American Economic Review: Papers & Proceedings 2014, 104(5): 310–316
http://dx.doi.org/10.1257/aer.104.5.310

Welfare and Trade without Pareto[†]

By Keith Head, Thierry Mayer, and Mathias Thoenig*

Heterogeneous firm papers which 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 Rodríguez-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..."

* Head: University of British Columbia, Sauder School of Business, 2053 Main Mall, Vancouver, B.C., V6T 1Z2, Canada, and CEPR (e-mail: keith.head@sauder.ubc.ca); Mayer: Sciences-Po, 28 rue des Saints-Pères, 75007 Paris, France, CEPII, and CEPR (e-mail: thierry.mayer@sciencespo.fr); Thoenig: Faculty of Business and Economics, University of Lausanne, Bâtiment Extranef, 1015 Lausanne, Switzerland, and CEPR (e-mail: mathias.thoenig@unil.ch). 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 with data, Jonathan Eaton, Andrés Rodrígues-Clare, and Arnaud Costinot for valuable insights, Marc Melitz and Stephen Redding for sharing code, and the Douanes Françaises for data.

[†] Go to http://dx.doi.org/10.1257/aer.104.5.310 to visit the article page for additional materials and author disclosure statement(s).

[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 procompetitive effects of trade.

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

I. Welfare Theory

We assume constant elasticity of substitution (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 online 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 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 $\hat{\mathbb{W}}_i = \hat{\pi}_{ii}^{1/\epsilon}$, where "hats" denote proportional 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:

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

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 $f_d \times G(\alpha_{dA}^*)[\mathcal{H}(\alpha_{dA}^*) - 1] = f^E$.

The last step is therefore to specify $G(\alpha)$. Pareto-distributed productivity $\varphi \equiv 1/\alpha$ implies a power law cumulative distribution function (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

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

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 Four 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 online 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.

[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.

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

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 microdata approach, on which we concentrate in the main text. The second, which we call **M3′** is the trade elasticity ϵ_x: the macrodata approach, covered in the online 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, $\mathbf{M4} = \frac{\tau_0^{1-\sigma}}{1+\tau_0^{1-\sigma}}$, 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 ($\overline{\alpha} \equiv 1/\underline{\varphi}$ 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 ($\underline{\varphi}$, θ) 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 $\tilde{\nu} = (\sigma - 1)\nu$. 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:

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

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

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

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:

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

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

The QQ estimator regresses the empirical quantile, $\mathbb{Q}_i^E$, on the theoretical quantiles, $\mathbb{Q}_i^P$ or $\mathbb{Q}_i^{LN}$. 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

TABLE 1—PARETO VERSUS 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%
Observations	34,751	17,376	8,688	1,737	1,390	1,042	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

Notes: 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 $\sigma = 4$.

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 online 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 sets of exporters we use a leading destination: Belgium for French firms and Japan for Chinese ones. The precise mapping between productivity and sales distributions only holds for individual destination markets. Nevertheless, we also show in the online 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 (middle rows) and Pareto (bottom rows) 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 percent of the variation in the untruncated empirical quantiles, compared to 80 percent for Pareto. In the log-normal case the slope coefficient remains stable even as increasingly high shares of small exporters are removed. This is 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 online 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. Microdata 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 untruncated 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 percent 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.

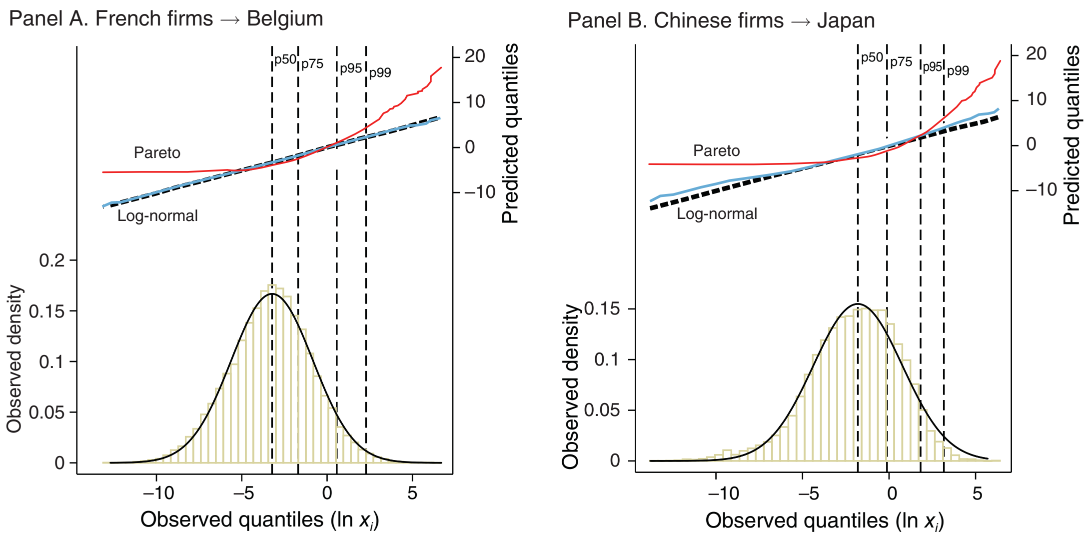

FIGURE 1. QQ GRAPHS

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 online Appendix shows that the impact of a rise in **M1** on GFT is in general ambiguous, depending on relative rates of changes 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 probability distribution function (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 online Appendix), we calibrate heterogeneity parameters on the macrodata 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

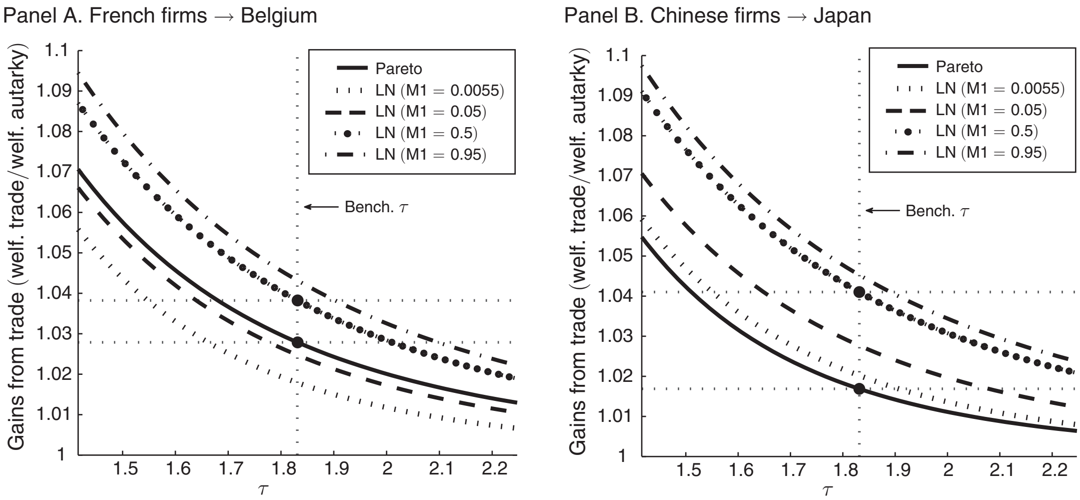

FIGURE 2. WELFARE GAINS, SENSITIVITY TO **M1** (*entry rate*)

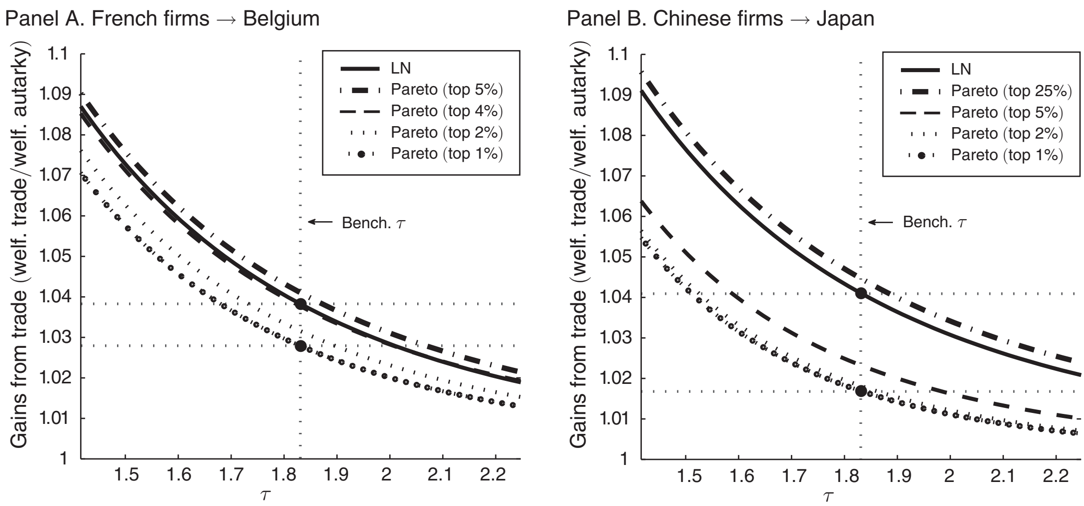

FIGURE 3. WELFARE GAINS, SENSITIVITY TO **M3** (*truncation*)

the distributional assumption. The microdata method points to large GFT differences when the macrodata method points to very similar welfare outcomes.

Which calibration should be preferred? ACR make a compelling case for the macrodata 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 misspecified 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 nonconstant.

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, Costas, Arnaud Costinot, and Andrés Rodríguez-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* 94 (5): 1429–51.

Feenstra, Robert C. 2013. "Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity." Unpublished.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quarterly Journal of Economics* 123 (2): 441–87.

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): 271–82.