Pareto Distributions in International Trade: Hard to Identify, Easy to Estimate.*

Marnix Amand†
HEC Lausanne

Florian Pelgrin‡
EDHEC Business School

November 13, 2016

Abstract

We show that, in heterogeneous-firm international trade models, common forms of heterogeneity and uncertainty drive a (multiplicative random) wedge between the observable exports distribution and the latent distribution of firm productivity. Even if the latter is exactly Pareto distributed, this wedge, correlated with firm productivity, distorts the exports data, often making it look log-normally distributed. We show this distortion to be quantitatively relevant, meaning that empirical evidence of a log-normal exports distribution does not preclude an underlying Pareto distribution for productivity. Furthermore, this wedge renders common maximum-likelihood and quantile-based estimators misspecified, hence inconsistent. We provide general conditions in a broad class of international trade models under which, despite this misspecification issue, the tail of the exports distribution can be used to estimate the power exponent of the productivity distribution consistently—provided the sample size of left-truncated export data is large enough.

Keywords: Pareto, power law, international trade, productivity distribution, misspecification.
JEL Classification Numbers: F12, F14.

*We thank Pascal St-Amour, Tony Smith, Mathias Thoenig, Thierry Mayer, Keith Head and the seminar and conference participants at HEC Lausanne, EDHEC Business School, Toulouse School of Economics, IAAE 2016 and EEA-ESEM 2016 for their feedback and insights, and Costas Arkolakis for an early discussion on the basic insight underlying this paper.

†Corresponding author. HEC Lausanne - University of Lausanne, Quartier Dorigny - Internef 524, CH-1015 Lausanne, Switzerland. Email: marnix.amand@gmail.com.

‡EDHEC Business School, 24 avenue Gustave Delory, 59057 Roubaix, France. Email: florian.pelgrin@edhec.edu.

1 Introduction

The assumption that firm productivity follows a Pareto distribution[1] has become widely used in theoretical work that builds on the Melitz (2003) heterogeneous firm international trade model. The main reason is tractability. The Pareto distribution is "scale-free", in the sense that its shape is invariant to left-truncation. With this assumption the shape of the productivity distribution of firms that export to a given destination does not depend on the exact value of the minimal productivity cut-off. For instance, if one assumes a Pareto distribution for firm productivity, trade between France and Germany on one hand and France and The Philippines on the other hand has the exact same structure, up to a rescaling depending on the size and distance of the export destination. In empirical work, this unknown rescaling factor can easily be eliminated through the use of country-pair fixed effects (see Head and Mayer, 2014, for an extensive literature review) and in theoretical work by concentrating on elasticities, as in Chaney (2008). It even allows for a micro-founded explanation of the gravity equation (Chaney, 2015; Arkolakis, Costinot, Donaldson, and Rodrìguez-Clare, 2015). Eaton, Kortum, and Kramarz (2011) show that the Pareto law assumption allows for a very sparse parametrization of a generalized version of the Melitz model such that it lends itself well to a structural estimation.[2]

All these results depend crucially on assuming a Pareto distribution for firm productivity. In the Melitz model, the aggregate effects of moving a cut-off depend on where exactly this cut-off is located on the productivity distribution. Hence, fixed cut-off's generically make the computation of trade elasticities and welfare effects difficult, and only a Pareto distribution

[1] A random variable X follows a Pareto distribution if its counter-cumulative distribution is a power law, i.e., $\forall x > x_{\min}, \mathbf{P}[X > x] = (x/x_{\min})^{-\alpha}$, where $x_{\min}$ is the lower bound of the distribution and α is called the Pareto, or power law, exponent. A random variable X follows a power law distribution, also called power law tail or Pareto tail, if its counter-cumulative distribution is a power law times a slowly-varying function, where f is defined as a slowly-varying function if and only if $\forall a > 0, \lim_{x\to\infty} \frac{f(ax)}{f(x)} = 0$.

[2] See footnote 22 in Arkolakis, Costinot, and Rodrìguez-Clare (2012) for a list of the main papers using this assumption in the international trade literature.

allows to solve this issue elegantly.

However, this model predicts that destination-specific exports are also Pareto distributed if productivity is, and recent empirical research has shown that the exports distribution is better modeled by a log-normal distribution, with possibly a power law right tail, than by an exact Pareto law. Head, Mayer, and Thoenig (2014) in particular argue that the empirical evidence rejects the Pareto assumption for productivity and show that this matters along an economically significant dimension, welfare, and thus cannot be ignored. In this respect, the empirical justification for the use of a Pareto distribution for productivity in the Melitz (2003) framework is increasingly seen as questionable, notwithstanding the Pareto's very useful properties for international trade models.

The motivation of this paper is to offer an alternative interpretation of the empirical evidence on exports, precisely one that allows us to reconcile the Pareto law assumption for productivity with an exports distribution that is shaped differently, and to estimate the underlying Pareto distribution using this (non-Pareto shaped) exports data.

We look at the problem as follows. We have exports data Y at hand and need to be informed about the underlying distribution X of productivity, which is not directly observable. One needs to resort to a structural intermediary to tease out information about productivity from the distribution of firm exports. We write this problem as

$$Y = \Omega \cdot X \tag{1}$$

where Ω is a multiplicative stochastic wedge of unknown distribution, possibly correlated with X.

We argue that little is known in practice about Ω, it is a nuisance factor, which rules out the distributional assumptions that deconvolution methods require.[3] Conversely, without any further structure on Ω, Y is useless as such to learn anything about X. For any strictly positive random variable X, an Ω can be found to accommodate the data Y. So identification assumptions are necessary. In the canonical Melitz (2003) model, one has (up to an irrelevant exponentiation and proportionality factor) $Y = X$, i.e. $\Omega \equiv 1$. Then of course, there are neither estimation nor identification problems, Y and X can be used interchangeably. However, it is a very strong claim to say that the (usually unobserved) productivity distribution is exactly identified and matched with available export data.[4] This one-to-one relationship between exports data Y and productivity X is weakened as soon as one introduces an element of heterogeneity to the Melitz model. Doing this drives a "wedge" between observed exports and the underlying productivity distribution. We do not need to take a strong stance on the nature of the deviation Ω from the Melitz model for our results. Quite the opposite, we build on the Melitz model and provide several reduced-form stochastic wedges, without making any strong assumptions on the underlying distribution of these wedges. It turns out that any such wedge (or any combination) boils down to a non-degenerate Ω.[5]

This is the first take-away point of our paper. The identification of the productivity distribution in the canonical Melitz model relies on very strong assumptions that are essential—identification fails if they are relaxed—but knife-edge and therefore difficult to justify in practice. One needs to be willing to assume away *all* forms of heterogeneity (other than productivity), uncertainty and plain measurement error to use exports directly to characterize

[3]If one were willing to assume that the distribution of Ω belongs to a given family and that it is independent from X, the estimation of the Pareto law exponent of X is a deconvolution problem written in multiplicative form. Maximum-likelihood methods would yield consistent estimates of the parameters of the distribution of X. If one cannot assume that Ω is independent from X, estimating X by way of deconvolution methods is a problem that has barely been studied in the literature (Meister, 2009; Carroll, Ruppert, Stefanski, and Crainiceanu, 2006).

[4]See, for instance, the complexity of the econometrics used in the industrial organization literature to obtain reliable estimates of productivity, a seminal example being Olley and Pakes (1996).

[5]The idea that the link between productivity and exports must be distended is not novel as such. Arkolakis (2010) has previously solved for an explicit example of a non-degenerate Ω (see equation 23 in his paper). Our approach is more general and not geared towards explaining a given set of empirical facts.

the productivity distribution. Adding any kind of heterogeneity, uncertainty or measurement error breaks down the tight one-to-one link between exports and productivity, and exports need no longer be Pareto distributed even if productivity is, and might well look log-normal. We argue that it is therefore not surprising but quite natural that exports have failed to be Pareto distributed over their whole range, this says very little about the productivity distribution.

To illustrate this, let us give a simple introductory example: an international trade model where firms have differing productivity and face heterogeneous fixed market access costs. Compared to the productivity distribution, the bottom of the exports distribution will be polluted by missing firms (i.e., high-enough productivity firms that "should" be exporting but had an unlucky draw of market access costs) and the presence of firms that "should not be there" (i.e., low productivity firms with a particularly favorable draw of market access costs). This distorts the shape of the exports distribution away from the productivity distribution.

What this example shows is that if the data is likely to be distorted by a non-degenerate Ω, X is no longer identified by Y and standard econometric techniques that ignore Ω are obviously misspecified. This issue is potentially very severe: one knows very little about unmodelled heterogeneity, misspecification of the underlying model and measurement error, to name but a few issues, so it is problematic to make precise assumptions about Ω. Furthermore, our model and our introductory example show that it is hard to rule out that Ω is not independent from X.

This matters in practice. Using QQ-plots, as is often done in the trade literature, we show that exports data simulated by the estimated model of Eaton, Kortum, and Kramarz (2011)—which uses a Pareto law for productivity—fits a log-normal almost perfectly and does very poorly when fitted to a Pareto distribution. In Appendix B, we use a formal testing procedure provided by Malevergne, Pisarenko, and Sornette (2011) and show that indeed, quite often a log-normal distribution is a better fit than a Pareto distribution for data simulated with a Pareto distribution for productivity X but with a non-degenerate wedge Ω.

This is our second take-away: in the presence of a wedge Ω, even a perfectly Pareto distributed X can yield observable data Y that "looks" log-normal, which is indeed what can be observed in practice. Our result allows to reconcile the Pareto law assumption for productivity with the empirical evidence that shows exports only have a power-law tail. In the presence of a wedge such as Ω, a power-law tail in the data is exactly what one would expect from Pareto distributed productivity X.

Finally, coming back to the misspecification issue when estimating the exponent of the Pareto distribution of X, the previous empirical literature has either ignored the issue and used exports data for the estimation procedure while relying on Melitz (2003) to make statements about productivity (Head, Mayer, and Thoenig, 2014, for instance), or used an explicit model to link these two, then relying on calibration (Arkolakis, 2010) or on a structural estimation (Eaton, Kortum, and Kramarz, 2011). In all cases, this comes down to making an explicit assumption on the link between productivity and exports (i.e., an explicit assumption on Ω). Our last point is that such restrictive assumptions can be avoided.

At an intuitive level, one would think that, even if productivity is not exactly identified by exports, surely exports are driven by productivity to a large enough extent that the former is informative about the latter even in the presence of some unknown Ω. Using a theorem from a companion paper (Amand and Pelgrin, 2016), we claim that a consistent estimation of a Pareto distribution for X is still possible using data Y, provided the sample size is large enough and the assumptions of our theorem are verified. Besides a technical condition that avoids a degenerate case, this requires two assumptions that have a simple economic interpretation in the trade literature, namely that Ω is not "too" heavy tailed (less than X) and not "too" correlated with X at the top. This allows one to apply the theorem using an economic justification. The key idea of the proof is to build on the heavy tail of the Pareto productivity distribution X: this tail will come to dominate the shape of the right of the exports distribution Y almost no matter the distribution or dependence-structure of Ω.

The paper is organized as follows. Section 2 briefly reviews the previous empirical work on identification and estimation of the Pareto distribution in the international trade context. Section 3 introduces a very general Melitz-type model with heterogeneous wedges which results in a data structure of the $Y = \Omega X$ type. In Section 4, we illustrate with an example that the presence of Ω makes identification and estimation problematic when the presence of Ω is ignored. Section 5 provides the formal results that allow one to identify and estimate a Pareto distribution despite the presence of this wedge Ω. Section 6 concludes.

2 Pareto distributions in international trade

As stated in the introduction, considerable attention has been given to the empirical justification of the Pareto assumption. The standard Melitz model predicts that the destination-specific exports of a firm are proportional to an exponent of that firm's productivity.[6] Hence the firm distribution of exports, denoted Y in this paper following the notation in (1), is equal to the firm distribution of (an exponent of) productivity, denoted X (including the constant proportionality factor). The implicit assumption here is that $\Omega \equiv 1$. Since a Pareto distribution is invariant by exponentiation (up to a change in exponent), this yields a testable implication: it is necessary and sufficient to observe that the distribution of firm exports

[6] To be precise, $\varphi^{\sigma-1}$ is proportional to destination-specific exports, where $1/\varphi$ is the unit cost of production ("productivity") of the firm and σ is the elasticity of substitution of the CES utility function of consumers.

from an origin country to a target country Y follows a Pareto distribution to conclude that using a Pareto distribution for the firm productivity distribution X in the origin country is indeed supported by the data.[7,8] Thus all one needs to do is to look at the distribution of firm destination-specific exports.

Notably, several papers have shown, still under the implicit assumption that $\Omega \equiv 1$, that the right tail of destination-specific exports does indeed follow a Pareto law (see di Giovanni, Levchenko, and Rancière, 2011, and references therein). It is much harder, though, to find evidence that the *complete* distribution of destination-specific exports follows a Pareto law. Until recently, it was unclear whether these mixed empirical results should be counted in favor or against the Pareto law assumption for X. In an important recent paper using very large micro data sets, Head, Mayer, and Thoenig (2014) show that the *complete* distributions of exports from France to Belgium and China to Japan fit a log-normal distribution much better than a Pareto law. Following Melitz (2003), this implies a log-normal distribution for firm productivity. The authors show that, from a theoretical perspective, this has non-negligible welfare consequences compared to a Pareto law distribution, and the left tail of the distribution (i.e., the smaller firms) matters for this welfare calculation. Lastly, they show that partial-equilibrium trade elasticities do also depend on the choice of the productivity distribution. For a Pareto law, these are constant, for a log-normal distribution they are not and hence estimation methods must be rethought (Bas, Mayer, and Thoenig, 2015). Said differently, these authors show that one cannot simply assume a Pareto law *in lieu* of a log-normal distribution for analytical convenience: the choice of the distribution to model firm

[7] di Giovanni, Levchenko, and Rancière (2011) make the important point that this reasoning does *not* apply to the total sales (or total exports) of these firms, which generically do not follow a Pareto law even if firm productivity does.

[8] Note that the exponent $\sigma - 1$ plays no important role, switching between φ and $\varphi^{\sigma-1}$ is straightforward. In particular, this exponent does not change the nature of a Pareto law or a log-normal law. This justifies that we drop all reference to this exponent in our explanations and make claims such as "productivity is equal to exports", by which we mean $Y = X$, instead of writing "an exponent of the productivity distribution is proportional to destination-specific exports". This is commonly done in the literature through a change of variables (in particular, see Head, Mayer, and Thoenig, 2014).

productivity heterogeneity matters along important dimensions. Their empirical evidence indicates firm exports to be log-normally distributed over the whole range of exports and they show that only considering the top of the distribution when evaluating the empirical relevance of the Pareto distribution is not a valid shortcut.

Subsequently, one direction of research has been to reconcile the empirical evidence that exports Y seem to be either log-normally distributed or, at best, power-law distributed in the right tail only with the assumption that productivity X is Pareto distributed over its whole range. To do this, one needs to break the tight relationship between exports and productivity as predicted by the Melitz model by arguing that exports have other determinants than just productivity. This is what Arkolakis (2010) does by adding "market penetration costs" that weigh heavier on larger firms. This allows for the existence of smaller exporting firms, which yields a firm export distribution that has a right power-law tail but a density that decreases at a lower rate than a Pareto density for smaller firms and which can even be hump-shaped, depending on parameter values. Interestingly, trade elasticities remain constant across destinations despite the non-Pareto aspect of exports. Although our paper shares a common purpose with Arkolakis (2010), we differ along two dimensions. We allow for more flexibility in the shape of the exports distribution whereas Arkolakis (2010) predicts an exports distribution that has a closed form solution. Second, we take a reduced-form approach with a more empirical emphasis.[9]

Lastly, a recent trend in the literature is to assume a right-truncated (i.e., bounded) Pareto law for productivity. Using a right-truncated Pareto law distribution, Helpman, Melitz, and Rubinstein (2008) obtain a gravity equation in trade that is consistent with the observed zero trade flows between countries. Feenstra (2014) uses a bounded Pareto law distribution and a novel class of preferences to build a tractable heterogeneous firm model with two additional

[9]The Arkolakis (2010) model is calibrated in the original paper and is structurally estimated by Eaton, Kortum, and Kramarz (2011).

"gains-of-trade margins", an expansion of product variety and a pro-competitive reduction in mark-ups, that are neutralized in the Melitz model. Capitalizing on our results, we discuss in Appendix E some issues regarding the identification and estimation of a bounded Pareto law for productivity in the presence of a wedge Ω.

3 Observed exports and productivity: a general model

In this section, our goal is to justify the presence of a non-degenerate Ω in the data. To do this, we provide a very general Melitz-type model in which observed exports Y are related to the productivity distribution X as in equation (1), with Ω correlated with X. In Appendix A, we provide an additional model where Ω is independent from X.

Setup. Our starting point is the canonical Melitz model with endogenous variable market access costs following Arkolakis (2010). We follow the standard notation of Melitz and Redding (2014), dropping the index i used for the home country as we are not interested in general equilibrium here and do not need to distinguish between home countries. The consumer side is CES with elasticity of substitution σ. The wage level is normalized to 1 and all costs are expressed in terms of domestic (i.e., exporting-country) wages. The firm side consists of firms producing horizontally-differentiated goods, each with firm-specific productivity φ. The variable production cost of producing quantity q is $\frac{q}{\varphi}$. Firms export from the home country to country n (possibly the same country). Firms are risk-neutral, and to exist, a firm has to pay f_E and subsequently draws its productivity φ from a known distribution. To export to country n, a firm has to pay an additional market access cost f_n composed of a fixed cost and a variable cost that increases with the proportion of the market targeted by the exporting firm. By assumption, firms must serve the domestic market before exporting.[10]

We introduce three sources of ex-ante destination- and firm-specific heterogeneity in addition to productivity: heterogeneity in fixed market access costs ϵ_{f}, heterogeneity in variable market access costs ϵ_{m} and heterogeneity in demand ϵ_{d}. Specifically, the market access costs are defined as follows: a firm accesses a fraction $m \in [0, 1]$ of a market n by paying the following cost:

$$f_n(m, \epsilon) = \epsilon_{\mathrm{f}} + \epsilon_{\mathrm{m}} \frac{1 - (1 - m)^{1-1/\lambda}}{1 - 1/\lambda}. \tag{2}$$

and the total market size (i.e., demand) for this firm in the target country is $\epsilon_{\mathrm{d}} R_n$ with

$$\mathbf{E}[\epsilon_{\mathrm{f}}] = \bar{\epsilon}_{\mathrm{f}}$$

$$\mathbf{E}[\epsilon_{\mathrm{m}}] = \bar{\epsilon}_{\mathrm{m}}$$

$$\mathbf{E}[\epsilon_{\mathrm{d}}] = 1.$$

The last equation is justified by the fact that total demand in country n accross all products is R_n. These heterogeneities are denoted collectively as ϵ, are all known in advance and vary per destination n for a given firm.[11]

[10] It is easy to endogenize this assumption by redefining the fixed cost slightly. It is clearer though to expose the model as is done here, with domestic sales and exports playing identical roles.

[11] It is straightforward to add (many) more sources of heterogeneity than just three, say, heterogeneity in marginal production costs, heterogeneity in iceberg costs, the need to pay a tax/bribe proportional to sales that varies per firm, etc. However, this does not add generality. A firm has only three degrees of liberty: whether to enter a market, what size of the market to target (m) and how much to sell (q). These three decisions are driven by three equations, an inequality (positive net profits) and two first-order conditions (profit maximization for m and q, given downward-sloping demand). So two firms can only differ along a maximum of three dimensions (in addition to productivity). Note that any one type of heterogeneity is enough to obtain our results, we introduce all three types to cover all possibilities. We have chosen these three heterogeneities such that the algebraic solutions are simple and (mostly) linear. Lastly, note that, aside from possible general equilibrium dimensions, our model contains the original Melitz (2003) model, the Arkolakis (2010) model, the Chaney (2008) model, the di Giovanni, Levchenko, and Rancière (2011) model and the Eaton, Kortum, and Kramarz (2011) model.

Optimal strategy for an exporting firm. Given the CES demand structure, the market-clearing price p charged by a firm selling q to fraction m of market n is given by the demand curve:

$$q = m\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} p^{-\sigma} \tag{3}$$

with P_n the standard CES price index. To streamline notation, we introduce the variable $\overline{\pi}_n$:

$$\overline{\pi}_n = \frac{1}{\sigma} R_n P_n^{\sigma-1} \tau^{1-\sigma}$$

which is such that $\overline{\pi}_n \varphi^{\sigma-1}$ is the optimal profit before market access costs (i.e., sales minus production and iceberg costs, but not marketing or fixed entry costs) while keeping demand heterogeneity at its expected value of 1. Notice that $\overline{\pi}_n$ only depends on parameters of the model, and not on any firm heterogeneity. With this new notation, we now have the following results for a firm's optimal policy:

$$\begin{aligned} p_n(\varphi,\epsilon) &= \frac{\sigma}{\sigma-1}\frac{\tau}{\varphi} \\ q_n(\varphi,\epsilon) &= m_n(\varphi,\epsilon)\epsilon_{\mathrm{d}}\frac{\varphi(\sigma-1)}{\tau}\overline{\pi}_n\varphi^{\sigma-1}, \\ \epsilon_{\mathrm{m}}\big(1-m_n(\varphi,\epsilon)\big)^{-\frac{1}{\lambda}} &= \epsilon_{\mathrm{d}}\overline{\pi}_n\varphi^{\sigma-1}. \end{aligned}$$

Notice that, as usual, firms charge a fixed mark-up above marginal cost. This results in sales in profits as follows:

Sales:	$r_n(\varphi,\epsilon) = \epsilon_{\mathrm{d}}\sigma m_n(\varphi,\epsilon)\overline{\pi}_n\varphi^{\sigma-1}$	(4)
Profits before market access costs:	$\pi_n(\varphi,\epsilon) = \epsilon_{\mathrm{d}} m_n(\varphi,\epsilon)\overline{\pi}_n\varphi^{\sigma-1}$	(5)

Entry decision. Firms enter a market n only if their profits $\pi_n(\varphi, \epsilon)$ minus market access costs are positive. Given π_n is strictly increasing in φ for all ϵ, the entry condition is

$$\varphi \geq \varphi_n(\epsilon) \qquad \text{with} \qquad \pi_n\big(\varphi_n(\epsilon), \epsilon\big) = f_n(m_n(\varphi, \epsilon), \epsilon),$$

where $\varphi_n(\epsilon)$ is the minimal productivity threshold. Lastly, the initial entry decision is

$$\mathbf{E}\left[\sum_{\substack{n \\ \varphi \geq \varphi_n(\epsilon)}} \pi_n(\varphi, \epsilon)\right] \geq f_E. \tag{6}$$

Taking both the optimal sales condition and the entry condition, we are now in position to write the model in the form $Y = \Omega X$. Let $\delta_n(\varphi, \epsilon)$ be defined as

$$\delta_n(\varphi, \epsilon) = \begin{cases} 1 & \text{if} \quad \varphi \geq \varphi_n(\epsilon) \\ 0 & \text{if not.} \end{cases} \tag{7}$$

Then for each draw (φ, ϵ), observed sales are:

$$r_n(\varphi, \epsilon) = \delta_n(\varphi, \epsilon)\epsilon_{\mathrm{d}} m_n(\varphi, \epsilon)\sigma\overline{\pi}_n\varphi^{\sigma-1}. \tag{8}$$

Over the space of realizations (φ, ϵ), let Y denote the random variable equal to the function r_n, X denote the random variable $\varphi^{\sigma-1}$ and Ω denote the random variable $\delta_n(\varphi, \epsilon)\epsilon_{\mathrm{d}} m_n(\varphi, \epsilon)\sigma\overline{\pi}_n$. Then, in the notation of our framework, sales are distributed as $Y = \Omega X$. Note that Ω is not independent of X and does not behave as a measurement error. It is composed of three stochastic processes: a "selection" process $\delta_n(\varphi, \epsilon)$, a "market share" process $m_n(\varphi, \epsilon)$ and a "market size" process ϵ_{d}. The first two are clearly dependent on X, both the decision to export and the decision to target a certain market size depend on productivity. Any of these three sources of heterogeneity is enough for a non-degenerate Ω.[12]

Welfare and trade elasticities. Interestingly, if one assumes that the fixed entry cost ϵ_{f} is 0, our model boils down to the Eaton, Kortum, and Kramarz (2011) model without any functional assumptions, and it is straightforward to show that the results from Arkolakis, Costinot, and Rodrìguez-Clare (2012) apply: if productivity is Pareto distributed, aggregate trade elasticities are constant and welfare changes can be computed using a simple formula. This illustrates again the importance (and convenience) of the Pareto assumption for productivity.

4 The Pareto/log-normal debate in the presence of misspecification: a numerical example

In order to shed some light on past empirical work, we now turn to the econometric implications of the general formulation $Y = \Omega X$ in the case where the presence of Ω is ignored by or unknown to the econometrician.

A QQ-plot exploration. To highlight the misspecification issue, we assume that Ω and X are correlated and distributed as in Eaton, Kortum, and Kramarz (2011). Specifically, we set the parameter of the Pareto law generating $\varphi^{\sigma-1}$ at 2.46, $\lambda = 0.91$, $\epsilon_{\mathrm{f}} = 0$ and assume that $\ln \epsilon_{\mathrm{d}}$ and $\ln \frac{\epsilon_{\mathrm{d}}}{\epsilon_{\mathrm{m}}}$ are joint normally distributed with variances of resp. 1.69 and 0.34 and correlation -0.65. These values come directly from Eaton, Kortum, and Kramarz (2011). We proceed with an horse-race between the log-normal and the Pareto distribution to best

[12]Note that even a stochastic ϵ is not necessary. Arkolakis (2010) introduces a market share process with finite non-zero λ to obtain a non-degenerate Ω. This is enough to distort Y towards log-normality using a Pareto-distributed X. Results are not reported here but are available upon request.

fit Y. To do this, we use a QQ-regression as suggested in the trade literature (Head, Mayer, and Thoenig, 2014). This comes down to visually comparing the QQ-plot of the best-fitting member of each family of distributions with the 45 degree line.[13] Note however that there are no tabulated test-statistics in the literature that allow to assess the goodness-of-fit of a QQ-plot or compare competing QQ-plots.

To avoid small-sample bias, we simulate $200,000$ draws of Y—a sample size somewhat larger than those most often encountered in international trade data. Then we estimate the log-normal and Pareto law distribution that best fit Y using a QQ-regression and show the QQ-plots in the top panel of Figure 1. Moreover, we also report the QQ-plot of Y using the true distribution of X. Following the literature, the conclusion is straightforward: Y is much closer to the best-fitting log-normal distribution (in blue) than to the best-fitting distributional Pareto law (in red), which is not the actual distribution of X (in green). As mentioned earlier, Head, Mayer, and Thoenig (2014) apply this procedure using actual firm-level data on exports from France to Belgium and from China to Japan and come to the same conclusion: the log-normal distribution is a far better fit than the Pareto law for their exports distributions.

The fact that Y "looks" log-normal and not Pareto distributed on a QQ-plot is driven by the thick base of the Pareto law. At the left, the Pareto law packs a lot of data points close to the minimum (the log of the minimum is 0.00, the log of the median is 1.33). This means that most of the variation on the left will be driven by Ω, even if it has a small standard deviation itself. Hence, even for a small-variance Ω, the bottom of the distribution of Y is essentially

[13] For a given dataset, a QQ-plot can help assess the goodness of fit of a candidate distribution by plotting the theoretical quantiles of the distribution v. the empirical quantiles of the data. A perfect fit would be the 45 degree line: each quantile of the candidate distribution aligns perfectly with the observed quantile in the data. The QQ-estimator (or QQ-regression) minimizes the distance (in the sense of least squares) between the 45 degree line and the theoretical QQ-plots of a parametric family of distributions. See Kratz and Resnick (1996), and Schultze and Steinebach (1996) for an extended explanation of the link between QQ-plots and the QQ-estimator, Head, Mayer, and Thoenig (2014) for a first use in the international trade literature and Section C in the appendix for a brief summary.

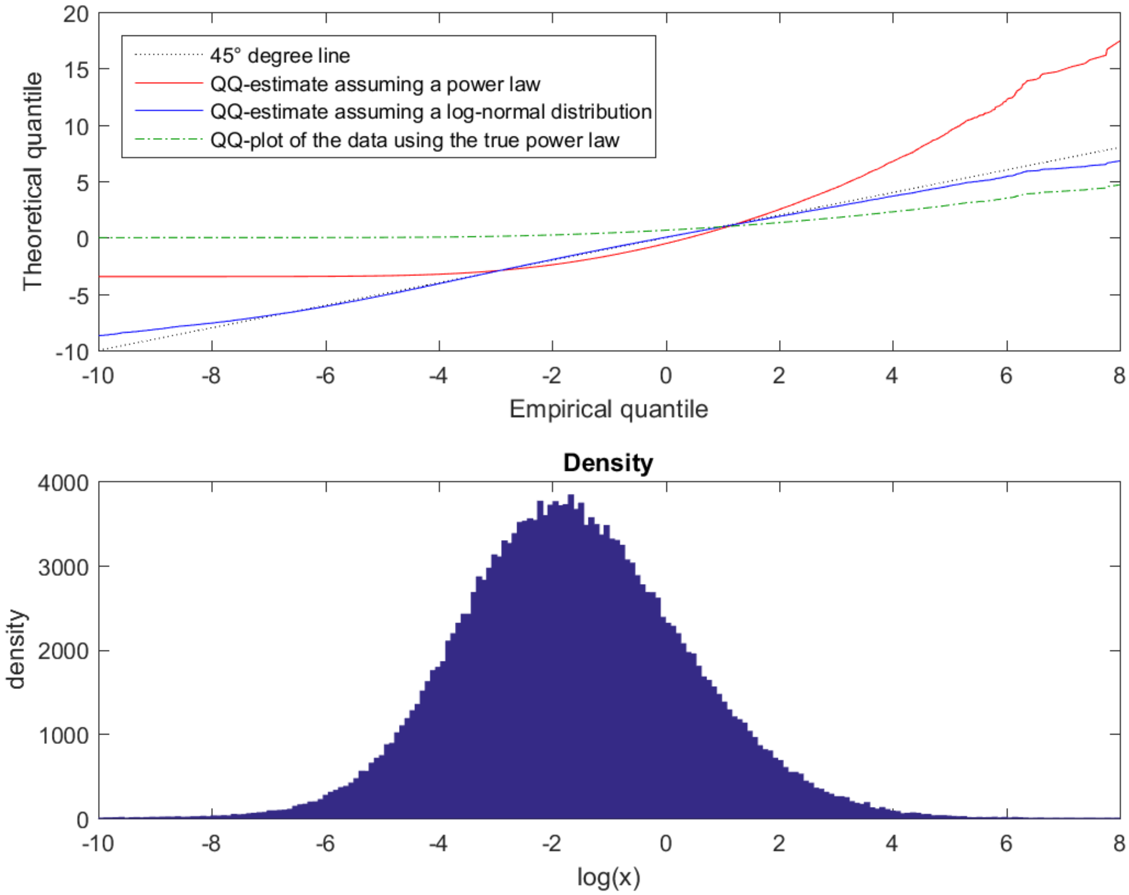

Figure 1: QQ-plots of data $Y = \Omega X$ generated by a Pareto law for X and a correlated Ω wedge as in Eaton, Kortum, and Kramarz (2011). See Section 4 for the exact data generating process.

shaped by Ω. This is what the left part of the QQ-plot picks up. Furthermore, the best fitting Pareto law (in red) is not the true Pareto law of X (in green), indicating the inconsistency of the QQ-regression. The conclusion of this numerical example is that mistaking Y for X can lead to conclude that X is probably log-normally distributed and, should one overcome this hurdle and model X and Y as Pareto laws, to consider the parameter of the Pareto law that best fits Y as a consistent estimate of the Pareto law underlying X, which is wrong. In practice, unless one is willing to assume the absence of Ω, this means one should not treat the failure to identify a Pareto law in Y (i.e., in exports) as a reason for rejecting the hypothesis of a Pareto distribution for productivity.

A formal test. One should note that our example is not driven by small sample bias. In Figure 4 in Appendix D, we show that our results are identical if one uses a very large sample of $20,000,000$ data points. Furthermore, these results hold even with a Pareto law distribution having infinite variance (i.e., with an exponent less than 2). Lastly, these results are not particular to the QQ-plot approach. To assess the reliability of the visual inspection, we conduct the uniformly most powerful unbiased test proposed by Malevergne, Pisarenko, and Sornette (2011).[14] This test is known as the Wilks' test and can be viewed as a likelihood ratio test in which the Pareto distribution is considered as a "limit case" of the log-normal distribution. The test proceeds as follows. In a first step, one needs to find the optimal threshold such that the profile (composite) likelihood (for the whole sample) of the maximum likelihood estimates of the distribution parameters is maximized. In a second step, the clipped sample coefficient of variation is used and a critical threshold (to reject the null hypothesis) can be obtained by a saddle point approximation (the method used in

[14]One may be tempted to look for tests that test Y directly for a given distribution (say, Pareto or log-normal). But without a distributional assumption on Ω, all tests would be misspecified by nature. This includes a formal goodness-of-fit test (e.g., Anderson-Darling) for Y; such a test would most likely reject a log-normal and a Pareto law, whether bounded or unbounded, given the presence of Ω (it does for our data). Lastly, note that, in any case, a QQ-plot horse race is not a valid statistical test.

our experiments) or by Monte Carlo simulations. In this respect, the results show that the Pareto distribution cannot be identified in Y even if X is exactly Pareto distribution as in our data generating process. Especially, there is strong evidence in favor of the log-normal distribution.[15]

Estimating a power law exponent using Y. Ignoring for one moment the identification issues raised in the preceding paragraphs, can Y be used to estimate the power law coefficient of X? We concentrate on four very common estimators: two versions of maximum likelihood, namely the unconditional Hill estimator (Hill, 1975) and the conditional Hill estimator (Aban, Meerschaert, and Panorska, 2006), the log-size log-rank regression (Gabaix and Ibragimov, 2011) and the QQ-regression (Kratz and Resnick, 1996; Schultze and Steinebach, 1996). We recapitulate these estimation methods in Appendix C. Misspecification implies that it is not econometrically valid to use Y to infer the parameter values of X's Pareto law, this leads to inconsistent (not merely biased) estimates. We illustrate this by running a Monte-Carlo simulation; we estimate the power law exponent of Y on $5,000$ independently drawn datasets drawn from the same data generating process as in the previous section with an attenuated Ω for expositional purposes.[16] Moreover, since a well-known recommendation of previous empirical work is to work with only the top of the distribution, we also run each estimation procedure again on left-truncated data by progressively dropping more and more (from 10% to 99.9%) of the leftmost data points. The results are shown in Figure 2.

If the goal is to estimate the power law exponent of X, it is clear from these results that all estimators are inconsistent when the data is Y and not X. However, our simulations

[15] Other results of this test are provided in Appendix B, where we look at different parameter values in the specific case of an uncorrelated Ω, i.e., measurement errors. Appendix A provides an alternative theoretical justification for the relevance of this case.

[16] Specifically, we set $\epsilon_{\mathrm{f}} = 0$ and the parameter of the Pareto law generating $\varphi^{\sigma-1}$ at 2.46 as before, but with $\lambda = 10$ and assume that $\ln \epsilon_{\mathrm{d}}$ and $\ln \frac{\epsilon_{\mathrm{d}}}{\epsilon_{\mathrm{m}}}$ are joint normally distributed with variances of resp. 0.4 and 0.1 and correlation -0.65. We do this strictly for the sake of the clarity of the graphics: the misspecification is actually worse with the original values.

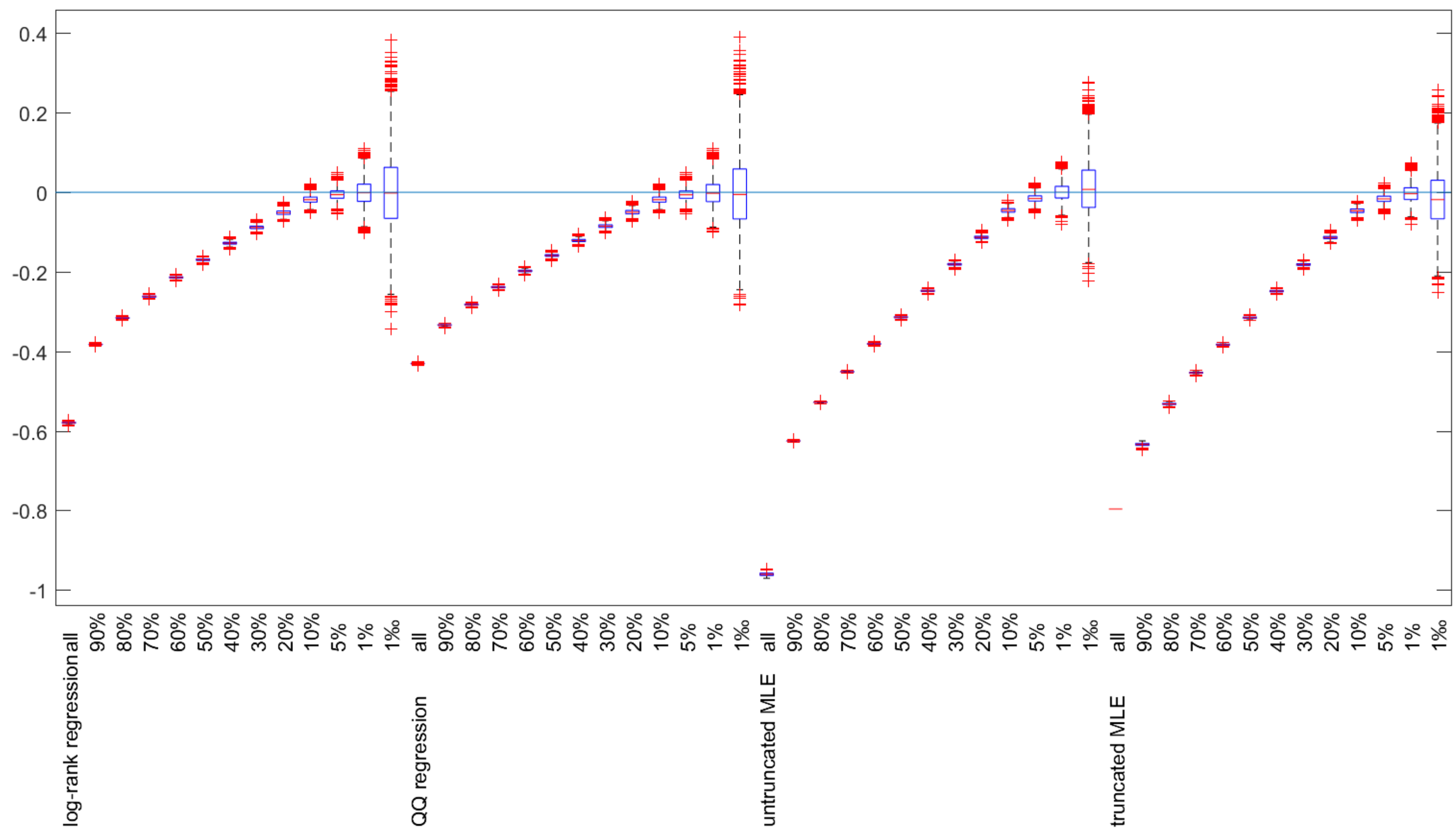

Figure 2: Monte-Carlo simulations of estimates of the power law exponent.

Notes: Boxplot of 5,000 estimates of the power law exponent α according to four estimation procedures (see main text), at different levels of left-truncation. Each estimation is done on data Y that is generated by a Pareto law for X and a correlated Ω wedge as in Eaton, Kortum, and Kramarz (2011). See footnote 16 for the exact data generating process. The size of each draw is $200,000$.

Results are centered and normalized around the true value of α: $\dfrac{\widehat{\alpha}-\alpha}{\alpha}$ is reported.

also illustrate that left-truncating the data (i.e., dropping the lowest points) does allow for consistent estimates, which is actually a general result (see Section 5). Note that this is done in practice by many papers without further justification.

5 Estimation and identification

The preceding section illustrates that Ω causes both estimation and identification issues. In this section, we ask these two questions more generally and formally. Once one has concluded that the data Y is distributed as in (1), with Ω determined by some version of our model, can the tail of Y be used to estimate X if one is willing to assume that X is exactly Pareto distributed? Second, if Y has a power law tail, can we use this to conclusively identify a power law for X?

5.1 Estimation

Regarding estimation, we build on a theorem result from our companion paper, Amand and Pelgrin (2016), where we show the following result.

Theorem 1. *Let $Y = \Omega X$, with X a Pareto distributed random variable on $[x_{min}, +\infty)$ with exponent α, and Ω a random variable on $\mathbb{R}^+$ with $\Phi_{\Omega|X}(\cdot|x)$ denoting the conditional counter-cumulative given $X = x$. If there exist constants $C > 0$ and $\kappa > 0$ and a function Φ_0 such that:*

$$\begin{array}{lll} \forall \omega \geq 0, \quad \forall x \geq x_{min} & \Phi_{\Omega|X}(\omega|x) < C\omega^{-\alpha-\kappa} & \textit{(thin-tailed condition)} \\ \forall \omega \geq 0 & \lim\limits_{x\to\infty} \Phi_{\Omega|X}(\omega|x) = \Phi_0(\omega) & \textit{(pseudo-independence condition)} \\ \exists \omega' > 0 & \Phi_0(\omega') > 0 & \textit{(non-degeneracy of } \Phi_0\textit{)} \end{array}$$

then there exists a non-random sequence k_n such that for any sequence of independent random draws $\{Y_1, \ldots, Y_n, \ldots\}$ of Y, we have

$$\lim_{n\to\infty} \frac{k_n}{n} = 0$$

$$\operatorname*{plim}_{n\to\infty} \hat{\alpha}_{k_n,n} = \alpha$$

where $\hat{\alpha}_{k_n,n}$ is any of the four most common estimators of α of Appendix C computed using the k_n highest order statistics of the n first observations.

The interpretation is that, provided Ω fulfills the conditions of the theorem, each estimator can get arbitrarily close (in probability) to the true value of α by just using the right tail (the highest order statistics) as long as one has a sufficiently large sample. Intuitively, this theorem allows one to estimate the power law exponent of X using Y as long as one assumes a underlying model that generates an appropriate Ω. The first two assumptions relate to the "top" of the conditional distribution of Ω; as far as the left end of Ω is concerned, anything is possible. The third assumption is there to avoid situations in which Φ_0 is degenerate. This means that no assumptions on the economic explanation or the exact structure of Ω are needed to apply this theorem, one can stay entirely agnostic about the causes of the presence of Ω. As long as Ω satisfies the three conditions, which only concern its limiting behavior, the estimates of the power law exponent α using the tail of Y will be consistent.[17]

The economic interpretation of the three conditions is as follows. The first condition is that, whatever the productivity draw ($X = x$ in our notation), the distribution of the wedge for a certain type of productivity x is thinner-tailed than X. The second condition is that, for sufficiently high values of productivity, the wedge is "almost" independent of productivity. Intuitively, these two conditions mean respectively that the top of the exports distribution is

[17]Note that the first two conditions are related to the concept of (quasi-) asymptotic independent random variables, see Resnick (2002, 2007) and Maulik, Resnick, and Rootzén (2002).

not polluted by low productivity firms that had an extremely favorable draw of Ω, and that we do not need to worry that the tail of Y is distorted by interactions between Ω and X. In our model, equation (8) shows that both market share and the selection process are bounded by 1. Thus the first condition is true as long as the demand heterogeneity is thinner tailed than X.[18] The second condition is also true given that shocks are either independent from x (such as market size) or converge to 1 (such as market access) for high productivity. The third condition is in this particular case merely technical and always verified.

More generally, how realistic is it that Ω fulfills all three criteria? Ω would need to be a wedge that is thicker tailed than productivity to contradict the first assumption. An example would be fixed market access costs that are thicker-tailed distributed than productivity. Then, the tail of the exports distribution will identify the tail of the fixed-cost distribution and productivity plays the role of the wedge, but it seems natural to assume that perturbations are thinner-tailed than a power law. To contradict the second assumption, Ω would have to not converge point-wise for $x \to \infty$. An example of the second case would be a wedge that is very different for very similar high-productivity firms. It is hard to come up with a realistic economic example of a cost structure where this is the case. To contradict the third assumption, Ω would have to converge point-wise to 0 for $x \to \infty$. This would be the case if the Pareto distribution is not truncated but the data is truncated by Ω. We deal with a truncated Pareto distribution X in Appendix E. It is possible to construct an example where X would be *untruncated* and Y truncated, but this would require assumptions (say, decreasing returns to scale and an upper bound to production) that are at a large distance from standard international trade models.

Our conclusion is that assumptions of Theorem 1 regarding the wedge Ω are very likely to be satisfied in our generalized version of the Melitz model. Therefore—provided that the

[18] The normal, log-normal, Laplace, exponential and gamma distribution all fulfill this condition.

sample size of the left-truncated export data is large enough—estimates of the exponent of the power-law tail of productivity X are consistent if the right tail of exports Y is used, even if the exports data is perturbated (compared to productivity) by all the heterogeneities we mentioned in Section 3. This absence of bias justifies the approach of theoretical papers that rely on precise estimates of the value of the exponent (e.g., Chaney, 2008, 2015).

5.2 Identification

The last question is identification. To what extent can Y help in identifying some distributional properties of X? The previous theorem provides an immediate corollary:

Corollary 1. *Let $Y = \Omega X$ with X a random variable on $[x_{min}, +\infty)$ and Ω a random variable on $\mathbb{R}^+$ such that Ω fulfils the three conditions of Theorem 1. Then X can only be power-law distributed with exponent α if Y displays a power law tail with the same exponent α.*

According to Corollary 1, if one is willing to make the aforementioned assumptions on the wedge Ω, then the absence of a power-law tail in Y rules out the possibility of a Pareto distribution (or even a power-law tail) in X. We have a "necessary-condition test", one that can rule out Pareto distributions. One should note, however, that this is a weak test: one needs to be sure that the absence of a power law tail in Y is not due to the limited size of the dataset at hand. We show in the next section that this can in practice be the case. Lastly, it is worth noting that sufficient conditions on Y that imply a power law tail for X cannot be obtained when Ω is not independent from X.[19]

[19] See lemma 4.3 in Jessen and Mikosch (2006) and the explanation in Amand and Pelgrin (2016).

5.3 Applications

One application of our theorem is di Giovanni, Levchenko, and Rancière (2011). They show that the firm sales distribution in a country is not Pareto distributed because it is the sum of Pareto-distributed sales in each exporting destination (including the home country), and a sum of Pareto distributions with identical exponent but different cut-offs is not a Pareto distribution. It has a much thicker base, which leads to inconsistent estimates of the Pareto exponent. However, if one defines Ω as the stochastic fraction of the world market targeted by a firm (which will depend on productivity, idiosyncratic market-access costs per destination country, etc.) and X as productivity then it is easy to see that $\lim_{x\to\infty} \mathbf{P}[\,\Omega = 1 \mid X = x\,] = 1$, meaning the largest firms export everywhere. Hence Ω trivially fulfills all conditions of the theorem and the *tail* of the firm size distribution is identical to the tail of the firm productivity distribution, *irrespective* of the fact that more productive firms export more and to more countries.

More generally, a practical question is how high the cut-off should be for the estimates of the Pareto exponent of X to be consistent. Heuristically, a good rule of thumb is to progressively drop more and more data points on the left until the estimates stabilize. Figure 2 illustrates this very well. Formally, using a weighted least squares algorithm, Beirlant et al. (1996) show that tail index estimates can be obtained from estimates of the slope at the right upper tail of a Pareto quantile plot. Notably, algorithms based on the root mean squared errors can be constructed in order to search for the optimal order statistic to the right of which one obtains an optimal linear fit of the quantile plot. There is, however, no guarantee that one has enough data points for the estimates to stabilize before the sample becomes too small. One very telling example is the model of Eaton, Kortum, and Kramarz (2011). Figure 3 shows the estimates of the exponent of X depending on which level of left-truncation is chosen. Even truncating 99.9% of the lowest data points does not yield a consistent estimate.

This means that it is not possible to estimate the power law coefficient of X using a single set of bilateral trade data.[20]

Interestingly, this discussion of Corollary 1 implies that identification of the Pareto distribution itself is actually inconclusive with just one set of bilateral trade data. Head, Mayer, and Thoenig (2014) do not find a power law tail in their French exports data and conclude from a QQ-plot that this rules out a Pareto distribution in productivity. Our results show that their conclusion is indeed formally warranted, and this in a very general setting, i.e. as long as Ω fulfills the conditions of the theorem. They conclude that the log-normal is a better fit for X. An alternative possibility for X would be a truncated Pareto distribution, which generates similar QQ-plots as we show in Appendix E. In any case, their discussion of the welfare implication of X being thinner-tailed than a Pareto distribution remains valid. But a last possible interpretation of Head, Mayer, and Thoenig's (2014) result is that their dataset may not be large enough to rule out the possibility that there is a power tail in their data generating process, only higher than their highest data points. As we have shown in the previous paragraph with our simulations of the (estimated) model of Eaton, Kortum, and Kramarz (2011), this is not an unlikely possibility, even with a large sample size.

6 Conclusion

The main message of this paper is that identifying and estimating the productivity distribution using the Melitz (2003) model and exports data requires knife-edge assumptions. These assumptions do not hold if one allows for heterogeneity or uncertainty in the model or measurement error in the data. Therefore, conclusions about productivity that rely solely

[20] Note that in their paper, the authors use a *complete* set of bilateral trade data, i.e. data on exports of each French firm to every possible destination. Heuristically, this allows for the identification of the power law exponent of φ because heterogeneity in ϵ "averages out" across countries, but heterogeneity in φ does not.

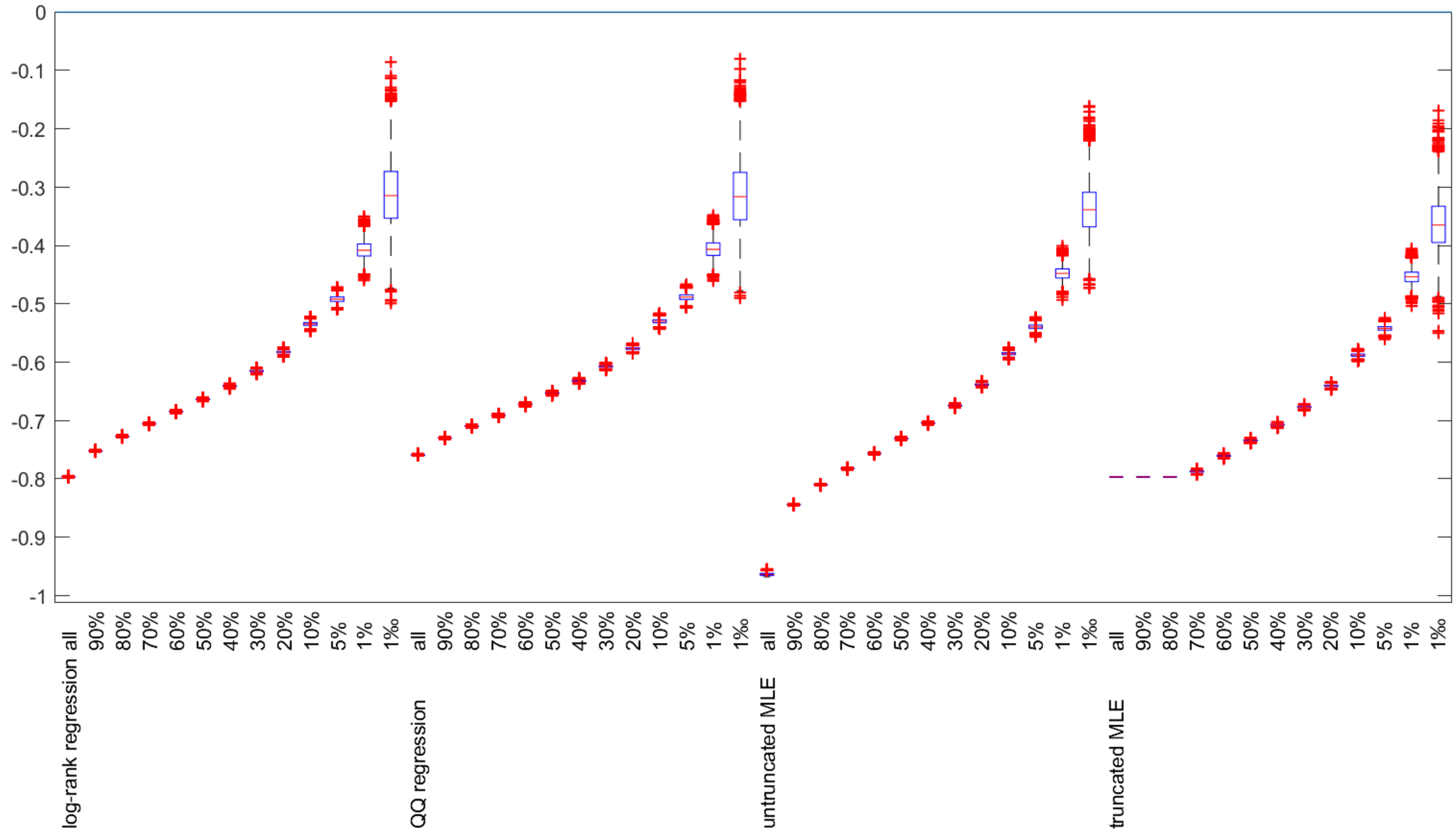

Figure 3: Monte-Carlo simulations of estimates of the power law exponent.

Notes: Boxplot of 5,000 estimates of the power law exponent α according to four estimation procedures (see main text), at different levels of left-truncation. Each estimation is done on data Y that is generated by a Pareto law for X and a correlated Ω wedge as in Eaton, Kortum, and Kramarz (2011). The parametrization and numerical values of the parameters are exactly as in Eaton, Kortum, and Kramarz (2011), see Section 4 for details. The size of each draw is $200,000$.

Results are centered and normalized around the true value of α: $\dfrac{\widehat{\alpha}-\alpha}{\alpha}$ is reported.

on exports data may be more fragile than first thought. However, we also show that, under assumptions that we claim should in practice be very reasonable from an economic viewpoint, it is possible to use the right tail of exports to identify and estimate the right tail of productivity without further knowledge of the wedge Ω. The question of identification (can we conclusively affirm or rule out the existence of a Pareto distribution for productivity?) remains in our view undecidable given the current state of knowledge. Even large datasets do not allow a conclusion to go one way or the other. Importantly, this means that rejecting the assumption of a Pareto distribution based on log-normal exports data is not warranted.

References

Aban, I. B., and M. M. Meerschaert (2004): "Generalized least-squares estimators for the thickness of heavy tails," *Journal of Statistical Planning and Inference*, 119(2), 341–352.

Aban, I. B., M. M. Meerschaert, and A. K. Panorska (2006): "Parameter Estimation for the Truncated Pareto Distribution," *Journal of the American Statistical Association*, 101(473), 270–277.

Amand, M., and F. Pelgrin (2016): "Pareto distributions and (very) imperfect data," Working paper.

Arkolakis, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118(6), 1151 – 1199.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodrìguez-Clare (2015): "The Elusive Pro-Competitive Effects of Trade," NBER Working Papers 21370, National Bureau of Economic Research, Inc.

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

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

Beirlant, J., M. Fraga Alves, M. Gomes, and M. Meerschaert (2014): "Extreme Value Statistics for Truncated Pareto-type Distributions," Discussion paper.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006): *Measurement Error in Nonlinear Models: A Modern Perspective*, vol. 105 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2 edn.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707–21.

——— (2015): "The Gravity Equation in International Trade: An Explanation," Discussion paper, Toulouse School of Economics.

di Giovanni, J., A. A. Levchenko, and R. Rancière (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.

Feenstra, R. C. (2014): "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," NBER Working Papers 19833, National Bureau of Economic Research, Inc.

Gabaix, X., and R. Ibragimov (2011): "Rank - 1 / 2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, 29(1), 24–39.

Head, K., and T. Mayer (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, chap. 3, pp. 131–195. Elsevier.

Head, K., T. Mayer, and M. Thoenig (2014): "Welfare and Trade without Pareto," *American Economic Review*, 104(5), 310–16.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123(2), 441–487.

Hill, B. M. (1975): "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3(5), 1163–1174.

Jessen, A. H., and T. Mikosch (2006): "Regularly varying functions," *Publications de l'Institut Mathématique*, 79(93), 1–23.

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

Malevergne, Y., V. Pisarenko, and D. Sornette (2011): "Testing the Pareto against the Log-normal Distributions with the Uniformly Most Powerful Unbiased Test Applied to the Distribution of Cities," *Physical Review*, E 83(3), 1–11.

Maulik, K., S. I. Resnick, and H. Rootzén (2002): "Asymptotic independence and a network traffic model," *J. Appl. Probab.*, 39(4), 671–699.

Meister, A. (2009): *Deconvolution Problems in Nonparametric Statistics*, vol. 193 of *Lecture Notes in Statistics*. Springer Verlag.

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 (2014): "Heterogeneous Firms and Trade," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, chap. 1, pp. 1–54. Elsevier.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–97.

Resnick, S. I. (2002): "Hidden Regular Variation, Second Order Regular Variation and Asymptotic Independence," *Extremes*, 5(4), 303–336.

——— (2007): *Heavy-Tail Phenomena*, Springer Series in Operations Research and Financial Engineering. Springer Verlag.

Schultze, J., and J. Steinebach (1996): "On least squares estimates of an exponential tail coefficient," *Statistics & Risk Modeling*, 14(4), 353–372.

A A special case: Ω behaving as measurement error

One possible interpretation of Ω in equation (1) is measurement error, there is no reason to assume that the reporting by firms is done flawlessly. If one assumes that the data is distorted by an additive measurement error that is normally distributed of mean 0 and with a standard deviation proportional to the true value X (i.e., the standard deviation is constant if expressed as a percentage of the true value), a first order approximation[21] shows that the observed data Y is $Y = X\Omega$ with Ω distributed log-normally and independent from X.

Within the confines of the Melitz model, measurement error is enough to generate observed exports that are not distributed as a Pareto law even if productivity is. We argue that such a "measurement error interpretation" of Ω can also be justified from a theoretical perspective.[22] We show this by introducing uncertainty about aggregate local demand and uncertainty about iceberg losses to the Melitz model. The story goes as follows. Firms decide whether and how much to export to a given country, and it is only after the entry and the exporting decisions have irrevocably been made that the firm discovers how high or low local demand is and how much of exports have been lost in transit. The firm can then adjust prices to account for the new demand curve and available quantities. This differs from previous literature, which has modelled ex-ante cost and demand heterogeneity (see Eaton, Kortum, and Kramarz, 2011), but not uncertainty.

This approach highlights the fact that the observed sales are the final result of a firm optimizing and re-optimizing over its different control variables sequentially, as it learns more about the environment it operates in. Only one of these optimizing decisions, the first one, which we call "intended sales", is purely driven by productivity and thus very informative on

[21]Let $Y = X + X\epsilon$ with ϵ normally distributed, centered in 0 and with a small variance. Then $\ln Y = \ln X + \ln(1+\epsilon)$ hence $Y \approx Xe^{\epsilon}$.

[22]Note that this does not rule out actual measurement error, which should still be a worry by itself for anyone using observed sales to identify or estimate the productivity distribution. Actual measurement in our set-up would compound multiplicatively with the "theoretical" Ω.

X. All subsequent firm decisions distort the observed data compared to this ideal variable. To be more precise, we show that intended sales are nil below a certain cut-off, and above this cut-off are indeed an exponent of the distribution of firm productivity, exactly as in the Melitz model. But the data does not give us intended exports. Only the realized exports are reported, after firms have re-optimized prices, and we show that, in our specification, realized exports are distributed according to intended sales X multiplied by a distribution Ω (driven by the local demand uncertainty and the iceberg cost shocks) that takes the structure of an independent multiplicative measurement error.

Setup. We again follow the notation of Melitz and Redding (2014). The canonical Melitz model is modified by adding multiplicative uncertainty about the exact value of demand in the target country and multiplicative uncertainty about the iceberg costs. Both of these uncertainties are revealed upon arrival in the destination country, when the quantity decision has already been made but prices can still be adjusted. More specifically, the demand in country n that a firm can access is not exactly aggregate demand R_n and is uncertain: $\epsilon_{\mathrm{d}} R_n$ with $\mathbf{E}[\epsilon_{\mathrm{d}}] = 1$, and the iceberg costs are also uncertain and written as $\epsilon_{\mathrm{s}} \tau_n$ with $\mathbf{E}[\epsilon_{\mathrm{s}}] = 1$. Aside from this uncertainty, iceberg costs have the usual effect: firms decide on how much to export, k, but upon arrival the available quantity diminishes to $q = \frac{k}{\epsilon_{\mathrm{s}} \tau_n}$, which is now uncertain. Only then do firms decide about the selling price p and earn revenue $r = pq$. Note that both sources of uncertainty are firm- and destination-dependent. Without loss of generality we normalize τ_n for n the home country to 1, and it is furthermore assumed that $\tau_n \geq 1$ for all n.

Optimal strategy for an exporting firm. We solve by backward induction. A firm that has already committed to exporting to country n needs to solve for the optimal level of exports k_n (before iceberg costs), bearing in mind it can adjust prices after uncertainty has been revealed.

We proceed with the standard resolution of demand and monopoly pricing in a CES framework. A firm selling a net quantity of exports q given (known) aggregate local demand $\epsilon_{\mathrm{d}} R_n$ sets its price p_n such that:

$$q = \epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} p_n(q, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}})^{-\sigma} \tag{9}$$

with P_n the standard CES price index. Hence for given gross exports k and realized shocks, variable profits are $\frac{k}{\epsilon_{\mathrm{s}}\tau_n} p_n\left(\frac{k}{\epsilon_{\mathrm{s}}\tau_n}, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}\right) - \frac{k}{\varphi}$.

When deciding on optimal gross exports k_n, the firm does not know ϵ_{d} nor ϵ_{s} yet, and thus maximizes expected variable profits:

$$k_n(\varphi) = \operatorname*{argmax}_k \mathbf{E}\left[\frac{k}{\epsilon_{\mathrm{s}}\tau_n} p\left(\frac{k}{\epsilon_{\mathrm{s}}\tau_n}, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}\right)\right] - \frac{k}{\varphi}.$$

Solving for optimal k using (9) for the optimal ex-post price:

$$\begin{aligned} k_n(\varphi) &= \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} R_n P_n^{\sigma-1} \varphi^{\sigma} \tau_n^{1-\sigma} \bar{\epsilon}^{\sigma} \\ \text{with} \quad \bar{\epsilon} &= \mathbf{E}\left[\epsilon_{\mathrm{d}}^{\frac{1}{\sigma}} \epsilon_{\mathrm{s}}^{\frac{1}{\sigma}-1}\right]. \end{aligned}$$

Note that in the non-stochastic case, $\epsilon_{\mathrm{s}} = \epsilon_{\mathrm{d}} = 1$, this result is identical to the one obtained in the standard Melitz model. Furthermore, using this result and the definition of sales, the observed sales are:

$$r_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} R_n P_n^{\sigma-1} \varphi^{\sigma-1} \tau_n^{1-\sigma} \bar{\epsilon}^{\sigma} \eta$$

with the following notation:

$$\eta = \frac{(\epsilon_{\mathrm{d}}\epsilon_{\mathrm{s}})^{\frac{1}{\sigma}}}{\epsilon_{\mathrm{s}}\bar{\epsilon}} \quad \text{, and } \mathbf{E}[\eta] = 1.$$

By introducing the quantity $B_n = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} R_n \bar{\epsilon}^{\sigma} P_n^{\sigma-1} \tau_{ni}^{1-\sigma}$ which only depends on parameters and the local aggregate demand structure, we highlight that the distribution of *intended* sales is proportional to $\varphi^{\sigma-1}$ but the distribution of *realized* sales is not, it is perturbed by η:

$$\text{intended sales:} \qquad \mathbf{E}[\, r_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi \,] = B_n \varphi^{\sigma-1} \tag{10}$$

$$\text{realized sales:} \qquad r_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) = B_n \varphi^{\sigma-1}\eta. \tag{11}$$

Finally, expected and realized variable profits for this particular trade destination are:

$$\text{expected profits:} \qquad \mathbf{E}[\, \pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi \,] = \frac{B_n}{\sigma}\varphi^{\sigma-1} \tag{12}$$

$$\text{realized profits:} \qquad \pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) = \frac{B_n}{\sigma}\varphi^{\sigma-1} \cdot \big(\sigma(\eta-1)+1\big). \tag{13}$$

Entry decision. Firms choose to enter a market before discovering the values of ϵ_{s} and ϵ_{d}. Thus, they reason in terms of expected profits instead of actual profits. This leads to the same entry selection criteria as in Melitz (2003): a firm enters market n if and only if

$$\varphi \geq \varphi_n^* \quad \text{with} \quad \mathbf{E}[\, \pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi = \varphi_n^* \,] = f_n. \tag{14}$$

The initial entry decision, i.e., the decision whether to draw a φ or not, is again based on expected profits. A potential firm with productivity φ pays the initial fixed cost to exist if and only if:

$$\mathbf{E}\left[\sum_{n} \max\Big(0, \mathbf{E}[\,\pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi\,] - f_n\Big)\right] \geq f_E. \tag{15}$$

Equation (10) establishes the tight link between productivity and intended sales: it is if and only if intended country-specific exports follow a distributional Pareto law that the same can be said about the distribution of firm productivity (at least, above the cut-off). This is true in both our set-up and in the Melitz model. What is not true in our set-up is that one can use realized sales as a stand-in for intended sales. As (11) shows, if we denote the distribution of expected sales as X and η as Ω, realized sales exactly follow our measurement error structure $Y = \Omega X$ with Ω independent from X. Hence no conclusion can be drawn *a priori* from the fact that realized sales do or do not follow a Pareto law distribution. Furthermore, cut-off's are determined by (14), which depend on unobserved expected profits (12) whereas realized profits are given by (13) and do not follow the same distribution as expected profits. Hence the data will show firms with sales below the productivity cut-off: these are firms that expected to be above the zero-profit cut-off but had an unlucky draw of η.

Lastly, the trade elasticities in this model follow the literature summarized by Head and Mayer (2014). The firm-level ("micro") elasticity of trade to a change in variable trade costs (τ) is obvious from equation (10), it is $1 - \sigma$. The aggregate ("macro") elasticity of trade to a change in trade costs, i.e. the percentage change of aggregate trade between the home country and country n when trade costs rise by 1%, is less obvious since a change in trade costs also implies an increase or decrease in the mass of exporters. In general, there is no reason for this aggregate elasticity to be independent of n. In the particular case where productivity is distributed as a Pareto law, Chaney (2008) shows that this macro-elasticity is independent of n and is equal to $-\alpha$, with α the exponent of the Pareto law density. This result is valid in our setting, as all entry decisions are made before uncertainty is revealed.

B A formal test when Ω and X are independent

Using the model of Appendix A, we assume that Ω and X are independent and are distributed as log-normal and Pareto, respectively. We proceed with an horse-race for Y between the log-normal and the Pareto distribution. To further highlight the misspecification issues, we conduct the uniformly most powerful unbiased test proposed by Malevergne, Pisarenko, and Sornette (2011) instead of a QQ-plot. More specifically, we consider a grid for (α,σ), where α takes values between 1 and 4 and σ between 0.2 and 1.2. For each couple, we run $1,000$ simulations and test the null hypothesis that, beyond some threshold, the upper tail of the size distribution is a Pareto law against the alternative that it is a (truncated from below) log-normal distribution. For each simulation, we run the test and count the number of times the null hypothesis is rejected. Interestingly, almost all couples (α,σ) lead to clear-cut conclusions, i.e. when the null hypothesis of a Pareto distribution is rejected for some values of the Pareto distribution exponent and the standard deviation of the log of the log-normal distribution of Ω, this concerns roughly 95% of the number of simulations. As Table 1 shows, using a formal testing procedure, our first result is robust; in the presence of an independent wedge, one can fail to identify a Pareto law in Y even if X is exactly Pareto distributed. As could be expected, the lower α, i.e., the thicker the tail of X, the higher σ needs to be for the log-normal to be a better fit of the data. But even for a low α such as 1, Y becomes more log-normal than Pareto distributed for no more than $\sigma = 0.8$.

$\alpha \setminus \sigma$	0.2	0.4	0.6	0.8	1
1	Pa (0.99)	Pa (0.92)	Pa (0.86)	Ln (0.36)	Ln (0.14)
1.5	Pa (0.98)	Pa (0.87)	Ln (0.28)	Ln (0.05)	Ln (0.01)
2	Pa (0.92)	Ln (0.37)	Ln (0.055)	Ln (0.01)	Ln (0.00)
2.5	Pa (0.89)	Ln (0.16)	Ln (0.02)	Ln (0.00)	Ln (0.00)
3	Pa (0.86)	Ln (0.03)	Ln (0.00)	Ln (0.00)	Ln (0.00)
4	Ln (0.35)	Ln (0.01)	Ln (0.00)	Ln (0.00)	Ln (0.00)

Table 1: Pareto versus Log-normal distribution in the presence of Ω

Note: `Pa` and `Ln` stand for the Pareto and log-normal distribution, respectively, and denote the evidence of the Wilks' test (Malevergne, Pisarenko, and Sornette, 2011). Values in parentheses denote the acceptance rate for the null hypothesis that data are Pareto-distributed.

C Estimators of a power law exponent

Assume X is a random variable that follows a Pareto distribution with exponent α and possibly truncated at $x_{\max}$:

$$\mathbf{P}[X > x] = \begin{cases} 1 & \text{if} \quad x < x_{\min} \\ \dfrac{x^{-\alpha} - (x_{\max})^{-\alpha}}{(x_{\min})^{-\alpha} - (x_{\max})^{-\alpha}} & \text{if} \quad x_{\min} \leq x \leq x_{\max} \\ 0 & \text{if} \quad x > x_{\max} \end{cases} \tag{16}$$

with $x_{\max}$ infinite except for the last estimation method, and assume we have a set of independent random draws $X_1, \ldots, X_N$ of X. We summarize the main estimation methods of the power law exponent α. In terms of notation, we denote the order statistics as $X_{(1)} \geq \cdots \geq X_{(N)}$.

Log-size log-rank regression. The basic idea underlying a log-rank regression is that $\mathbf{P}\big[X > X_{(i)}\big] \approx \frac{i}{N}$ for any $i \in \{1, N\}$, where the right hand side is simply the empirical cumulative distribution function. Hence, using the definition of a Pareto law and taking logs:

$$\ln \frac{i}{N} \approx \ln C - \alpha \ln X_{(i)}. \tag{17}$$

In other words: if X follows a Pareto distribution, one can simply regress log-size on log-rank to obtain a estimate of α. This estimate is consistent and Gabaix and Ibragimov (2011) show that by using $\ln\left(\text{rank} - \frac{1}{2}\right)$, one minimizes bias.

QQ-regression. A QQ ("quantile-quantile") estimation finds the parameter(s) that minimize the sum of squared errors between the N empirical quantiles of the data and the N theoretical quantiles predicted by the parametrized distribution one wishes to estimate (Kratz and Resnick, 1996). In turns out that in the case of a Pareto law, the relationship between empirical quantiles and the parameter of interest, α, is linear. Indeed, the i'th quantile (out of N) is $X_{(i)}$ in the data and Q_i according to a Pareto law, with Q_i solving $\mathbf{P}[X > Q_i] = \frac{i}{N+1}$. This is straightforward to solve:

$$\ln Q_i = \frac{\ln C}{\alpha} - \frac{1}{\alpha} \ln \frac{i}{N+1} \tag{18}$$

Minimizing the sum of squared errors between $\ln Q_i$ and $\ln X_{(i)}$ is by definition a regression of $\ln X_{(i)}$ on $\ln Q_i$, meaning we regress $\ln X_{(i)}$ on $\ln \frac{i}{N+1}$ and a constant. This is the QQ-regression. Two points worth noticing: first, the QQ-regression is nothing more than the reciprocal regression of the log-rank regression, which should be obvious from the two equations (17) and (18). This explains why the standard errors and biases shown in Figures 6 and 2 are very similar. Second, the relationship between the empirical quantiles and the parameters of a log-normal distribution is also linear, which makes a QQ-regression of a log-normal distribution equally easy.

The Hill estimator (maximum likelihood for an non right-truncated Pareto law). The Hill estimator (Hill, 1975) $\widehat{\alpha}$ is the maximum-likelihood estimator of the power law exponent, which has a closed-form expression:

$$\widehat{\alpha} = \frac{N}{\sum_{i=1}^{N} \left[\ln X_{(i)} - \ln X_{(N)}\right]} \tag{19}$$

Note that Aban and Meerschaert (2004) show that $\widehat{\alpha}^{-1}$, the inverse of the Hill estimator, is the best linear unbiased estimator and the best uniformly minimum variance unbiased estimator of α^{-1}.

Maximum likelihood for a right-truncated Pareto law. In the event one suspects the data at hand to be generated by a right-truncated Pareto law with unknown upper bound (i.e., $x_{\max}$ is possibly finite), Aban, Meerschaert, and Panorska (2006) show that the maximum likelihood estimator of the power law exponent is the solution of the following equation:

$$\frac{N}{\widehat{\alpha}} + \frac{N\left[X_{(N)}/X_{(1)}\right]^{\widehat{\alpha}} \ln\left[X_{(N)}/X_{(1)}\right]}{1 - \left[X_{(N)}/X_{(1)}\right]^{\widehat{\alpha}}} - \sum_{i=1}^{N} \left[\ln X_{(i)} - \ln X_{(N)}\right] = 0 \tag{20}$$

D Extensions of the simulated example of Section 4

See Figure 4.

E Productivity as a truncated Pareto law

As mentioned in the literature review in Section 2, a recent trend in international trade is to use a right-truncated Pareto law distribution to model productivity in order to generate

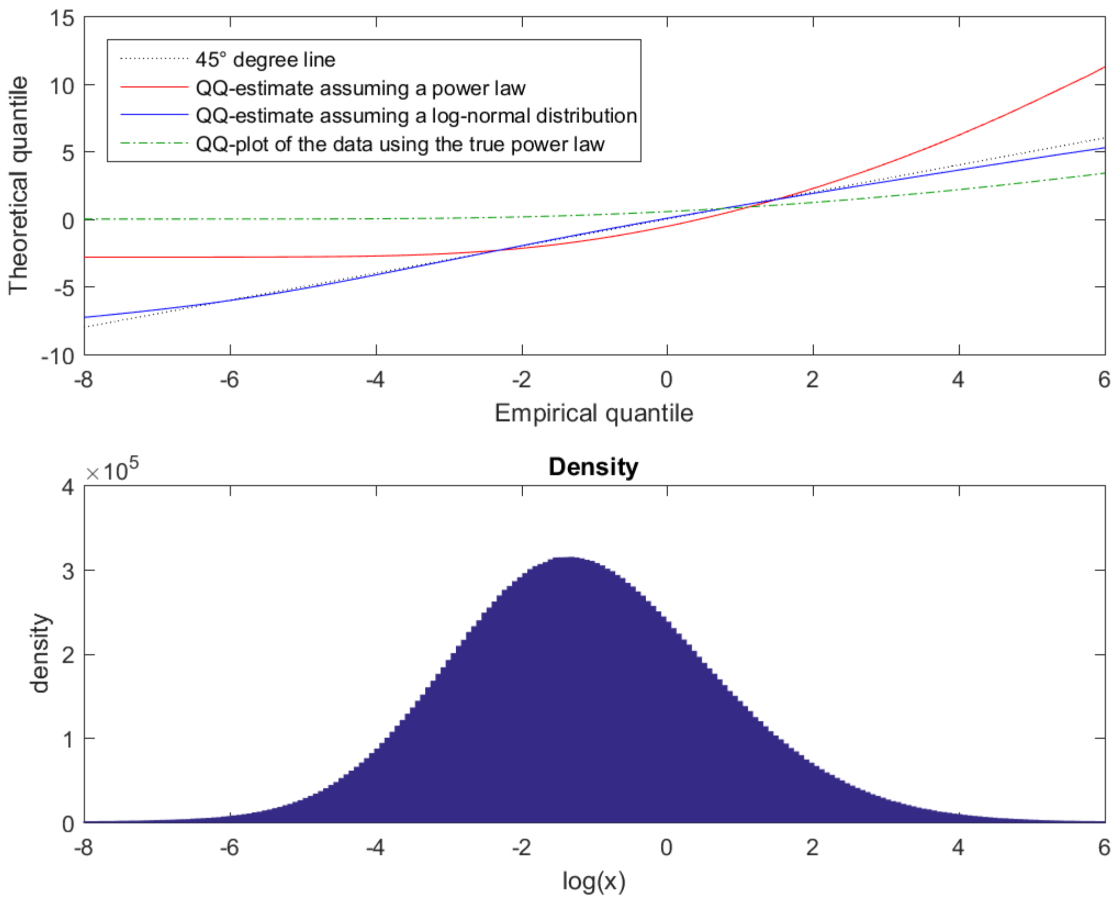

Figure 4: QQ-plots of data $Y = \Omega X$ generated by a Pareto law for X and a correlated Ω wedge as in Eaton, Kortum, and Kramarz (2011), using a very large sample (20,000,000 data points). See Section 4 for the exact data generating process.

effects—such as an expansion of product variety and a pro-competitive reduction in markups—that are absent in the original Melitz model with an unbounded Pareto law. Given none of our work in Section 3 relies on any specific assumption for X, we explore in this appendix whether anything can be inferred about X in the presence of Ω if one assumes X follows a bounded Pareto law and X and Ω are uncorrelated.

Our answer is that the assumption of a bounded Pareto law for productivity is even more difficult to test using exports data, and identification might often not be feasible. Formally, without any assumptions on Ω nothing can be said about X. But contrary to Section 5, Theorem 1 does not apply,[23] so even with Ω not too heavy-tailed (say, log-normal) there is no formal result that allows us to consistently estimate X using only the right tail of Y.[24]

The following illustration may be revealing and useful. We look at a simple example that mimics the simulations from Appendix B, where we generate the data for X using a bounded Pareto law. We use only a small variance for Ω (1.50) and we bound X at the very top, we only drop the top 0.1% of the distribution. The results are in Figures 5 and 6. In the first figure, we run a horse-race between a log-normal and a Pareto distribution on the data Y. Of course, as in the previous section, this horse-race is misspecified. What is interesting is that the log-normal clearly outperforms the Pareto law again. In other words, evidence (even strong evidence) in favor of a log-normal distribution in exports should not be construed as evidence against the assumption of a bounded Pareto law in productivity. With a wedge Ω in the data, it is entirely possible for the log-normal to be a much better fit.

In the second figure, we estimate the exponent of the bounded Pareto law driving X using Y. Given the misspecification, it is unsurprising that all results are inconsistent, including

[23] See our companion paper. The proof fails if X is bounded. Using the notation from that proof, g will never be a slowly-variating function.

[24] In the statistics literature, Beirlant, Fraga Alves, Gomes, and Meerschaert (2014) do offer some guidance for the estimation of bounded-Pareto-type distributions, but their statistical framework, although similar, is not identical to ours and their results do not carry over in a simple manner.

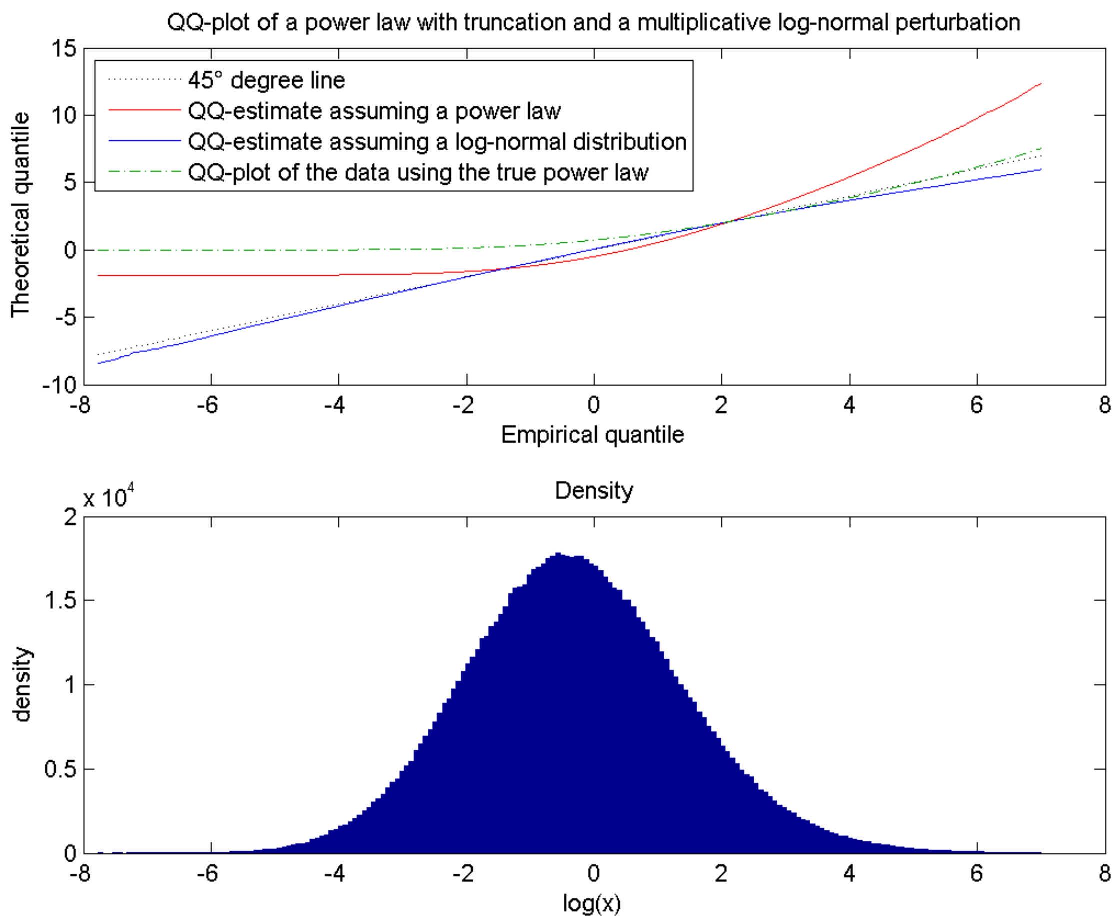

Figure 5: QQ-plots of a truncated Pareto law with a multiplicative log-normal error (large sample)

Note: Data Y is generated according to the following generating process: $Y = X \cdot \Omega$, where X is a Pareto law with $\alpha = 1.2$ and truncated at the top 0.1%, $x_{\min} = 1.0$ and Ω the exponential of a normal distribution of mean 0 and standard deviation 1.50. The size of the dataset is large, $1,000,000$ draws.

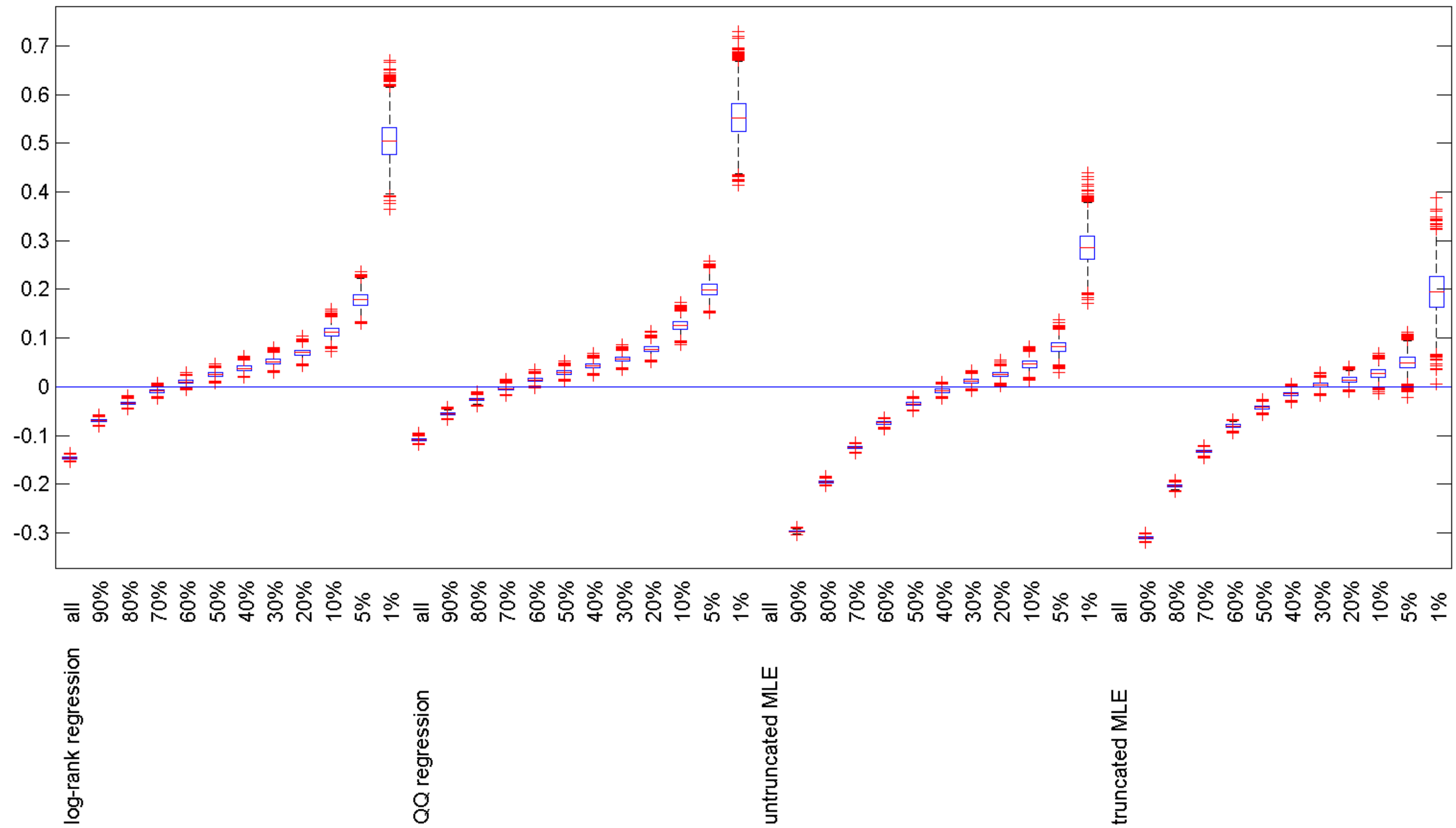

Figure 6: Monte-Carlo simulations of estimates of the power law exponent (truncated case).

Notes: Boxplot of 5,000 estimates of the power law exponent α according to four estimation procedures (see main text), at different levels of left-truncation. Each estimation is done on data Y that is generated according to the following generating process: $Y = X \cdot \Omega$, where X is a Pareto law with $\alpha = 1.2$, $x_{\min} = 1.0$ truncated at the top 0.1%, and Ω the exponential of a normal distribution of mean 0 and standard deviation 0.60. The size of each draw is $100,000$.

Results are centered and normalized around the true value of α, i.e. we show $\dfrac{\widehat{\alpha} - \alpha}{\alpha}$.

Results that are not shown lie outside the graph.

if one only uses the top of the distribution, and including the maximum likelihood method of Aban, Meerschaert, and Panorska (2006) that supposedly accounts for right-truncation. There is currently no known estimation method of a bounded Pareto law in the presence of a wedge Ω.[25] In practice, a warning sign that X may be truncated-Pareto should be the fact that one runs the estimation on data that is successively more left-truncated (as in the previous subsection) and observes that the estimated value α does not stabilize (contrary to what one sees in Figure 2 in the case of a non-truncated distribution for X). If the estimate of α keeps increasing, one is dealing with data Y that is thinner tailed than a Pareto law. According to Corollary 1, this means that X is also thin-tailed, one possibility being a truncated-Pareto,[26] although we know of no identification strategy to decide between different thin-tailed candidate distributions for X. In addition, as indicated in Section 5.3, a thin tail in Y could also be the consequence of a too-small data set. In conclusion, we do not see in practice how the identification of a truncated Pareto can be done conclusively if the data is perturbated by a non-degenerate Ω.

[25] Preliminary simulations show that, if Ω is "sufficiently" thin-tailed and X not "too" bounded, it may be possible to find a window in the data that allows for the estimation of α. This is left for future research.

[26] Note that a truncated Pareto law is by definition thin-tailed, since the density is equal to 0 for high enough values.