Can Productivity Follow a Pareto distribution if Exports "Look" Log-normal?

Marnix Amand[*]
HEC Lausanne

Florian Pelgrin[†]
EDHEC Business School

February 16, 2016

Abstract

Yes it can. We show that many forms of heterogeneity and uncertainty that are common in the heterogeneous firm international trade literature drive a wedge between observed exports in the data and productivity. This wedge can distort the left end of the observed exports to "look" log-normal even if productivity is exactly Pareto distributed. Hence identifying exports with productivity one-for-one, as is often done, relies on knife-edge assumptions. Furthermore, the presence of this wedge, which is possibly correlated with the productivity distribution, means that econometric methods that rely on exports to directly identify and estimate the productivity distribution are misspecified. However, under reasonable assumptions, we prove that a power law tail in productivity will imply a power law tail in observed exports. This allows for consistent estimates of the productivity distribution and for a test for a power law tail in productivity without needing to rely on strong modelling assumptions. Lastly, we show that testing for an exact Pareto distribution for productivity is impossible, barring unreasonable assumptions.

Keywords: Pareto, power law, international trade, measurement error, heterogeneous firms.
JEL Classification Numbers: F12, F14,

1 Introduction

Is the use of a Pareto distribution[1] to model productivity heterogeneity at all justified empirically? This modeling assumption has become increasingly popular in the international trade literature,

[*]HEC Lausanne - University of Lausanne, Quartier Dorigny - Internef 524, CH-1015 Lausanne, Switzerland. Email: marnix.amand@unil.ch.

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

[1]A random variable X follows a Pareto distribution, sometimes called Pareto law, if its couter-cumulative distribution decays at polynomial speed, 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.

in particular in research that builds on the seminal Melitz (2003) heterogeneous firm international trade model.[2] The main reason for this popularity is tractability. A Pareto distribution is "scale-free", in the sense that it is invariant to left-truncation, and this turns out to be an extremely useful property, both empirically and theoretically, for models that have as one of their drivers the self-selection of more productive firms into export markets.

The problem is, recent empirical results are challenging this modeling assumption, however convenient and popular it may be for theorists. The standard Melitz model predicts that destination-specific exports are distributed as an exponent of productivity. Given that a Pareto distribution is invariant to exponentiation (up to a simple change in parameters), it is simple enough within this framework to test the Pareto assumption for productivity using exports data. Unfortunately, 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 a Pareto law. Head, Mayer, and Thoenig (2014) argue that this discrepancy is welfare-relevant and thus cannot be ignored. In other words, they argue that the empirical evidence rejects the Pareto assumption along an economically significant dimension, notwithstanding the Pareto's oh-so-useful properties for theorists. Our reading of the dominant interpretation of the recent empirical work is that it is becoming harder and harder to answer the introductory question in an affirmative way.

In this paper, we wish to offer an alternative interpretation of the empirical evidence on exports that allows to reconcile the Pareto law assumption for productivity with an exports distribution that is shaped differently.

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. Without further structure on Ω, for any strictly positive random variable X, an Ω can be found to accommodate the data Y. Hence Y is of no use to learn anything about X unless one knows (or makes strong assumptions about) Ω. In the canonical Melitz (2003) model, one has (up to an

[2] Note that the analytic convenience of the Pareto distribution has made it very popular outside of international trade too. To give a brief overview, Pareto distributions have been used to model the productivity distribution of firms in an endogenous growth context (Perla and Tonetti, 2014; Lucas and Moll, 2014); the firm size distribution in a business cycles fluctuations context (Gabaix, 2011); the market value distribution of firms (Gabaix and Landier, 2008); the distribution of ideas (Jones, 2005) and the city-site distribution (Gabaix, 1999; Behrens, Duranton, and Robert-Nicoud, 2014). For further papers that use this assumption, see the reviews in Gabaix (2009) and Gabaix (2016).

irrelevant exponentiation and proportionality factor) $Y = X$, i.e. $\Omega = 1$. Of course, in that case, identification is not an issue, if Y is not a Pareto law, neither can X be. But from an empiricist's view, it is a very strong claim to say that the (usually unobserved) productivity distribution is exactly identified and matched with available export data. We will show that all it takes to break this one-to-one relationship between exports data Y and productivity X is to add an element of heterogeneity to the Melitz model. Doing this drives a "wedge" between observed exports and the underlying productivity distribution.

The idea that the link between productivity and exports must be distended is not novel as such. Arkolakis (2010) has previously solved[3] for an explicit example of a non-degenerate Ω. Our approach is more general and less geared towards explaining a given set of empirical facts. 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 such wedges (or any combination thereof) boil down to a non-degenerate Ω.

This is the first take-away point of our paper. The identification of productivity 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. Only if one is willing to assume away *all* forms of heterogeneity (other than productivity), uncertainty or plain measurement error can one use exports directly to characterize 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 are no longer 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. Under specific assumptions, we furthermore show that these wedges behave like a multiplicative measurement error independent of firm size. This orthogonality allows one to conserve the main properties regarding welfare and aggregate trade elasticities of the Melitz model with Pareto distributed productivity (see Chaney, 2008) whilst offering a flexible explanation of the non-Pareto distribution of observed exports data.[4]

Furthermore, in the presence of Ω, X is no longer identified by Y and standard econometric techniques that ignore Ω are obviously misspecified. In principle, one then cannot use Y anymore to infer anything about X. This has practical consequences: (misspecified) methods that look for a *complete* Pareto distribution in exports data or that use exports data to estimate a Pareto law yield misleading results. We even show that, in a horse race using data simulated with a Pareto distribution for productivity X but with a non-degenerate wedge Ω, it is very likely for the

[3]Equation 23 in his paper.

[4]As a matter of fact, under mild regularity conditions, *any* exports distribution can be explained in our model using a Pareto productivity distribution and a suitably chosen "wedge", this is a simple deconvolution problem.

log-normal distribution to be a better fit for Y than a Pareto distribution.

Finally, our last research question is whether, despite Ω, exports data contains some information about productivity. At an intuitive level, it should: 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. The previous literature either used the exports distribution 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 both cases, this comes down to making an explicit assumption on the link between productivity and exports (i.e., an explicit assumption on Ω). In practice, this might be problematic as one knows very little about unmodelled heterogeneity, measurement error etc. We show formally that precise assumptions on the structure of Ω are not needed for estimation purposes because the heavy tail of the productivity distribution will come to dominate the shape of the right of the exports distribution. For this to hold one needs to be willing to assume that the wedge between productivity and exports is not "too" heavy tailed and not "too" correlated with X for high values of X (both reasonable assumptions, we argue).

To the best of our knowledge, this is new in economics. Our interpretation is that this is an "(almost) anything goes" result. A Pareto distribution X can be perturbated by any type of Ω, as long as Ω is well behaved at the top (not "too" heavy tailed and not "too" correlated with X) estimation of the power law exponent of X is consistent. Our approach is in principle applicable to any situation where one has data Y but needs inferences about a Pareto distribution X with $Y = \Omega X$. This result also allows us to reconcile previous empirical evidence that shows exports only had a power law tail with the Pareto law assumption. 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.

The paper is organized as follows. Section 2 highlights the theoretical importance of the Pareto assumption for the productivity distribution and the subsequent empirical work. Section 3 introduces two versions of our Melitz-type model with stochastic wedges, one where wedges reduce to measurement error (with beneficial consequences), and one, more general, where this is not the case. Both models result in a data structure of the $Y = \Omega X$ type. Section 4 illustrates with an example that the presence of wedges makes identification and estimation problematic, and shows formally that under some circumstances the right tail of exports can be used. Section 5 concludes.

2 Pareto distributions in international trade

The main explaining factor of the patterns of international trade in heterogeneous-firm models that are built on the seminal paper by Melitz (2003) is the assumption that, for a given export destination, all firms face an inverted U-shaped total profit curve in function of the quantity of exports. This is generated by a fixed entry cost and a downward-sloping demand curve. The more productive a firm is, the higher and further to the right the highest (i.e., optimal) point on that curve. Only those firms with positive profits actually export, which leads to a minimal (”cut-off”) productivity that depends on the destination. This assumption was initially motivated by the stylized fact that more productive firms export more and more often (Bernard and Jensen, 1999; Melitz and Redding, 2014), which is indeed a theoretical prediction of this model. However, these fixed cut-off's generically make the computation of trade elasticities and welfare effects difficult when studying trade policy, as the aggregate effects of moving a cut-off depend on where exactly this cut-off is located on the productivity distribution.

To tackle this problem, the use of a Pareto law to model the distribution of firm productivity has become increasingly popular. A Pareto distribution is "scale-free", in the sense that its shape is invariant to left-truncation. Consequently, with this assumption the shape of the productivity distribution of firms that export to a given destination does not depend on where the cut-off is. Thus, the theoretical variables of interest generally become, up to a multiplicative factor, independent of country-pair characteristics. Said differently, this means that if one assumes a Pareto distribution for firm productivity, trade between France and Germany on one hand and France and the Galapagos Islands 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). 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.[5]

As stated in the introduction, considerable attention has been given to the empirical justification of this 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). Since a

[5] For further references, see the long footnote 22 in Arkolakis, Costinot, and Rodrìguez-Clare (2012) for a list of the main papers using this assumption.

[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.

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 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.

Several papers have done exactly that and have shown that the left 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 are more likely to be log-normally distributed than Pareto-law distributed. 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). Simply put, 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 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 looking only at the top of the 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

[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 not novel and is commonly done in the literature through a change of variables (in particular, see Head, Mayer, and Thoenig, 2014).

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. This approach is very elegant from a modelling perspective and has the very nice property that 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 new 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 "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 D 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: two models

In this section, we build on the Melitz model and provide two variants of this model in which observed exports Y and underlying productivity X are linked as in (1). In the first model, Ω is independent of X which, under the Pareto law assumption for X, implies a constant partial-equilibrium aggregate trade elasticity and simple welfare calculations, both desirable properties as described in the literature review. In our less restrictive case, X and Ω are no longer independent, and the previous elasticity and welfare results do not hold anymore. Both variants have some critical implications regarding the identification of the Pareto law distribution.

3.1 A special case: Ω behaving as measurement error

As argued in the introduction, one possible interpretation of Ω is measurement error, there is no reason to assume that the reporting by firms is done flawlessly. If one assumes that the data is

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

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[10] shows that the observed data Y is $Y = X\Omega$ with Ω distributed log-normally and independent from X.

As such, this is an interesting result: 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. But relying solely on measurement error to generate these results—and not heterogeneities or shocks embedded in the model—would be very restrictive. We argue that such a "measurement error interpretation" of Ω can also be justified from a theoretical perspective.[11] 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 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 follow the (very standard) notation of Melitz and Redding (2014), dropping the index (customarily 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

[10] 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 $Y \approx Xe^{\epsilon}$.

[11] 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.

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 fixed cost f_n. By assumption, firms must serve the domestic market before exporting.[12]

This 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. Finally, ϵ_{s} and ϵ_{d}'s joint cumulative distribution function is H, defined on $\mathbb{R}^+ \times \mathbb{R}^+$.

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:

$$q = \epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} p_n(q, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}})^{-\sigma} \tag{2}$$

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

[12] 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.

expected variable profits:

$$k_n(\varphi) = \underset{k}{\text{argmax}} \iint\limits_{\mathbb{R}^+ \times \mathbb{R}^+} \frac{k}{\epsilon_\text{s}\tau_n} p\left(\frac{k}{\epsilon_\text{s}\tau_n}, \epsilon_\text{d}, \epsilon_\text{s}\right) dH(\epsilon_\text{d}, \epsilon_\text{s}) - \frac{k}{\varphi}.$$

Solving for optimal k using (2) for the optimal ex-post price:

$$k_n(\varphi) = \left(\frac{\sigma - 1}{\sigma}\right)^\sigma R_n P_n^{\sigma-1} \varphi^\sigma \tau_n^{1-\sigma} \overline{\epsilon}^\sigma$$
$$\text{with} \quad \overline{\epsilon} = \iint\limits_{\mathbb{R}^+ \times \mathbb{R}^+} \epsilon_\text{d}^{\frac{1}{\sigma}} \epsilon_\text{s}^{\frac{1}{\sigma}-1} dH(\epsilon_\text{s}, \epsilon_\text{d}).$$

Note that in the non-stochastic case, $\epsilon_\text{s} = \epsilon_\text{d} = 1$, 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:

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

with the following notation:

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

By introducing the quantity $B_n = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} R_n \overline{\epsilon}^\sigma P_n^{\sigma-1} \tau_{ni}^{1-\sigma}$ 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:

$$\text{intended sales:} \qquad \mathbf{E}[\, r_n(\varphi, \epsilon_\text{d}, \epsilon_\text{s}) \,|\, \varphi \,] = B_n \varphi^{\sigma-1} \tag{3}$$

$$\text{realized sales:} \qquad r_n(\varphi, \epsilon_\text{d}, \epsilon_\text{s}) = B_n \varphi^{\sigma-1} \eta. \tag{4}$$

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

$$\text{expected profits:} \qquad \mathbf{E}[\, \pi_n(\varphi, \epsilon_\text{d}, \epsilon_\text{s}) \,|\, \varphi \,] = \frac{B_n}{\sigma} \varphi^{\sigma-1} \tag{5}$$

$$\text{realized profits:} \qquad \pi_n(\varphi, \epsilon_\text{d}, \epsilon_\text{s}) = \frac{B_n}{\sigma} \varphi^{\sigma-1} \cdot \big(\sigma(\eta - 1) + 1\big). \tag{6}$$

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

$$\varphi \geq \varphi_n^* \quad \text{with} \quad \mathbf{E}[\,\pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi = \varphi_n^*\,] = f_n. \tag{7}$$

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:

$$\mathbf{E}\left[\sum_n \max\Big(0, \mathbf{E}[\,\pi_n(\varphi, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{s}}) \,|\, \varphi\,] - f_n\Big)\right] \geq f_E. \tag{8}$$

Equation (3) 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 (4) 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 (7), which depend on unobserved expected profits (5) whereas realized profits are given by (6) 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 (3), 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.

3.2 General case: Ω as heterogeneity in more than just productivity

Our second variant of the Melitz model stays close to the recent literature and builds on Arkolakis (2010) and Eaton, Kortum, and Kramarz (2011), but without any specific assumption on the distribution of productivity draws.

Setup. Our starting point is the same canonical Melitz model as in the previous section without uncertainty and with endogenous variable market access costs following Arkolakis (2010). 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}.[13] These heterogeneities are denoted collectively as ϵ. Note that all of these heterogeneities are known in advance, and vary per destination n for a given firm. Lastly, market access costs are defined as follows: a firm can access a fraction $m \in [0, 1]$ of a market n by paying the following cost:

$$f_n(m, \epsilon) = \epsilon_{\mathrm{f}} f_n + \epsilon_{\mathrm{m}} \frac{1 - (1 - m)^{1-1/\lambda}}{1 - 1/\lambda}. \tag{9}$$

Optimal strategy for an exporting firm. We follow the previous literature and the previous section. Given the CES demand structure, the optimal price $p_n(q, m, \epsilon)$ charged by a firm selling q to fraction m of market n is given by the demand curve:

$$q = m\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} p_n(q, m, \epsilon)^{-\sigma} \tag{10}$$

with P_n the standard CES price index. Hence the profits (or losses) of such a firm are $p_n(q, m, \epsilon)q - f_n(m, \epsilon) - (q\tau/\varphi)$. Maximal profits are obtained using first order conditions and yield the following optimal choices of q and m:

$$q_n(\varphi, \epsilon) = m_n(\varphi, \epsilon)\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} \left(\frac{\varphi}{\tau}\right)^{\sigma} \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma}$$

$$m_n(\varphi, \epsilon) = 1 - \left[\frac{\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1}}{\epsilon_{\mathrm{m}}} \left(\frac{\varphi}{\tau}\right)^{\sigma-1} \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma-1}\right]^{-\lambda}$$

We assume for now that $m_n(\varphi, \epsilon) \geq 0$. This will later be one of the entry conditions. Note that $m_n(\varphi, \epsilon) \leq 1$ is always verified.

To streamline notation, we introduce the variable $\overline{\pi}_n$ defined such that $\overline{\pi}_n \varphi^{\sigma-1}$ is the optimal profits before entry costs (i.e., sales minus production and iceberg costs) while keeping all heterogeneities

[13] 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 is unnecessary. 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 and two first-order conditions. So two firms can only differ along a maximum of three dimensions (in addition to productivity). To keep things simple, we have chosen these three heterogeneities such that the algebraic solutions are simple and (mostly) linear.

at their expected value of 1. We also introduce variables ξ_1, ξ_2 and ξ_3, collectively referred to as ξ, which are simple combinations of the heterogeneities ϵ:

$$\overline{\pi}_n = \frac{1}{\sigma} R_n P_n^{\sigma-1} \tau^{1-\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}$$
$$\text{with} \qquad \xi_1 = \epsilon_{\mathrm{d}}, \qquad \xi_2 = \left[\frac{\epsilon_{\mathrm{d}}}{\epsilon_{\mathrm{m}}}\right]^{-\lambda}, \qquad \xi_3 = \left(\epsilon_{\mathrm{f}} f_n + \frac{\epsilon_{\mathrm{m}}}{1-1/\lambda}\right)$$

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 following its optimal policy:

$$\begin{aligned}
&\text{Target market size:} && m_n(\varphi,\xi) = 1 - \xi_2 \overline{\pi}_n^{-\lambda} \varphi^{\lambda(1-\sigma)} \\
&\text{Quantity exported:} && q_n(\varphi,\xi) = \xi_1 m_n(\varphi,\xi) \frac{\varphi(\sigma-1)}{\tau} \overline{\pi}_n \varphi^{\sigma-1} \\
&\text{Sales:} && r_n(\varphi,\xi) = \xi_1 \sigma m_n(\varphi,\xi) \overline{\pi}_n \varphi^{\sigma-1} && (11) \\
&\text{Profits:} && \pi_n(\varphi,\xi) = \xi_1 \overline{\pi}_n \varphi^{\sigma-1} \frac{\lambda - m_n(\varphi,\xi)}{\lambda-1} - \xi_3 && (12)
\end{aligned}$$

Entry decision. Firms enter a market n only if their profits $\pi_n(\varphi, \xi)$ are positive. There are two conditions for this to be true. A firm must at least cover all its costs, given by equation (12), and, as indicated previously, a firm must target a non-negative market size:

$$\pi_n(\varphi,\xi) \geq 0$$
$$m_n(\varphi,\xi) \geq 0.$$

Given both m_n and π_n are increasing in φ for all ξ, the entry condition is

$$\varphi \geq \varphi_n(\xi)$$

with the minimal productivity threshold $\varphi_n(\xi)$ defined as:

$$\begin{aligned}
& && \varphi_n(\xi) = \max\left\{\varphi_{1,n}(\xi), \varphi_{2,n}(\xi)\right\} \\
&\text{with} && \pi_n\big(\varphi_{2,n}(\xi), \xi\big) = 0 \\
&\text{and} && \left[\overline{\pi}_n \cdot (\varphi_{1,n}(\xi))^{\sigma-1}\right]^{\lambda} = \xi_2
\end{aligned}$$

Note that in the case of $\lambda \geq 1$, we always have $\varphi_{2,n}(\xi) \geq \varphi_{1,n}(\xi)$. Lastly, the initial entry decision is

$$\mathbf{E}\left[\sum_{\substack{n \\ \varphi \geq \varphi_n(\xi)}} \max\Big(0, \pi_n(\varphi, \xi)\Big)\right] \geq f_E \tag{13}$$

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

$$\delta_n(\varphi, \xi) = \begin{cases} 1 & \text{if} \quad \varphi \geq \varphi_n(\xi) \\ 0 & \text{if not} \end{cases} \tag{14}$$

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

$$r_n(\varphi, \xi) = \delta_n(\varphi, \xi)\xi_1 m_n(\varphi, \xi)\sigma\overline{\pi}_n\varphi^{\sigma-1}. \tag{15}$$

Over the space of realizations (φ, ξ), 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 $\delta_n(\varphi, \xi)\xi_1 m_n(\varphi, \xi)\sigma\overline{\pi}_n$. Then, in the notation of our framework, sales are distributed as $Y = \Omega X$. Note that, contrary to the previous section, Ω is not independent of X and does not behave as a measurement error, it is composed of three separate stochastic processes: a "selection" process $\delta_n(\varphi, \xi)$, a "market share" process $m_n(\varphi, \xi)$ and a "market size" process ξ_1. 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.

4 Practical and econometric implications of Ω

We now turn to the econometric implications of the general formulation $Y = X\Omega$. Once one has assumed that the data Y is distributed as in (1), with Ω determined by one of our models, we simply ask whether (global or local) identification as well as inference about X are feasible and, if so, under what conditions. Following Section 3, we subsequently assume that X and Ω are either independent (section 3.1) or not (section 3.2). As to be expected, the standard misspecification problem—using Y *in lieu* of X—leads to several issues. Identification of the distribution of X is problematic and the power law exponent can generally only be identified and consistently estimated by looking at the right tail of Y under certain conditions. Conversely, this yields a simple test: the absence of a power law tail in Y does imply the absence of a Pareto law tail in X under those same conditions.

4.1 Pareto/log-normal debate in the presence of misspecification

A qq-plot exploration To highlight the misspecification issue, we first 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—the true data generating process of X being a Pareto distribution. To do this, we make use of a QQ-regression as suggested in the trade literature (Head, Mayer, and Thoenig, 2014). Visually, this comes down to comparing the QQ-plot of best-fitting member of each family of distributions with the 45 degree line.[14] 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.[15]

We simulate 10,000 draws of Y—a sample size often encountered in international trade data—with a Pareto law distribution for X with exponent 1.2 and an independent log-normal distribution for Ω with standard deviation of the log of Ω of 1.50. Then we estimate the log-normal and Pareto law distribution that best fits 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 than to the best-fitting distributional Pareto law.

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 1.79). This means that most of the variation on the left will be driven by the log-normal distribution, even if it has a small standard deviation itself. Hence, even for a small-variance Ω, the bottom of the distribution is essentially 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 fist Y as a consistent estimate of the Pareto law underlying X.[16] In practice, unless one is willing to assume

[14] 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 B in the Appendix for a brief summary.

[15] Head, Mayer, and Thoenig (2014) apply this procedure on firm-level data on exports from France to Belgium and from China to Japan and conclude that the log-normal distribution is a far better fit than the Pareto law.

[16] Knowing the distribution of Ω, the estimation of the Pareto law exponent would lead to a deconvolution problem written in multiplicative form. Maximum-likelihood methods would then yield consistent estimates. However, little is known in practise about Ω, which rules out standard deconvolution methods.

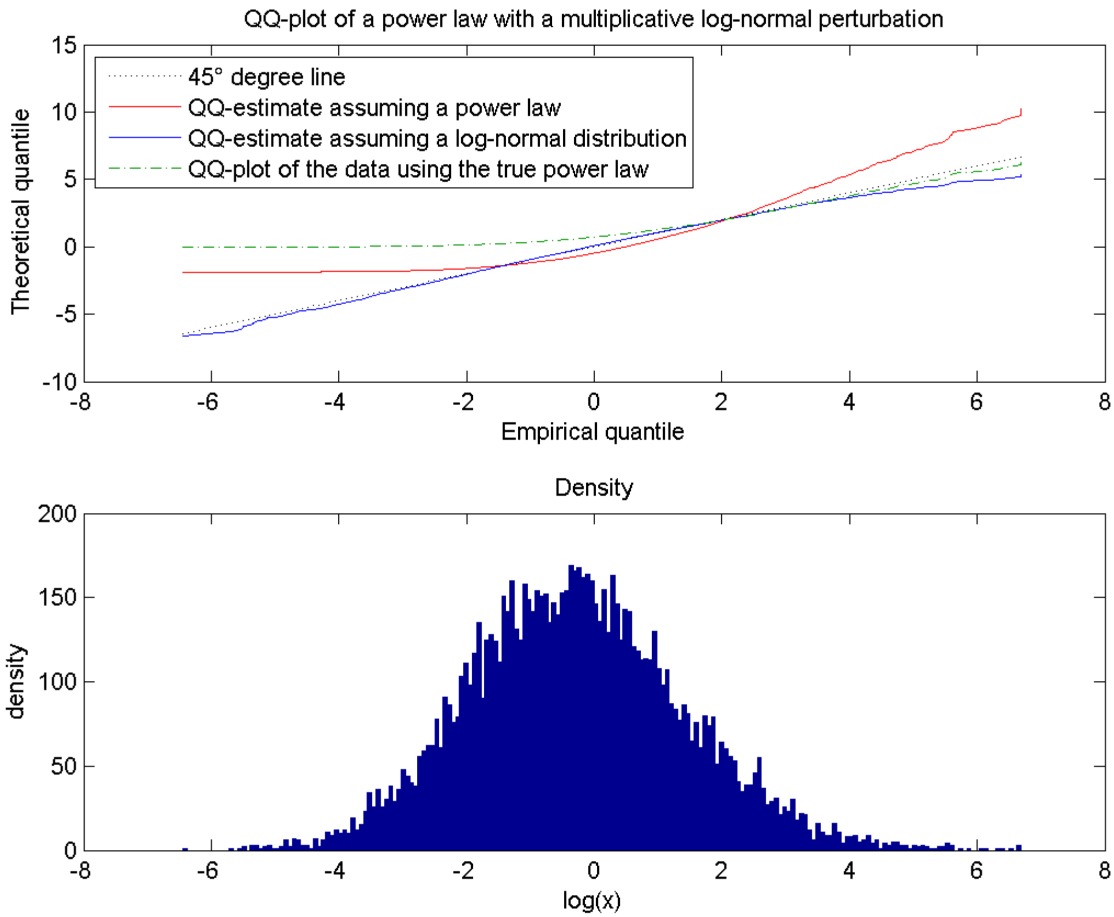

Figure 1: QQ-plots of a Pareto law with a multiplicative log-normal error (small sample)

Note: Data Y is generated according to the following generating process: $Y = X \cdot \Omega$, where X is a Pareto law with $\alpha = 1.2$, $x_{\min} = 1.0$ and Ω the exponential of a normal distribution of mean 0 and standard deviation 1.50.

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.

One should note that our example is not driven by small sample bias. In Figure 3 in Appendix C, we show that our results are identical if one uses a very large sample of $1,000,000$ points. Nor are these results driven by the higher variance of the log-normal distribution of Ω "overwhelming" the Pareto law distribution of X. The Pareto law distribution in our example has infinite variance. Lastly, these results are not particular to the QQ-plot approach. To assess the reliability of the visual inspection, we proceed with the uniformly most powerful unbiased test proposed by Malevergne, Pisarenko and Sornette (2011).[17] More specifically, we consider a grid for (α,σ), where α takes its 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.[18] 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 power-law distribution exponent and the standard deviation of the log of 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 to be expected, the lower α, i.e., the thiner the tail of X, the higher σ needs to be for the log-normal to get a better fit of the data. But even for a low α such as 1 (Zipf's law), Y becomes more log-normal than Pareto distributed for $\sigma = 0.8$ (and actually, even a little less).

Estimating a power law exponent using Y 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 Monte-Carlo simulations on $5,000$ independently drawn datasets Y following the same generating process as in the previous section ($\alpha = 1.2$ and $\sigma = 0.60$, following the same notation), and we estimate a power law exponent using two versions of maximum likelihood, namely the unconditional Hill's estimator (Hill, 1975) and the conditional Hill's estimator (Aban, Meerschaert, and Panorska, 2006), a log-rank regression (Gabaix and Ibragimov, 2011) and a QQ-regression (Kratz and Resnick, 1996; Schultze and

[17] This test is known as the Wilks' test and can be viewed as a likelihood ratio test (see Malevergne, Pisarenko and Sornette, 2011).

[18] The Pareto distribution can be viewed as a "limit case" of the log-normal distribution as in a "nested" test. The test is 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 our experiments) or by Monte Carlo simulations.

$\alpha \setminus \sigma$	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 et al., 2011). Values in parentheses denote the acceptance rate for the null hypothesis that data are Pareto-distributed.

Steinebach, 1996) without adjusting for the presence of Ω.[19] Since the statistics literature recommends to left-truncate the data for better results when there is a perturbation for the lower values (a so-called slowly varying function), we run each estimation procedure again on left-truncated data by progressively dropping more and more (from 10% to 99%) 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, interestingly, our simulations also illustrate that left-truncating the data (i.e., dropping the lowest points) does allow for consistent estimates. Note that this is done in practice by many papers without further justification. In the next section, we show that this is actually a general result. It is straightforward to see where this comes from. Let X be distributed according to a Pareto distribution with exponent α:

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

Let Ω be independent of X, of mean 1. We denote by f_X the probability density function of X, by f_Ω the probability density function of Ω. Then, following (1) and given the independence of X

[19]We recapitulate the main estimation methods in Appendix B.

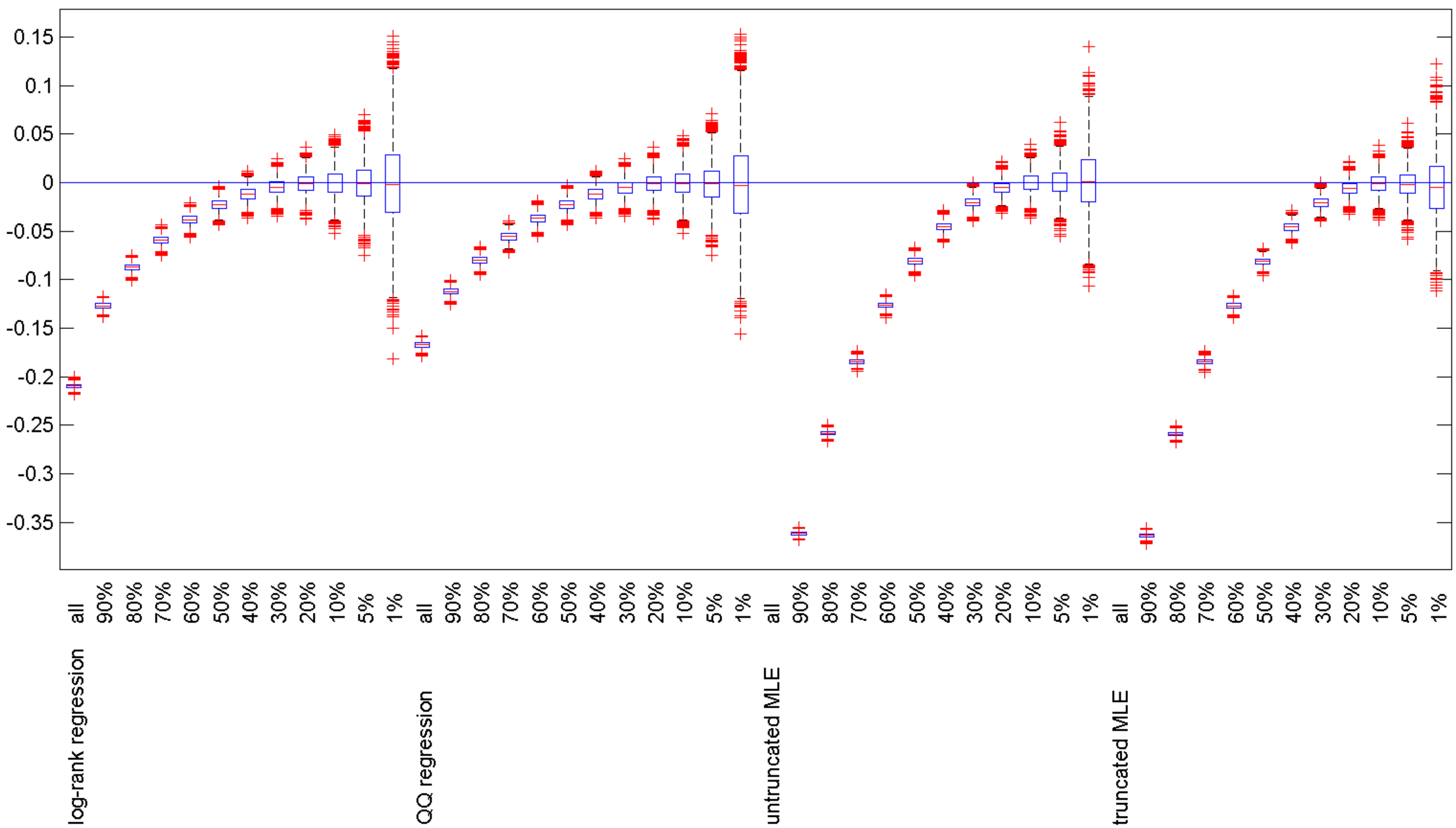

Figure 2: Monte-Carlo simulations of estimates of the power law exponent (untruncated case).

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 $\alpha = 1.2$, $x_{\min} = 1.0$ 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: $\dfrac{\widehat{\alpha} - \alpha}{\alpha}$ is reported.

Results that are not shown lie outside the graph.

and Ω, we have for all $y \geq 0$:

$$\begin{aligned}
\mathbf{P}[Y > y] &= \mathbf{P}[X\Omega > y] \\
&= \int_0^{\frac{y}{x_{\min}}} \left(\int_{\frac{y}{\omega}}^{\infty} f_X(x)dx \right) f_\Omega(\omega)d\omega + \int_{\frac{y}{x_{\min}}}^{\infty} 1 \cdot f_\Omega(\omega)d\omega \\
&= \left(\frac{y}{x_{\min}}\right)^{-\alpha} \underbrace{\left[\int_0^{\frac{y}{x_{\min}}} \omega^\alpha f_\Omega(\omega)d\omega + \left(\frac{y}{x_{\min}}\right)^{\alpha} \mathbf{P}[\Omega x_{\min} > y] \right]}_{g(y)}
\end{aligned} \tag{17}$$

In the log-normal case, it is straightforward to show that we have, using the Landau notation for $y \to \infty$, for a given constant C:

$$\mathbf{P}[Y > y] = Cy^{-\alpha} + o(y^{-\alpha}) \tag{18}$$

This heuristically explains why left-truncating the data Y works when estimating the power law exponent α of X: the upper part of the data behaves like the correct Pareto distribution of X, up to a negligible term.

4.2 Using the right tail of the data to estimate a Pareto law

We have shown that the complete productivity distribution X cannot be generally inferred in the presence of a wedge Ω. In other words, the assumption that X is an exact Pareto distribution is generically untestable (unless one is willing to make very strong assumptions on Ω). This leaves two open questions: (1) if one is willing to assume that X is exactly Pareto distributed, can the tail of Y be used to estimate X? (2) If Y has a power law tail (i.e., is regularly varying), does this imply the same for X?

4.2.1 Estimation

As illustrated in Section 4.1, using the right tail of the distribution is perfectly legitimate, irrespective of the chosen estimation method in Appendix B. The literature in statistics has shown the following result:

Result 1. *Let Y be a random variable with a distribution function such that, for all y:*

$$\mathbf{P}[Y > y] = y^{-\alpha} g(y),$$

with g a slowly-varying function[20] *and $\alpha > 0$. 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 estimators of α described in Appendix B, computed using the k_n highest order statistics of the n first observations.

Proof. See Appendix A □

The interpretation is that each estimator can get arbitrarily close (in probability) to the true value of α by just using the right tail (the highest order statistics) provided one has a sufficiently large sample.[21] Simply put, all the usual estimators of a power law exponent remain consistent if used on a distribution that is not a Pareto distribution but that does have a power law tail.

This result does not at first sight help us with Ω, the relationship between Ω and the function g is not immediately apparent, so it is unclear whether this theorem can be applied to $Y = \Omega X$. The following theorem provides some guidance.

Proposition 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 $X = x$. The function g defined for all $y > 0$ as :*

$$\mathbf{P}[Y > y] = y^{-\alpha} g(y)$$

is a slowly varying function if there exists a constant $\kappa > 0$ and a function Φ_0 such that there is a thresholds $\overline{\omega}$ such that:

$$\begin{array}{lll} \forall \omega > \overline{\omega}, \quad \forall x > x_{min} & \Phi_{\Omega|X}(\omega|x) < \omega^{-\alpha-\kappa} & \textit{(thin-tailed condition)} \\ \forall x > \overline{x} & \lim_{x\to\infty} \Phi_{\Omega|X}(\omega|x) = \Phi_0(\omega) & \textit{(pseudo-independence condition)} \end{array}$$

Proof. See Appendix A □

[20] A function f is defined as slowly varying if and only if $\forall a > 0 \lim_{x\to\infty} \frac{f(ax)}{f(x)} = 1$.

[21] Beirlant, Vynckier, and Teugels (1996) offer guidance on choosing the optimal cut-off k_N of the tail for a limited sample of size N. Their procedure is to chose the optimal threshold using the minimization of the asymptotic mean square error.

Intuitively, this results allows one to estimate the power law exponent of X using Y as long as one is willing to make two assumptions on Ω. Both assumptions relate to the "top" of the conditional distribution of Ω, as far as the left end of Ω is concerned, anything is possible. No assumptions on the economic explanation or the exact structure of Ω are needed, one can stay entirely agnostic about the causes of the presence of Ω; as long as Ω satisfies the two conditions, the estimates of the power law exponent α using the tail of Y will be consistent.

The economic interpretation of the two conditions is as follows. The first condition is that, whatever the productivity draw (x in our notation), the distribution of the wedge for a certain type of productivity x is thinner-tailed than X. Intuitively, this means than the top of the exports distribution is not polluted by low productivity firms that had an extremely favorable draw of Ω. The second condition is that, for sufficiently high values of productivity, the wedge is "almost" independent of productivity. Intuitively, this means that we do not need to worry that the tail of Y is distorted by interactions between Ω and X. Is this the case in our models? In our first model, see equation (4), the first equation is true as long as we choose a distribution for η that is thinner tailed than X, which is the case for any realistic distribution of shocks.[22] The second equation is always verified given Ω is independent from X. In our second model, see equation (15), one should note 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. 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.

4.3 Identification

The question here is to what extend Y can help in identifying some distributional properties of X. The previous theorem has an immediate corollary.

Corollary 1. *X can only be Pareto distributed if the data Y is regularly-varying[23] with the same exponent as X or if Ω does not fulfill one of the two conditions in Proposition 1.*

This is simply the converse of Proposition 1. If Ω does not fulfill one of the two conditions in Proposition 1, there is little we can say about Y but we do have the following result if X and Ω are independent random variables.

Proposition 2. *Suppose that (i) X and Ω are independent non-negative random variables and (ii) X is regularly-varying at infinity with index $\alpha > 0$.*

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

[23]A positive function is regularly varying at infinity with index α if there exists a finite real α such that $\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha}$, $\forall t > 0$. Note the log-normal distribution is not regularly-varying at infinity since this limit can take three values: 0 (if $t > 1$), 1 (if $t = 1$), and ∞ (if $t < 1$).

- *If Ω is regularly-varying at infinity with index β with $\beta > \alpha > 0$, then X can not be identified in the upper tail of Y;*
- *If Ω is regularly-varying at infinity with index β with $\alpha \geq \beta > 0$, then X can be identified in the upper tail. In particular, if $\beta = \alpha$:*

$$\mathbf{P}[X\Omega > y] \underset{y\to\infty}{\sim} 2\mathbf{E}\left[X^{\alpha-1}\right]\mathbf{P}[X > y]$$

- *More generally, if $\mathbf{E}[\Omega^\alpha] < +\infty$, then*

$$\mathbf{P}[X\Omega > y] \underset{y\to\infty}{\sim} \mathbf{E}[\Omega^\alpha]\,\mathbf{P}[X > y]\,.$$

Proof. See Appendix A. □

Intuitively, if the upper tail of the product convolution ($Y > y$) displays a power law tail and if one assumes the α'th non-central moment of Ω exists, then the upper tail of the X distribution behaves as a power law with the same exponent as Y. Proposition 3 goes on step further and establishes some necessary conditions on Y and Ω such that X belongs to the power law family of distributions. Specifically, if the product convolution density is regularly-varying at infinity and Ω belongs to the exponential family, then the distribution X is also regularly-varying at infinity with the same index as Y.

Proposition 3. *Suppose that:*

(i) X and Ω are independent non-negative random variables;

(ii) The product convolution $X\Omega$ is regularly-varying with index $\alpha > 0$, $X\Omega \sim \mathtt{RV}(\alpha)$;

(iii) Ω or some transformations of Ω belongs to the exponential family (for $\omega > 0$):

$$f_\Omega \propto h(\omega)exp(C_2T(\omega)).$$

(iv) $h(\omega)\mathbf{P}\left[X > \omega^{-1}\right]$ is monotone in ω.

Then X is regularly-varying of index $\alpha > 0$ and X can be Pareto-type distributed.

Proof. See Appendix A. □

Relaxing the independence assumption requires that the conditional distribution $\Omega \mid X$ behaves like the marginal distribution of Ω up to a sufficiently large x. This is left for future work, we suspect that this is indeed a possile generalization of our theorem.

5 Conclusion

The main message of this paper is that the strategy that the Melitz (2003) model employs to identify productivity using exports data is knife-edge; it relies on assumptions that 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 on exports data may be more fragile than first thought. However, we also show that, under assumptions that we claim should be straightforward in practice, 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. This allows for a simple test: under the same assumptions, the right tail of productivity follows a Pareto law if and only if the right tail of exports does.

However, in the presence of Ω, it is not possible to identify the lower end of the distribution of productivity. Furthermore, it is also not possible to identify a truncated Pareto distribution for productivity using exports data. In both cases, estimating such a law can therefore only be done in two ways. On can make strong assumptions on the exact statistical nature of the wedge between exports and productivity and use deconvolution methods or on can make a parametric assumption on Ω, which is weaker, and use a structural estimation.

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.

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, 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.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122(3), 507 – 553.

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

Beirlant, J., P. Vynckier, and J. L. Teugels (1996): "Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics," *Journal of the American Statistical Association*, 91(436), 1659–1667.

Bernard, A. B., and B. J. Jensen (1999): "Exceptional exporter performance: cause, effect, or both?," *Journal of International Economics*, 47(1), 1–25.

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

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. (1999): "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 114(3), 739–767.

——— (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1(1), 255–294.

——— (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79(3), 733–772.

——— (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30(1), 185–206.

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.

GABAIX, X., AND A. LANDIER (2008): "Why Has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

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.

JONES, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *The Quarterly Journal of Economics*, 120(2), 517–549.

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

LUCAS, JR., R. E., AND B. MOLL (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1), 1 – 51.

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.

Perla, J., and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122(1), 52 – 76.

A Proof

A.1 Proof of Result 1

See Beirlant, Vynckier, and Teugels (1996) for the Hill estimator, Kratz and Resnick (1996) for the qq-estimator, Aban, Meerschaert, and Panorska (2006) for the "truncated" MLE estimator and Gabaix and Ibragimov (2011) for the log-rank estimator. $\square$

A.2 Proof of Proposition 1

We use the notation defined in the theorem. Let $f_X(x) = \alpha \frac{(x_{\min})^\alpha}{x^{\alpha+1}}$ be the probability density function of X defined for all of $\mathbb{R}^+$. For $y > 0$ and any $\lambda > 0$, we have by definition of g

$$
\begin{aligned}
& g(y) = y^\alpha \int_{x_{\min}}^{\infty} \Phi_{\Omega|X}\left(\frac{y}{x}; x\right) f_X(x) dx \\
\Rightarrow \qquad & \frac{g(\lambda y)}{g(y)} = \lambda^\alpha \cdot \frac{\displaystyle\int_{x_{\min}}^{\infty} \Phi_{\Omega|X}\left(\frac{\lambda y}{x}; x\right) x^{-\alpha-1} dx}{\displaystyle\int_{x_{\min}}^{\infty} \Phi_{\Omega|X}\left(\frac{y}{x}; x\right) x^{-\alpha-1} dx} \\
\Rightarrow \qquad & \frac{g(\lambda y)}{g(y)} = \frac{\displaystyle\int_{0}^{\frac{\lambda y}{x_{\min}}} \Phi_{\Omega|X}\left(\omega; \frac{\lambda y}{\omega}\right) \omega^{\alpha-1} d\omega}{\displaystyle\int_{0}^{\frac{y}{x_{\min}}} \Phi_{\Omega|X}\left(\omega; \frac{y}{\omega}\right) \omega^{\alpha-1} d\omega}
\end{aligned}
$$

where we have used the changes of variables $\omega = \frac{\lambda y}{x}$ and $\omega = \frac{y}{x}$ in resp. the numerator and the denominator. The pseudo-independence property assures us that the limit at $y \to \infty$ of $\Phi_{\Omega|X}\left(\omega; \frac{\lambda y}{\omega}\right)$ is well defined for any λ and ω. The thin-tailed property assures us that the integrand is dominated by $\omega^{-1-\kappa}$ for any y and λ, which is integrable at ∞. This allows us to invoke dominated convergence and take the limit "under the integral", and hence we have for all $\lambda > 0$:

$$
\lim_{y \to \infty} \frac{g(\lambda y)}{g(y)} = \frac{\displaystyle\int_0^\infty \Phi_0(\omega)\, \omega^{\alpha-1} d\omega}{\displaystyle\int_0^\infty \Phi_0(\omega)\, \omega^{\alpha-1} d\omega} = 1 \quad \square.
$$

A.3 proof of proposition 2

[To be completed...]

A.4 proof of proposition 3

[To be completed...]

B Estimators of a power law exponent

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

$$\mathbf{P}[X > x] = \begin{cases} 1 & \text{if} \quad x < x_{\min} \\ \dfrac{x^{-\alpha} - (x_{\max})^{-\alpha}}{(x_{\min})^{-\alpha} - (x_{\max})^{-\alpha}} & \text{if} \quad x_{\min} \leq x \leq x_{\max} \\ 0 & \text{if} \quad x > x_{\max} \end{cases} \tag{19}$$

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 quickly 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-rank regression The basic idea underlying a log-rank regression is that $\mathbf{P}\left[X > X_{(i)}\right] \approx \frac{i}{N}$ 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:

$$\ln \frac{i}{N} \approx \ln C - \alpha \ln X_{(i)}. \tag{20}$$

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:

$$\ln Q_i = \frac{\ln C}{\alpha} - \frac{1}{\alpha} \ln \frac{i}{N+1} \tag{21}$$

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 important things to note: first, the QQ-regression is nothing more than the reciprocal regression of the log-rank regression, which should be obvious from the two equations (20) and (21). This explains why the standard errors, biases etc. shown in figures 5 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:

$$\widehat{\alpha} = \frac{N}{\sum_{i=1}^{N} \left[\ln X_{(i)} - \ln X_{(N)}\right]} \tag{22}$$

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:

$$\frac{N}{\widehat{\alpha}} + \frac{N\left[X_{(N)}/X_{(1)}\right]^{\widehat{\alpha}} \ln\left[X_{(N)}/X_{(1)}\right]}{1-\left[X_{(N)}/X_{(1)}\right]^{\widehat{\alpha}}} - \sum_{i=1}^{N}\left[\ln X_{(i)} - \ln X_{(N)}\right] = 0 \tag{23}$$

C Extensions of the simulated example of section 4.1

D 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 effects—such as an expansion of product variety and a pro-competitive reduction in mark-ups—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. The purpose is to see to what extent the available data might be used to validate or invalidate this assumption.

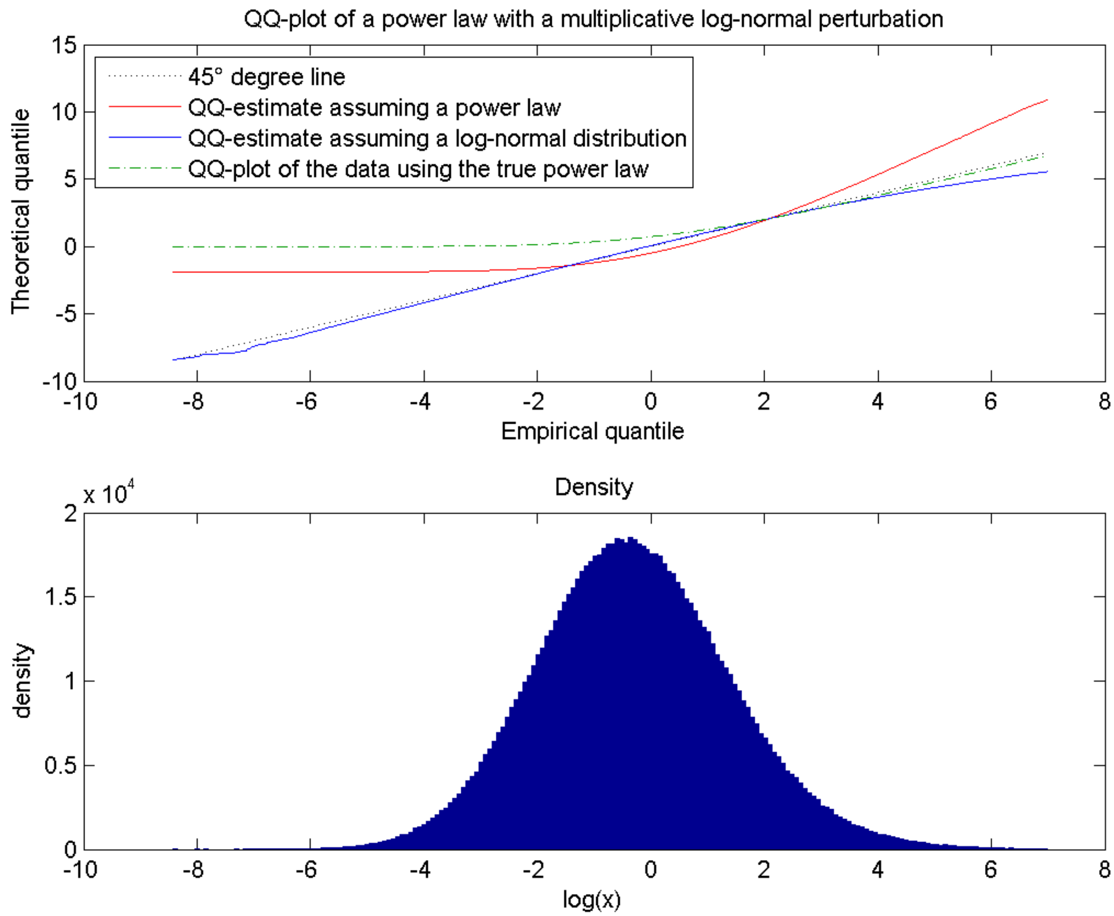

Figure 3: QQ-plots of a Pareto law with a multiplicative log-normal error (large sample)

Note: Data Y is generated according to the following generating process: $Y = X \cdot \Omega$, where X is a Pareto law with $\alpha = 1.2$, $x_{\min} = 1.0$ 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.

Our answer is that the assumption of a bounded Pareto law for productivity is more critical to test using exports data, and might be not feasible. Formally, as in Section 4, without any assumptions on Ω nothing can be said about X. But contrary to Section 4, Propositions 1-3 and the truncation-based estimation result in the statistics literature do not apply,[24] so even with Ω no 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.[25]

The following illustration may be revealing and useful. We look at a simple example that mimics the simulations from Section 4.1, 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 4 and 5. 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, we know the true distribution of Y. The only reason we do this is that the horse race Pareto law v. log-normal has gotten some traction in the literature. What is interesting is that the log-normal clearly outperforms the Pareto law. 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.[26]

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 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.[27] In practice, a warning sign 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 5 in the case of a non-truncated distribution for X). If the estimate of α keeps increasing,[28] one is dealing with data Y that is thinner tailed than a Pareto law. In practice (barring pathological distributions for Ω) this means that X is also thin-tailed, but this does not rule out the possibility that X is a truncated Pareto law or, heuristically, "almost" thick tailed.[29]

[24] g will never be slowly-varying if X is bounded.

[25] 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.

[26] One may be tempted to look for better tests. But without a distributional assumption on Ω, all tests would be misspecified by nature. This includes a formal goodness-of-fit test (say, 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, just a heuristic.

[27] This does not mean it is impossible. 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

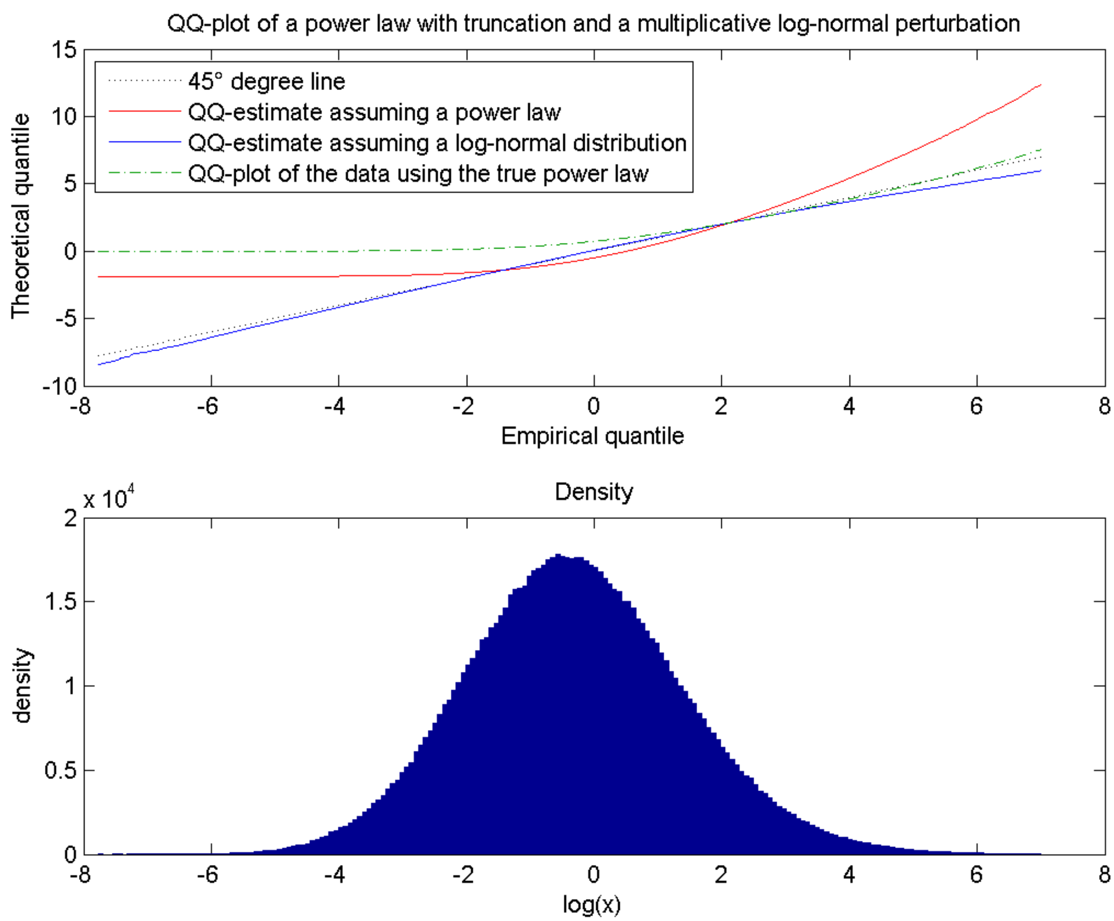

Figure 4: QQ-plots of a truncated Pareto law with a multiplicative log-normal error (large sample)

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.

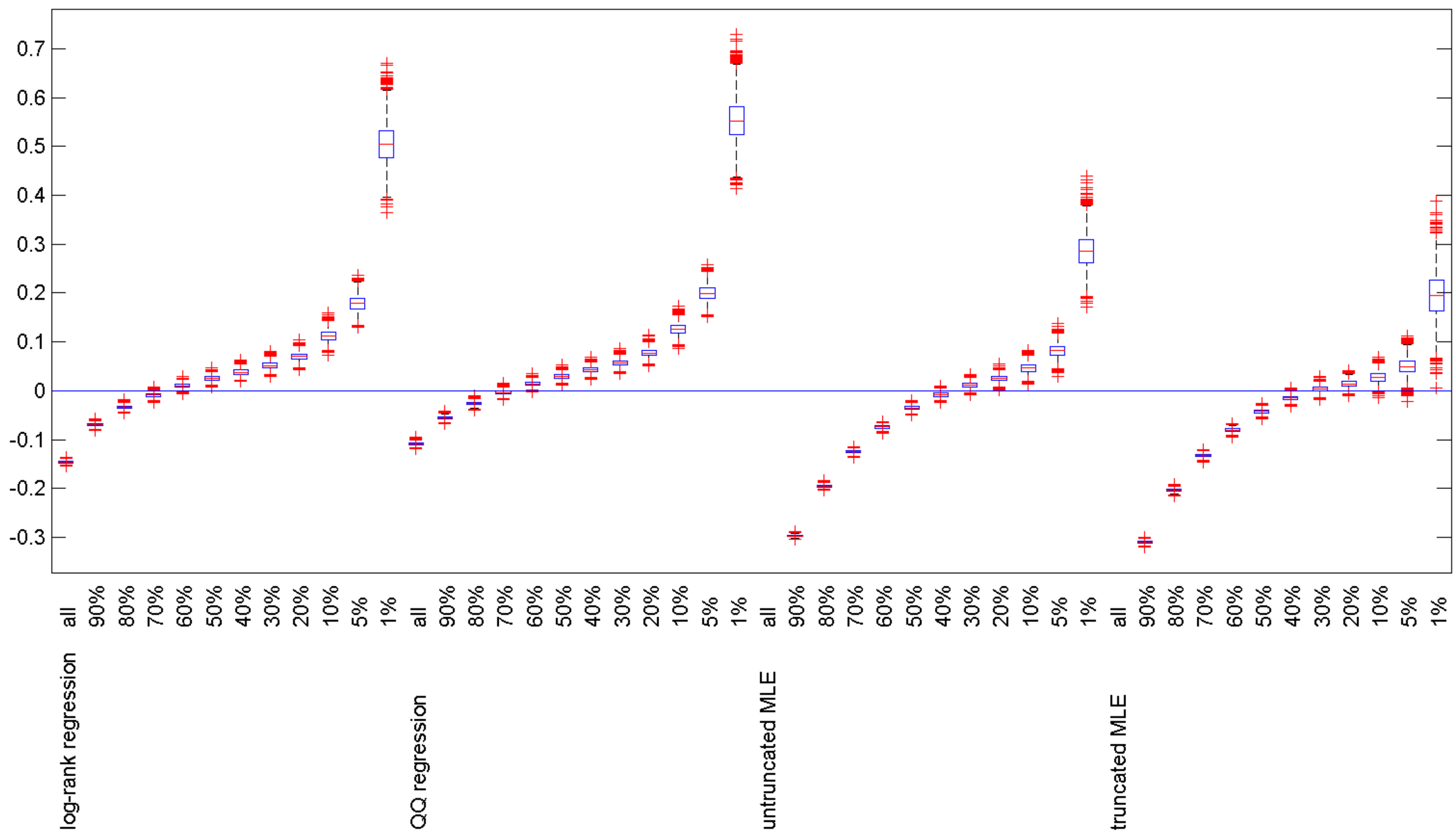

Figure 5: Monte-Carlo simulations of estimates of the power law exponent (truncated case).

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.

for future research.

[28]A good example of where this is happening is table 1 in Head, Mayer, and Thoenig (2014).

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