Pareto Distributions in International Trade:

Hard to Identify, Easy to Estimate.

Marnix Amand* Florian Pelgrin†

HEC Lausanne EDHEC Business School

April 30, 2016

Abstract

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 firm productivity. This implies that testing for an exact Pareto distribution for firm productivity is impossible, barring unreasonable 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 estimate the productivity distribution are misspecified. To overcome this, we provide a formal result with a straightforward economic interpretation that allows to estimate the tail of a power law irrespective of the presence of such a (correlated) wedge. Lastly, we show that this wedge can distort the left end of the observed exports, often making it "look" log-normal, even if productivity is exactly Pareto distributed. Hence our interpretation of recent empirical work that rejects a Pareto distribution in exports data is that these results do not necessarily reject the assumption of a Pareto distribution for productivity.

Keywords: Pareto, power law, international trade, productivity distribution, misspecification.

JEL Classification Numbers: F12, F14.

*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 Introduction

Due to its tractability, the Melitz (2003) heterogeneous firm international trade model augmented with a Pareto distribution[1] for productivity has become widely used in theoretical work. However, this model predicts that destination-specific exports are also Pareto distributed, 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 the empirical evidence rejects the Pareto assumption for productivity along an economically significant dimension, welfare, and thus cannot be ignored. Given these results and further recent empirical work, the empirical justification of the use of a Pareto distribution for productivity in the Melitz (2003) framework is increasingly seen as questionable, notwithstanding the Pareto's oh-so-useful properties for theorists.

This matters because the Pareto distribution allows to jump a major hurdle when solving the Melitz (2003) model. In this model, firms face a destination-dependent fixed entry cost and a downward-sloping demand curve. Only the more productive firms are able to export profitably, which leads to a minimal ("cut-off") productivity that determines which firms export to any given destination.[2] 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.

Not so with a Pareto distribution. 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

[1] A random variable X follows a Pareto distribution, also called Pareto law, 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, Pareto tail or Zipf's law, 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] This set-up 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.

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). 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.[3]

Hence the importance of the Pareto distribution in the Melitz framework. If indeed the Pareto distribution is ruled out as a good distributional model of productivity by the empirical evidence, none of the results cited in the previous paragraph would hold. 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. 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

[3] 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 in the international trade literature.

(or makes strong assumptions about) Ω.[4] 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, 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. This paper shows 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. 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. Only if one is willing to assume away *all* forms of heterogeneity (other than productivity), uncertainty and 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 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.

Once one is convinced that 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.

[4] If one knows the distribution of Ω, the estimation of the Pareto law exponent leads to a deconvolution problem written in multiplicative form. Maximum-likelihood methods would then yield consistent estimates of the parameters of the distribution of X. However, our whole point is that little is known in practice about Ω, which rules out standard deconvolution methods.

[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.

This misspecification issue is potentially very severe: one knows very little about unmodelled heterogeneity, misspecification of the underlying model, measurement error etc., so it is problematic to make precise assumptions about Ω. Furthermore, our model shows that it is hard to rule out that Ω is not independent from X.[6] With the exports data Y polluted by an unknown Ω possibly correlated with productivity X, can we still learn anything about X?

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. The second and main result of this paper is that this intuition is mostly correct: we provide a theorem that shows that, despite all these issues with Ω, a consistent estimation of a Pareto distribution for X is still often possible using data Y. The key idea 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. All one needs for this to be true is that, heuristically, Ω is not "too" heavy tailed (less than X) and not "too" correlated with X at the top. We show that these conditions have a simple economic interpretation, which allows one to apply the theorem using an economic (rather than a purely mathematical) justification.

To the best of our knowledge, this is new in economics. Our interpretation is that this is an "(almost) anything goes" result. To restate our result: the data of a Pareto distribution X can be perturbated by *any* process Ω, possibly correlated with X, as long as Ω is well behaved at the top, estimation of the power law exponent of X using the perturbated data is consistent.[7] This allows to overcome a severe misspecification issue with very weak (and economically easy to justify)

[6] A simple example of this possible dependence is stochastic market access costs: low-productivity firms will enter markets based on their draw, high-productivity firms will enter no matter their draw. To solve these issues, the previous empirical literature either 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 point is precisely that these (very restrictive) assumptions can be avoided.

[7] Note that Gabaix (2009, p. 259) makes the point that, when estimating a power law tail, "multiplying by normal variables, adding non-fat tail noise, or summing over independent and identically distributed (i.i.d.) variables preserves the exponent." However, he is writing about *independent* wedges, whereas our result applies to dependent wegdes too (which, we claim, is the relevant case in practice).

assumptions. To put this result in a larger context, we would like to highlight that the analytic convenience of the Pareto distribution has made its use very popular outside of international trade too, as we show in Section 2.2. Our main result is in principle applicable to any situation where one has data Y but needs inferences about a Pareto distribution X with $Y = \Omega X$. We argue that this could be the case in other fields and show in Appendix D three further possible applications of our theorem, giving our result a wider relevance.

Finally, coming back to the initial motivation of this paper, our result also 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. Furthermore, using both an informal but popular method, qq-plots, and a formal testing procedure provided by Malevergne, Pisarenko, and Sornette (2011), we show that 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 last 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.

The paper is organized as follows. Section 2 briefly sums up the previous empirical work on identification and estimation of the Pareto distribution in the international trade context and in a larger 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. Section 4 provides the formal results that allow one to identify and estimate a Pareto distribution despite the presence of this wedge Ω. In Section 5, we illustrate with an example that the presence of Ω makes identification and estimation problematic without the previous result, and Section 6 concludes. Most proofs are gathered in an appendix.

2 Pareto distributions in economics

In this section, we briefly discuss the use of the Pareto distribution in the international trade literature and in the broader economics literature.

2.1 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.[8] 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 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.[9,10] 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 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

[8] 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.

[9] 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.

[10] 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).

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 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.[11]

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 "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 F some issues regarding the identification and estimation of a bounded Pareto law for productivity in the presence of a wedge Ω.

2.2 Other fields

The Pareto distribution has also widely gained in popularity outside of the international trade literature. 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). Closely related to these papers are Benhabib, Bisin, and Zhu (2011) and Gabaix, Lasry, Lions, and Moll (2015) who use a multiplicative growth process perturbated by an income process to generate a Pareto tail for the wealth distribution, but they do not predict an exact Pareto distribution in the data.[12]

Just as with the Melitz model, theorists get a lot of mileage out of the Pareto assumption, in particular its scale-free aspect. Using an exact Pareto distribution yields several useful properties and elegant results that are difficult to micro-found otherwise, such as exponential aggregate growth (irrespective of the level of knowledge, there are always enough better ideas to search for, Kortum,

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

[12]For further papers that use this assumption, see the reviews in Gabaix (2009, 2016).

1997), the avoidance of a wealth-effect on saving decisions (even the very rich continue investing, given the high rewards they face, Benhabib, Bisin, and Zhu, 2011) and the micro-founding of the Cobb-Douglas production function (Jones, 2005). And just as in the empirical international trade literature, it is unclear whether these distributions are indeed Pareto. One observes very similar Pareto v. log-normal debates across JEL codes. In the city-size literature, Eeckhout (2004, 2009) argues that the city-size distribution is log-normal, not Pareto, whereas Gabaix (1999), Levy (2009) and Rozenfeld, Rybski, Gabaix, and Makse (2011) argue that it is Pareto-tailed. In the firm-size literature, Segarra and Teruel (2012) provide a thorough literature review, with again papers in favor of a lognormal distribution and papers in favor of a pronounced Pareto tail. In the wealth distribution literature, good data is harder to come by, but the available evidence (Piketty and Zucman, 2014a,b; Klass, Biham, Levy, Malcai, and Solomon, 2006; Brzezinski, 2014) points towards a Pareto tail but a log-normal lower end (which, not incidentally, is also what Benhabib, Bisin, and Zhu, 2011, aim to model). Reinforcing this, Toda and Walsh (2015) show that consumption also has a Pareto tail.[13]

In all these examples, the empirical results raise the possibility that the "true" hidden Pareto distribution (e.g., ideas or productivity) X might be perturbated by a wedge Ω, i.e., that the data Y is structured as $Y = \Omega X$. We apply our formal results to three examples from the above-cited literature in Appendix D.

3 Observed exports and productivity: a general model

In this section, our goal is to explain why the presence of Ω in the data is ubiquitous. To do this, we provide a 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 B, we provide an additional model where Ω is independent from X.

[13] In Gabaix (2009, figures 4 and 5), further graphical examples can be found from the finance literature where again the right tail is Pareto but the lower end of the distribution is not.

Setup. Our starting point is the canonical Melitz model with endogenous variable market access costs following Arkolakis (2010). 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 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.[14]

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}.[15] 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:

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

[14] 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.

[15] 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.

Optimal strategy for an exporting firm. 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:

$$q = m\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} p_n(q, m, \epsilon)^{-\sigma} \tag{3}$$

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:

$$\begin{aligned} q_n(\varphi, \epsilon) &= m_n(\varphi, \epsilon)\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1} \left(\frac{\varphi}{\tau}\right)^{\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \\ \text{and} \quad m_n(\varphi, \epsilon) &= 1 - \left[\frac{\epsilon_{\mathrm{d}} R_n P_n^{\sigma-1}}{\epsilon_{\mathrm{m}}} \left(\frac{\varphi}{\tau}\right)^{\sigma-1} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\right]^{-\lambda}. \end{aligned}$$

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 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:

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

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:

$$
\begin{aligned}
&\text{Target market size:} && m_n(\varphi,\xi) = 1 - \xi_2 \overline{\pi}_n^{-\lambda} \varphi^{\lambda(1-\sigma)} \\
&\text{Quantity exported:} && q_n(\varphi,\xi) = \xi_1 m_n(\varphi,\xi) \frac{\varphi(\sigma-1)}{\tau} \overline{\pi}_n \varphi^{\sigma-1}, \\
&\text{Sales:} && r_n(\varphi,\xi) = \xi_1 \sigma m_n(\varphi,\xi) \overline{\pi}_n \varphi^{\sigma-1} && (4) \\
\text{and}\quad &\text{Profits:} && \pi_n(\varphi,\xi) = \xi_1 \overline{\pi}_n \varphi^{\sigma-1} \frac{\lambda - m_n(\varphi,\xi)}{\lambda - 1} - \xi_3. && (5)
\end{aligned}
$$

Entry decision. Firms enter a market n only if their profits $\pi_n(\varphi, \xi)$ are positive. Two conditions are necessary for this. First, a firm must at least cover all its costs, given by equation (5), and second, as indicated previously, a firm must target a non-negative market size:

$$
\begin{aligned}
\pi_n(\varphi,\xi) &\geq 0 \\
m_n(\varphi,\xi) &\geq 0.
\end{aligned}
$$

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:

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

Lastly, the initial entry decision is

$$\mathbf{E}\left[\sum_{\substack{n \\ \varphi \geq \varphi_n(\xi)}} \max\Big(0, \pi_n(\varphi,\xi)\Big)\right] \geq f_E. \tag{6}$$

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 $\delta_n(\varphi, \xi)$ be defined as

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

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

$$r_n(\varphi, \xi) = \delta_n(\varphi, \xi)\xi_1 m_n(\varphi, \xi)\sigma\overline{\pi}_n\varphi^{\sigma-1}. \tag{8}$$

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 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 Estimation and identification

Once one has concluded that the data Y is distributed as in (1), with Ω determined by some version of our model (or possibly by some other process), this leads to two questions, which we both address in this section. If one is willing to assume that X is exactly Pareto distributed, can the tail of Y be used to estimate X, and if Y has a power law tail, does this imply the same for X?[16]

[16] It may seem more natural to look at identification before estimation. We chose the reverse order because our main result pertains to estimation, identification remains an open question and the sole result we have builds on our estimation theorem.

4.1 Main result

Regarding estimation, we build on an important result from the statistics literature that proves that power laws (i.e., "approximate" Pareto distributions) can be consistently estimated using a large class of estimators of exact Pareto distributions using only the right tail of the data. We consider 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).[17] For easy reference, we state this result formally:

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[18] *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 most common estimators of α mentioned above 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.[19] Simply put, if one is estimating the power law tail of a distribution that might not be

[17] We recapitulate these estimation methods in Appendix C.

[18] See footnote 1.

[19] Among others, 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.

exactly Pareto distributed, all the usual estimators of a power law exponent remain consistent if used on the right tail only.

There is however no existing result that allows one to use the previous theorem with data $Y = \Omega X$ with X Pareto distributed. The difficulty lies in the correlation between Ω and X, which means there is no obvious one-to-one link between X and $y^{-\alpha}$ on one hand and Ω and the slowly-varying function g on the other hand. We solve this problem with the following theorem:

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

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

is a slowly varying function if there exist constants $C > 0$ and $\kappa > 0$ and a function Φ_0 such that:

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

Proof. See Appendix A □

Intuitively, Theorem 1 allows one to estimate the power law exponent of X using Y as long as one is willing to make three assumptions on 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 a pathological case where 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.

Interestingly, the first two conditions are related to the concept of (quasi) asymptotic independent random variables. Resnick (2002, 2007) and Maulik, Resnick, and Rootzén (2002) have introduced the weaker notion of asymptotic independence (for two identically distributed random variables), generalized to quasi-asymptotic independence for any two random variables by Chen and Yuen (2009).[20] It is straightforward to show that the two first conditions of the theorem imply that X and Ω are quasi-asymptotically independent. However, these papers work on a larger class of functions and their results are not directly applicable to our context. Notice also that a result similar (and somehwat more general) to ours has be proven by Jessen and Mikosch (2006, lemma 4.2), but only for the case of two independent random variables.

4.2 Application to international trade

The economic interpretation of the three 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. 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 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. Is this the case in our model? In our model, see equation (8), 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.[21] 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

[20] Two non-negative random variables A and B with distributions F_A and F_B are said to be quasi-asymptotically independent if and only if

$$\lim_{z\to\infty} \frac{\mathbf{P}[A > z, B > z]}{(1 - F_A(z)) + (1 - F_B(z))} = 0$$

Intuitively, this means that an extremely large value for A (resp. B) is unlikely to occur at the same time as an extremely large value for B (resp. A)

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

particular case merely technical and always verified.

More generally, we would argue that it is difficult to economically justify an Ω that would not fulfil all three criteria. Ω would need to be a wedge that either (i) is thicker tailed than productivity, (ii) does not converge point-wise for $x \to \infty$ or (iii) converges point-wise to 0 for $x \to \infty$. An example of the first case 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. 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. An example of the third case would be a case where the Pareto distribution is not truncated but the data is truncated by Ω. We deal with a truncated Pareto distribution X in Appendix F. It is possible to construct an realistic 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.

An interesting 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 $\lim_{x\to\infty} \mathbf{P}[\,\Omega = 1 \,|\, X = x\,] = 1$, meaning the largest firms export everywhere. Hence Ω trivially fulfils 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.

4.3 identification

The last question is identification. To what extent can Y help in identifying some distributional properties of X? We assume that the data at hand Y displays a power law tail[22]. 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.*

In simpler words, 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. An immediate application would be Head, Mayer, and Thoenig (2014): they do not find a power law tail in their exports data. They conclude, heuristically, that this rules out a Pareto distribution in productivity. Our results shows that their conclusion is indeed formally warranted, and this is in a very general setting, i.e. when Ω fulfils the conditions of the theorem.[23]

5 The Pareto/log-normal debate in the presence of misspecification

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 is ignored by or unbeknownst to the econometrician. Essentially, what are the consequences of misspecification given our Theorem 1 and Corollary 1?

[22]By this we mean that the counter-cumulative distribution of Y displays a power law tail. See footnote 1 for the definition.

[23]Another interesting question would be whether there are any *sufficient* conditions on Y that imply a power law tail for X. Not exactly (to the best of our knowledge and ability) but we refer the interested reader to Lemma 4.3 in Jessen and Mikosch (2006), which shows that if Y is a regularly varying random variable with index α, a class of distributions that includes but is larger than Pareto tails with exponent α, and if Ω is both independent from X and sufficiently thin-tailed, then X is also a regularly varying function with index α. There are no known results if Ω is not independent from X or if Ω is not thin-tailed enough (we refer to the cited paper for the precise conditions).

A qq-plot exploration To highlight the misspecification issue, we assume that Ω and X are independent and are distributed as log-normal and Pareto, respectively. In essence, a multiplicative measurement error model.[24] We proceed with an horse-race for Y between the log-normal and the Pareto distribution. 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.[25] 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.[26]

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

[24]See Appendix B for a model that can generate such a data structure. In preliminary work, we have also looked at a case where Ω is correlated with X. We obtain results that yield the same conclusions as the ones we obtain below (as could be anticipated given Theorem 1).

[25]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.

[26]As mentioned earlier, Head, Mayer, and Thoenig (2014) apply this procedure using 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.

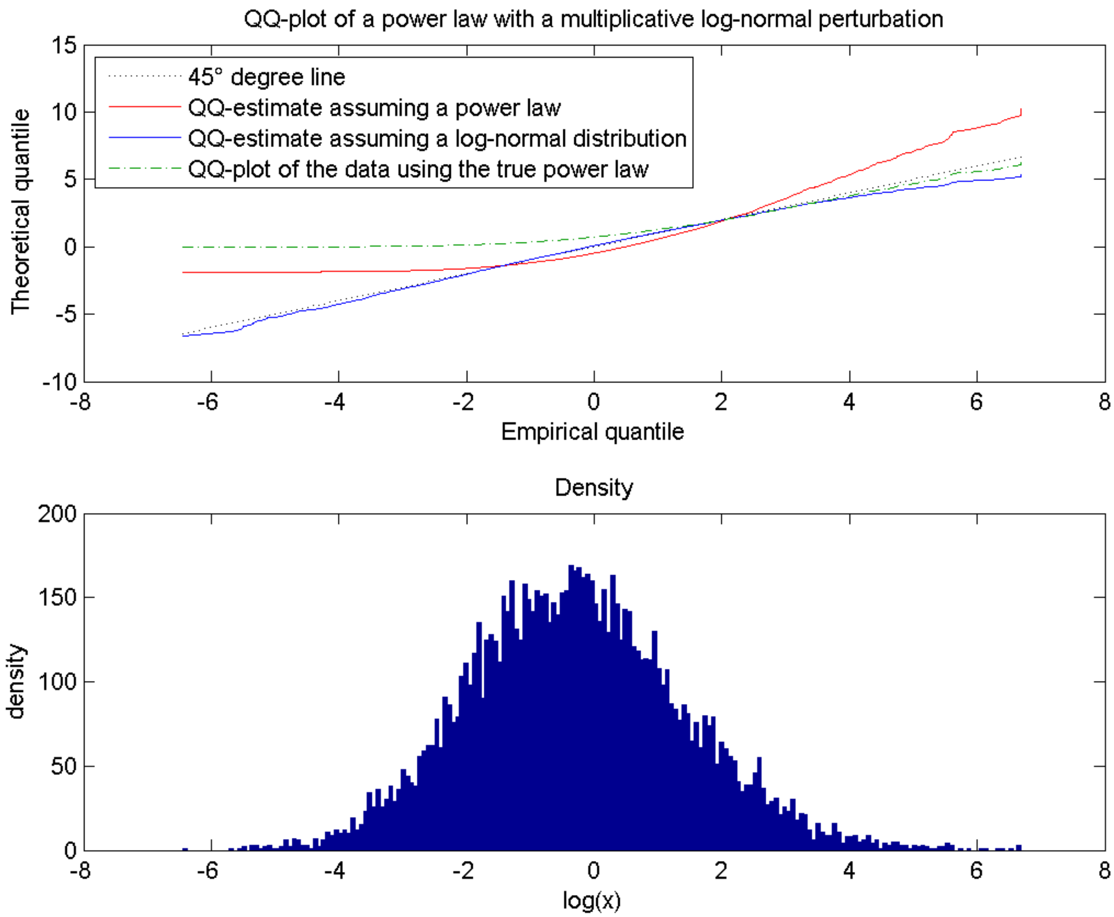

Figure 1: qq-plots of a Pareto law with a multiplicative log-normal error (small sample)

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.

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, 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.

One should note that our example is not driven by small sample bias. In Figure 3 in Appendix E, 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 conduct the uniformly most powerful unbiased test proposed by Malevergne, Pisarenko, and Sornette (2011).[27] 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.[28] 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

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

[28] 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.

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.

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 data 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 Result 1. Since Result 1 and Theorem 1 impose to left-truncate the data for consistent estimation, 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, our simulations also illustrate that left-truncating the data (i.e., dropping the lowest points) does allow for consistent estimates, which we have proven in Theorem 1. Note that this is done in practice by many papers without further justification.

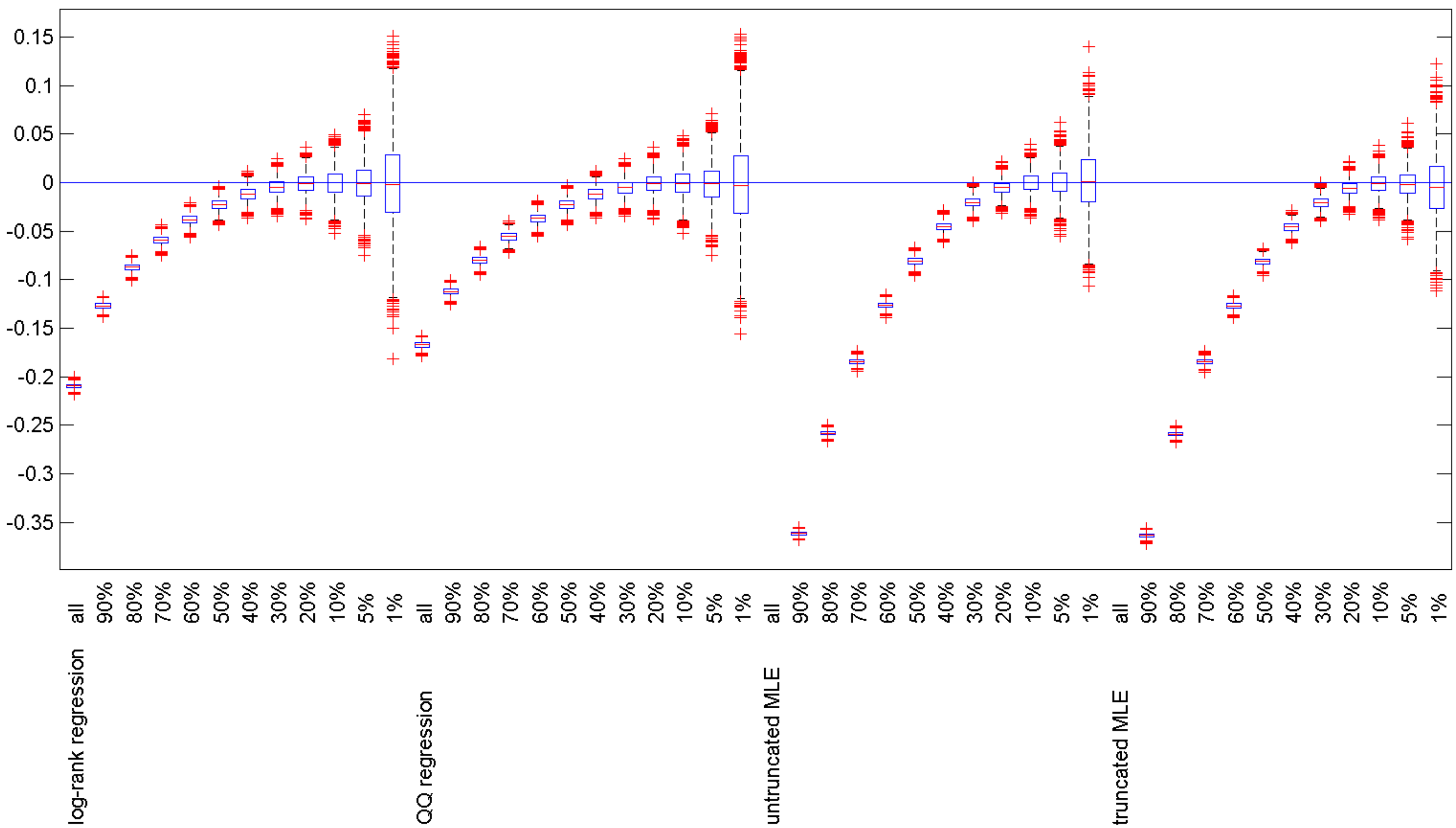

Figure 2: Monte-Carlo simulations of estimates of the power law exponent (untruncated case).

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.

Summing up, we have shown that the complete productivity distribution X cannot be generally inferred in the presence of a wedge Ω if one ignores the presence of 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, such as $\Omega \equiv 1$).

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 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. 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. These results are not limited to the international trade context but are in principle valid anytime one wishes to identify or estimate a Pareto distribution.

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, 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.

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.

BENHABIB, J., A. BISIN, AND S. ZHU (2011): "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, 79(1), 123–157.

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

Brzezinski, M. (2014): "Do wealth distributions follow power laws? Evidence from rich lists," *Physica A: Statistical Mechanics and its Applications*, 406(C), 155–162.

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.

Chen, Y., and K. C. Yuen (2009): "Sums of Pairwise Quasi-Asymptotically Independent Random Variables with Consistent Variation," *Stochastic Models*, 25(1), 76–89.

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.

Eeckhout, J. (2004): "Gibrat's Law for (All) Cities," *American Economic Review*, 94(5), 1429–1451.

——— (2009): "Gibrat's Law for (All) Cities: Reply," *American Economic Review*, 99(4), 1676–83.

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.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2015): "The Dynamics of Inequality," NBER Working Papers 21363, National Bureau of Economic Research, Inc.

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.

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

Klass, O. S., O. Biham, M. Levy, O. Malcai, and S. Solomon (2006): "The Forbes 400 and the Pareto wealth distribution," *Economics Letters*, 90(2), 290–295.

Kortum, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65(6), 1389–1420.

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

Levy, M. (2009): "Gibrat's Law for (All) Cities: Comment," *American Economic Review*, 99(4), 1672–75.

Lucas, Jr., R. E., and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1), 1 – 51.

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.

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.

Piketty, T., and G. Zucman (2014a): "Capital is Back: Wealth-Income Ratios in Rich Countries 17002010," *The Quarterly Journal of Economics*, 129(3), 1255–1310.

——— (2014b): "Wealth and Inheritance in the Long Run," in *Handbook of Income Distribution*, ed. by A. Atkinson, and F. Bourguignon, vol. 2B of *Handbook of Income Distribution*, chap. 20, pp. 1303–1368. Elsevier.

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.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011): "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review*, 101(5), 2205–25.

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

Segarra, A., and M. Teruel (2012): "An appraisal of firm size distribution: Does sample size matter?," *Journal of Economic Behavior & Organization*, 82(1), 314–328.

Toda, A. A., and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123(5), 1177 – 1200.

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

$$g(y) = y^\alpha \int_{x_{\min}}^{\infty} \Phi_{\Omega|X}\left(\frac{y}{x}; x\right) f_X(x) dx$$

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

$$\Rightarrow \qquad \frac{g(\lambda y)}{g(y)} = \frac{\displaystyle\int_{0}^{\frac{\lambda y}{x_{\min}}} \Phi_{\Omega|X}\left(\omega; \frac{\lambda y}{\omega}\right) \omega^{\alpha-1} d\omega}{\displaystyle\int_{0}^{\frac{y}{x_{\min}}} \Phi_{\Omega|X}\left(\omega; \frac{y}{\omega}\right) \omega^{\alpha-1} d\omega}$$

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 $\lambda > 0$:

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

The third condition is used in the last equality, to avoid the pathological case of having a limit of $\frac{0}{0}$.

B A special case: Ω behaving as measurement error

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 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[29] 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. We argue that such a "measurement error interpretation" of Ω can also be justified from a theoretical perspective.[30] 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 "in-

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

[30] 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.

tended 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 again follow the (very standard) 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:

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

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

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

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

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

$$k_n(\varphi) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} R_n P_n^{\sigma-1} \varphi^{\sigma} \tau_n^{1-\sigma} \overline{\epsilon}^{\sigma}$$
$$\text{with} \quad \overline{\epsilon} = \mathbf{E}\left[\epsilon_{\rm d}^{\frac{1}{\sigma}} \epsilon_{\rm s}^{\frac{1}{\sigma}-1}\right].$$

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

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

with the following notation:

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

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:

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

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

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

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

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

Entry decision. Firms choose to enter a market before discovering the values of $\epsilon_{\rm s}$ and $\epsilon_{\rm d}$. 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

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

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:

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

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.

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 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-size 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:

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

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 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 (17) and (18). This explains why the standard errors and biases 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:

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

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 Further applications of Theorem 1

We look at three examples from Section 2.2.

Wealth distribution Data on wealth is always tax data, meaning we observe wealth after tax optimization (and possibly, evasion). Obviously, the amount and percentage of wealth hidden (legally or not) from the taxman is correlated with pre-tax wealth—the rich have access to better tax optimization schemes—and stochastic, it is not entirely determined by the wealth level. If Y is the data, X the true wealth distribution and Ω the percentage of wealth not in the data, we have exactly $Y = \Omega X$. Can Y still be used to infer anything about X given our lack of knowledge about tax avoidance (i.e., Ω)?

To apply Theorem 1, one needs to assume that Φ_0 is not degenerate, i.e. equal to 0 for $\omega > 0$. If this were the case, this would mean for high enough wealth, tax evasion would be 100%. As long as one is willing to assume that *some* wealth is reported, and that the tax evasion process converges for high wealth and is thin tailed (two innocuous assumptions), Y can indeed be used to estimate the power exponent of X without any further knowledge of Ω.

Consumption Toda and Walsh (2015) show that consumption has a power law tail. However, data sources on consumption are notoriously imperfect, there are missing values, imputations, durables is often missing etc. If Y is the consumption data, X the true consumption distribution and Ω the effect of all these imperfections, we again have $Y = \Omega X$. The same question applies: can Y still be used to infer anything about the "true" consumption distribution X given our lack of knowledge about all the data imperfections summarized in Ω?

One needs to assume that these imperfections are not thicker-tailed than the "true" consumption distribution, that there is no truncation and that for high levels of consumption the data imperfections become independent of X. Neither of these assumptions seems particularly difficult to make, hence our theorem allows to conclude that the choice of consumption data of Toda and Walsh (2015) is not an obstacle to consistent estimates (we cannot say that the estimates are necessarily consistent, since there may, of course, be other issues with the data that preclude this).

City-size distribution One issue with the city-size distribution is that some datasets use administrative boundaries (Boston and Cambridge, MA, are two different cities) which caps most city sizes at arbitrary levels (the city boundaries, mostly set decades if not centuries ago). If this random cap is Ω, Y is the data and X the "true" city-size distribution, with a city is defined as a unified urbanized area irrespective of administrative boundaries, we again have the structure $Y = \Omega X$.

However, in this case Ω does not fulfil the third condition of the theorem, hence there is no guarantee that a Pareto distribution in X "carries over" to Y. One cannot use the identification of Y as log-normal as informative about X. Interestingly, empirical studies that do not use administrative city limits (such as Rozenfeld, Rybski, Gabaix, and Makse, 2011) do actually find a Pareto tail for Y (and hence, per our corollary, for X).

E Extensions of the simulated example of Section 5

See Figure 3.

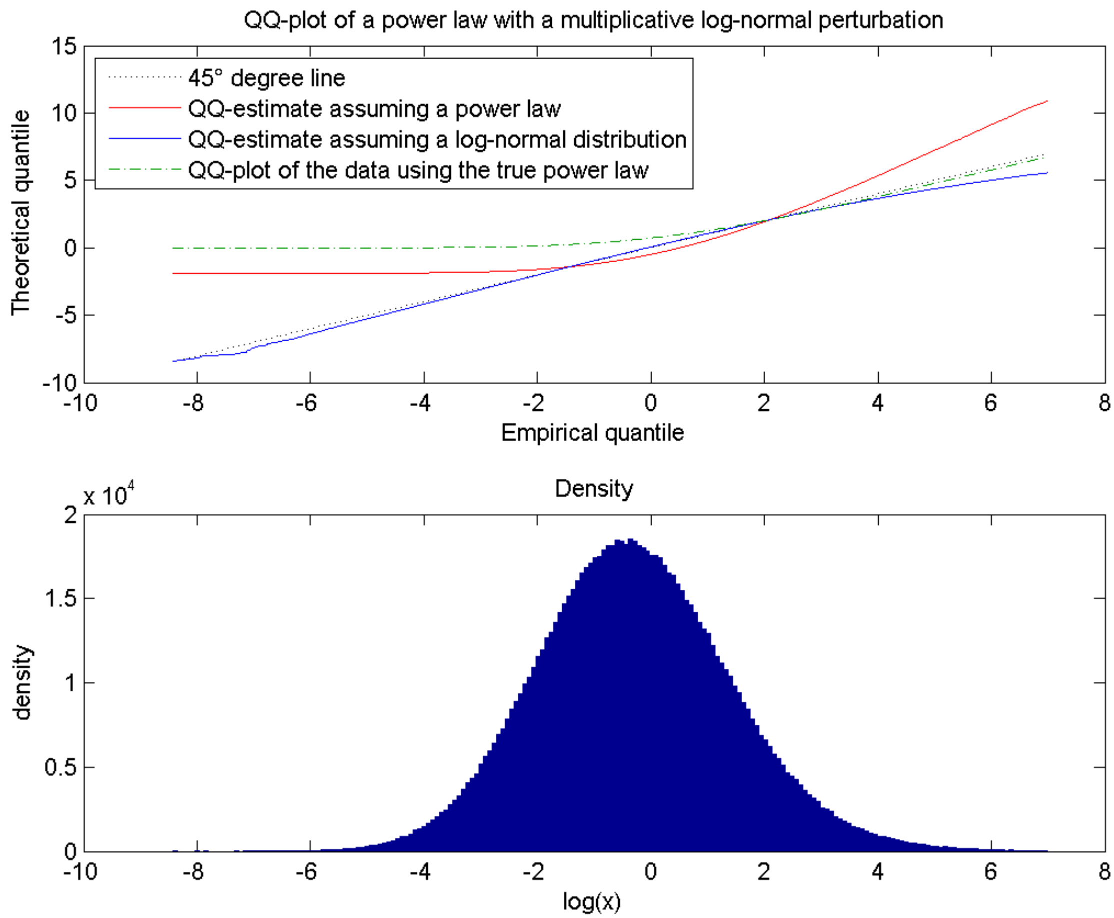

Figure 3: qq-plots of a Pareto law with a multiplicative log-normal error (large sample)

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.

F Productivity as a truncated Pareto law

As mentioned in the literature review in Section 2.1, 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.

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 4, Result 1 and Theorem 1 do not apply,[31] 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.[32]

The following illustration may be revealing and useful. We look at a simple example that mimics the simulations from Section 5, 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. 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

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

[32] 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.

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.[33] 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 2 in the case of a non-truncated distribution for X). If the estimate of α keeps increasing,[34] 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.[35]

[33]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.

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

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

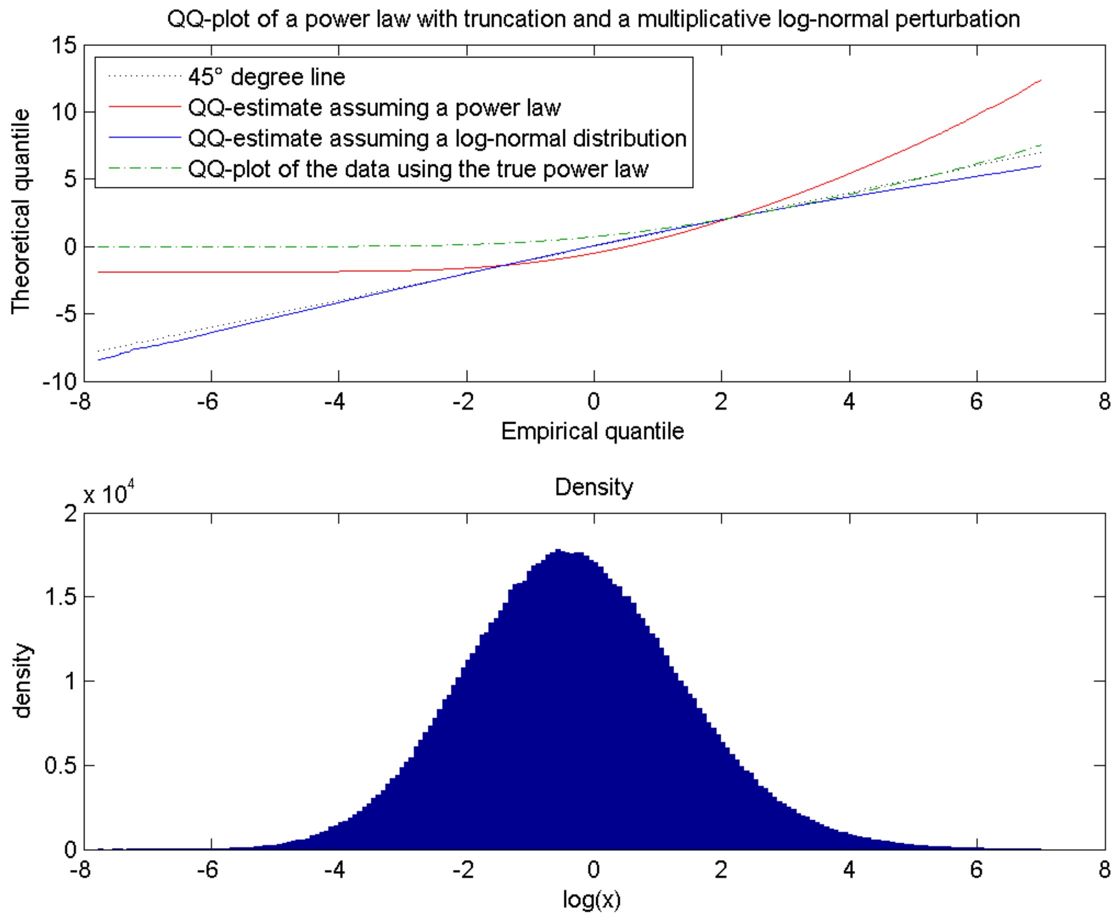

Figure 4: qq-plots of a truncated Pareto law with a multiplicative log-normal error (large sample)

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.

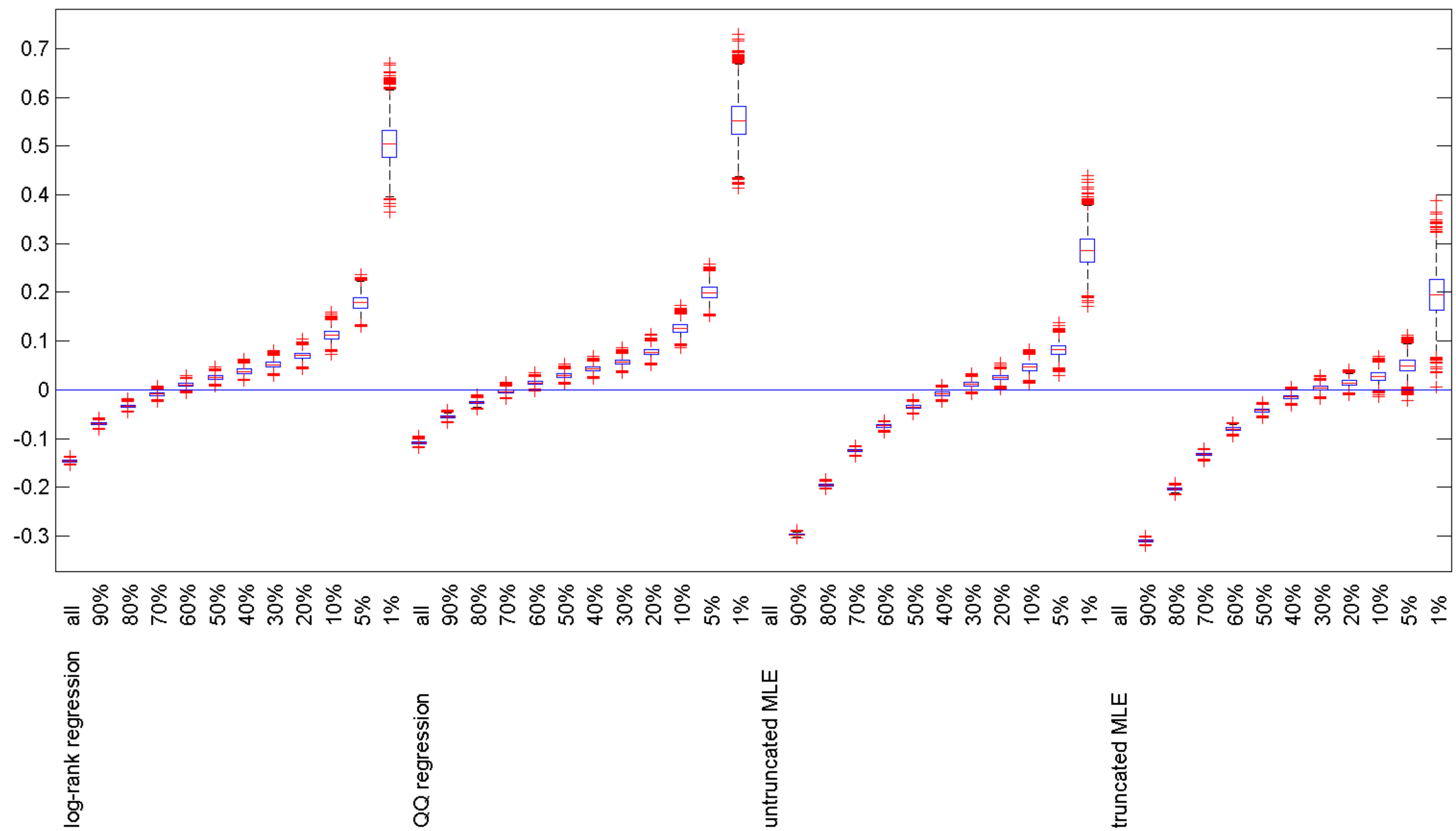

Figure 5: Monte-Carlo simulations of estimates of the power law exponent (truncated case).

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 $\frac{\widehat{\alpha} - \alpha}{\alpha}$.

Results that are not shown lie outside the graph.