{n\ell}} - \frac{dV_{nn}}{V_{nn}}.$$

Let us focus now in the difference in V term. From (A11), we can write:

$$\frac{dV_{n\ell}}{V_{n\ell}} - \frac{dV_{nn}}{V_{nn}} = \gamma_{n\ell}\frac{d\alpha_{n\ell}^*}{\alpha_{n\ell}^*} - \gamma_{nn}\frac{d\alpha_{nn}^*}{\alpha_{nn}^*}$$

$$= \gamma_{n\ell}\left[\frac{1}{\sigma-1}\frac{dY_n}{Y_n} - \frac{\sigma}{\sigma-1}\frac{dY_\ell}{Y_\ell} - \frac{d\tau_{n\ell}}{\tau_{n\ell}} + \frac{dP_n}{P_n}\right]$$

$$- \gamma_{nn}\left[-\frac{dY_n}{Y_n} + \frac{dP_n}{P_n}\right].$$

$$(A16) \qquad = (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} + \gamma_{n\ell}\left[\frac{\sigma}{\sigma-1}\left(\frac{dY_n}{Y_n} - \frac{dY_\ell}{Y_\ell}\right) - \frac{d\tau_{n\ell}}{\tau_{n\ell}}\right].$$

We then plug (A16) into (A15) to obtain

$$\frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} - (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} = \frac{dM_\ell^e}{M_\ell^e} + (1 - \sigma - \gamma_{n\ell})\frac{d\tau_{n\ell}}{\tau_{n\ell}}$$

$$(A17) \qquad\qquad\qquad + \left(1 - \sigma - \frac{\sigma\gamma_{n\ell}}{\sigma-1}\right)\left[\frac{dY_\ell}{Y_\ell} - \frac{dY_n}{Y_n}\right].$$

Therefore the term in square brackets inside (A12) is equal to

$$(A18) \qquad \frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} - (\gamma_{n\ell} - \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} + (1 - \sigma - \gamma_{n\ell})\left[\frac{dY_n}{Y_n} - \frac{dP_n}{P_n}\right].$$

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 (A12) to obtain

$$(A19) \qquad \sum_\ell \pi_{n\ell}\left[\frac{d\pi_{n\ell}}{\pi_{n\ell}} - \frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} + (\sigma - 1 + \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*}\right] = 0$$

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

$$(A20) \qquad -(\sigma - 1 + \gamma_{nn})\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} = -\frac{d\pi_{nn}}{\pi_{nn}} + \frac{dM_n^e}{M_n^e} + \sum_\ell \pi_{n\ell}\frac{d\pi_{n\ell}}{\pi_{n\ell}}$$

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

$$(A21) \qquad \frac{d\mathbb{W}_n}{\mathbb{W}_n} = -\frac{d\alpha_{nn}^*}{\alpha_{nn}^*} = \frac{-d\pi_{nn}/\pi_{nn} + dM_n^e/M_n^e}{(\sigma - 1 + \gamma_{nn})},$$

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

A2. *How **M1** (entry share) affects welfare in the symmetric model*

Under the trading regime, our 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

$$(A22) \qquad \frac{d\alpha_d^*/\alpha_d^*}{d\mathbf{M1}/\mathbf{M1}} = \frac{G(\alpha_d^*)}{\alpha_d^* G'(\alpha_d^*)} > 0,$$

$$(A23) \qquad \frac{d\alpha_x^*/\alpha_x^*}{d\mathbf{M1}/\mathbf{M1}} = \frac{G(\alpha_x^*)}{\alpha_x^* G'(\alpha_x^*)} > 0,$$

Simple manipulations of the differentiated system also yields

$$(A24) \qquad \frac{df_x}{f_x} = (\sigma - 1) \times \left[\frac{G(\alpha_d^*)}{\alpha_d^* G'(\alpha_d^*)} - \frac{G(\alpha_x^*)}{\alpha_x^* G'(\alpha_x^*)} \right] \times \frac{d\mathbf{M1}}{\mathbf{M1}},$$

$$(A25) \qquad \frac{df^E}{f^E} = A_1^+ \frac{d\alpha_d^*}{\alpha_d^*} + A_2^+ \frac{d\alpha_x^*}{\alpha_x^*} + A_3^+ \frac{df_x}{f_x},$$

where (A_1^+, A_2^+, A_3^+) are positive parameters. Looking at the Pareto version of definition (5), it is clear that $\frac{G(\alpha_d^*)}{\alpha_d^* G'(\alpha_d^*)} - \frac{G(\alpha_x^*)}{\alpha_x^* G'(\alpha_x^*)} = 0$, which means that the right hand side of (A24) is zero under Pareto. Therefore, a change of $\mathbf{M1}$ 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 (5)). Hence in the LN case,

$$(A26) \qquad \frac{df_x/f_x}{d\mathbf{M1}/\mathbf{M1}} \geqslant 0.$$

Combined with (A22), equations (A23), (A24) and (A25) thus imply that

$$(A27) \qquad \frac{df^E/f^E}{d\mathbf{M1}/\mathbf{M1}} > 0.$$

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

$$(A28) \qquad \frac{d\alpha_{dA}^*/\alpha_{dA}^*}{d\mathbf{M1}/\mathbf{M1}} > 0$$

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

$$(A29) \qquad \frac{d\mathbb{T}}{\mathbb{T}} = \left[\frac{d\alpha_{dA}^*}{\alpha_{dA}^*} - \frac{d\alpha_d^*}{\alpha_d^*} \right] \times \frac{d\mathbf{M1}}{\mathbf{M1}}$$

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.

A3. *Distribution parameters for Chinese exports to Japan*

Table A1 replicates Table 1 for the case of Chinese exports to Japan in 2000.

A4. *Distributions of total sales*

Some of the prior literature asserting Pareto is based on firm size distribution, 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)).

The mapping between productivity distribution parameters and sales distributions is less clear when considering total sales of firms (domestic sales plus exports to all destinations). Chaney (2013)

TABLE A1—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.

and Di Giovanni et al. (2011) are examples using total exports and sales, respectively, for French firms. Both papers truncate the samples. In Tables A2 and A3, and figure A1, we corroborate 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.

TABLE A2—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.

### A5. Comparison of QQ estimator to other methods

One alternative to the QQ estimators is to use method of moments. In this case, we infer the distributional parameters from the means and standard deviations of log sales. We can use equations (6) and (7) 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 $\ln \tilde{\varphi} + 1/\tilde{\theta}$, and a standard deviation of $1/\tilde{\theta}$. In this sample, the standard deviation of log sales is 2.393, hence predicted coefficients in Table 1 are 2.393 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.

TABLE A3—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 A1. QQ GRAPHS ON TOTAL SALES

(a) French firms   (b) Spanish firms
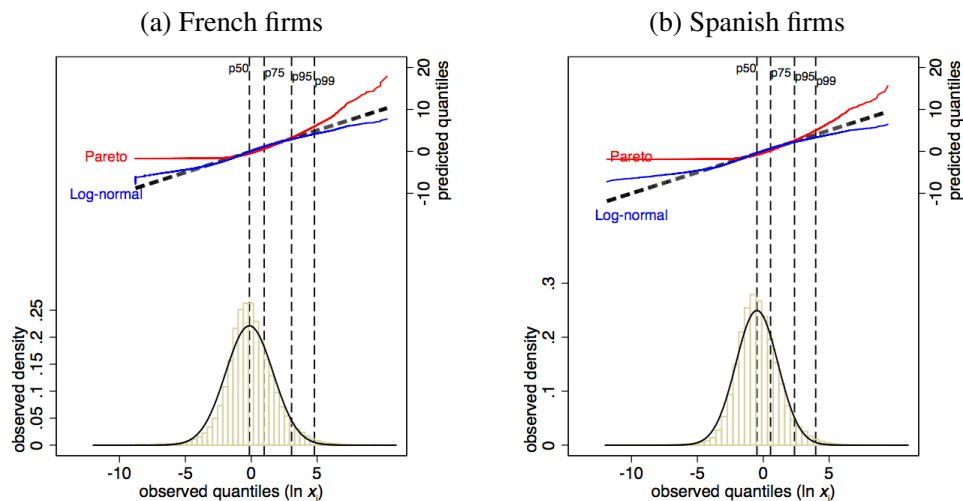
There is a close relationship between the QQ estimator for the Pareto and the familiar log rank-size regressions examined by Gabaix and Ioannides (2004) since both rank, $1 + (n - i)$, and one minus the empirical CDF are linear in i. This closely resembles the QQ estimator since, following the suggestion of Bury (1999), we estimate the empirical CDF as $\hat{F}_i = (i - 0.3)/(n + 0.4)$. Thus, the empirical CDF is an affine transformation of the rank. 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**.

A6. Macro-data simulations

In this section, we adopt the **M3′** approach where the underlying micro parameters ν and θ are calibrated to match the international 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. In this case, the calibration procedure will therefore select values for f^E, f_x and ν such that target values for **M1**, **M2**, and **M3′** are matched.

The most obvious empirical target value for **M3′** (recommended by Arkolakis, Costinot and Rodriguez-Clare (2012)) comes from estimates of 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 A2 plots the GFT as in figures 2 and 3, 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}. 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 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 some parameter targets, in particular the share of exporters. Third, this calibration searches for parameters 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 in figure A3, the Pareto and log-normal GFT again diverge from each other.

A7. 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

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

FIGURE A2. WELFARE GAINS CALIBRATED ON TRADE ELASTICITY
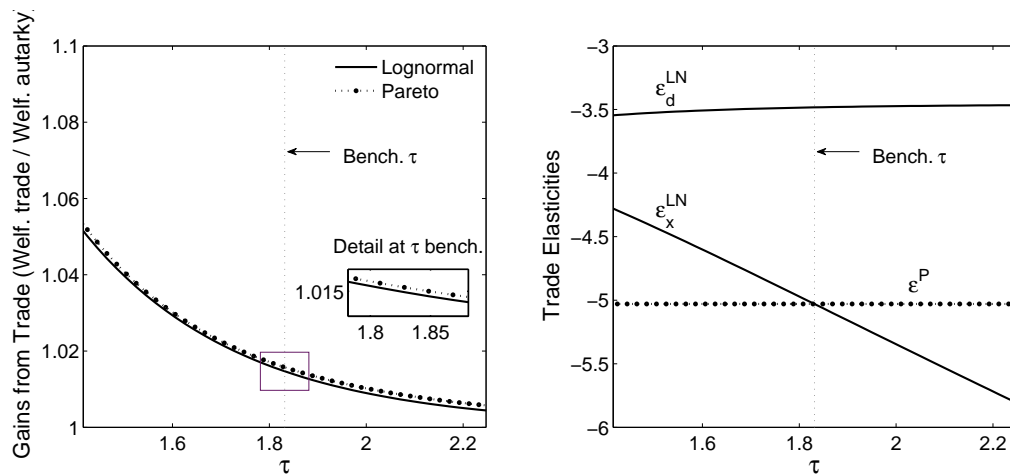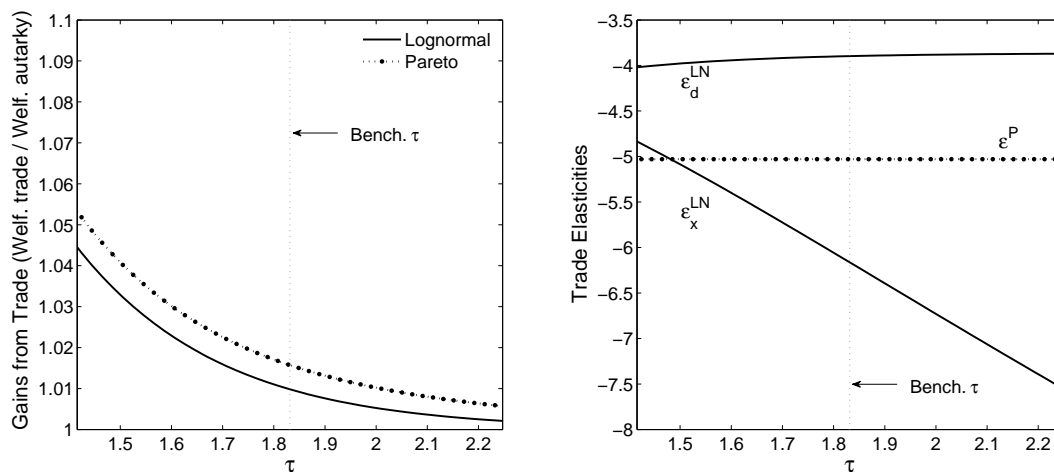

FIGURE A3. WELFARE GAINS CALIBRATED ON AVERAGE TRADE ELASTICITY

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.

<center>*</center>

Appendix References

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

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.

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

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.

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

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

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.