

Pareto or not? The Intensive and Extensive Margins of Trade and Trade Elasticity

Vladimir Tyazhelnikov^{*†}

July 29, 2016

Abstract

Firms differ in size and productivity with important implications for trade policy and measuring gains from trade. Distribution of firms' productivities is then a central object in a model with heterogenous firms. I introduce a new way to estimate the shape parameter of the Pareto distribution for firm productivity using the data on firm-level imports to the US. I provide my estimates for about 600 US industries at the HS-4 level under assumptions of both CES and translog utility functions and offer few alternative specifications. I improve this estimator by allowing for the case of bounded Pareto distribution, making it robust to misspecification. In order to check the validity of distributional assumptions I provide a new way to test them. My first finding is that it is important to allow parameters of productivity distributions to vary by industry. Then I show that for most industries the distribution is Pareto rather than log-normal and bounded Pareto rather than unbounded. Finally I show that the left tail of firms' productivities distribution is unlikely to belong to Pareto distribution. I propose a new composite distribution consistent with the data that allows productivities of small and large firms to follow different distributions. This flexible distribution matches the facts on intensive and extensive margins of trade and is consistent with firm level evidence.

^{*}I am grateful to Robert Feenstra, Katheryn Russ, Deborah Swenson, Ina Simonovska, Brian Kovak, Vasco Yasenov, Mingzhi Xu and Konstantin Egorov for helpful comments, as well as seminar participants at UC Davis.

[†]University of California, Davis. E-mail: vtyazhelnikov@ucdavis.edu. Website: vtyazhelnikov.com

1 Introduction

Heterogeneity of firms is important for international trade. It is well documented that there is significant amount of redistribution between firms in case of trade liberalization. To know the consequences of this redistribution one has to know which firms gain and which lose. One of the most popular ways to take into account firms' heterogeneity is to assume their productivities follow some statistical distribution. Non-parametric approach to distributions is getting more popular (e.g. Hottman et al. (2014)), still classic parametric approach is relevant: distributional assumptions allow for analytical solution, require less computational time and feasible when aggregated data on trade flows are used. There are still debates what distribution accounts for firms' heterogeneity better. In this paper I propose a novel way to choose between distributions, introduce a new way to estimate parameters of the most popular Pareto distribution on fairly disaggregated level and propose a new distribution that combines strengths of currently used ones and hence can be a compromise in a current debate. My findings allow to accurately describe the impact of trade shocks on both intensive and extensive margins of trade.

Melitz (2003) proposed a model where in the case of trade liberalization the least productive firms leave the market and the most productive firms benefit. It shows how trade costs affect intensive and extensive margins of trade. Chaney (2008) showed that Melitz (2003) has a simple analytical solution under assumption of Pareto distribution. The reason it happens is its curious property: Pareto distribution trimmed from below is also a Pareto distribution with the same shape parameter. In Chaney (2008) lower bound of distribution is an endogenous object that depends on trade costs and other parameters of the model, this property allows for a simple analytical expression. Chaney (2008) shows that elasticity of trade with respect to trade costs is equal to the value of shape parameter. Besides Arkolakis et al. (2012) show that gains from trade in this model has a simple expression are inversely related to the shape parameter. Due to its simplicity and tractability unbounded Pareto is the most popular assumption in the class of Dixit-Stiglitz heterogeneous firms models. But elegance comes with a price. There are four main problems associated with Pareto distribution in Chaney (2008) model:

1. As there is non-zero probability that a firm of arbitrary size exists, the model fails to generate zero trade flows.
2. Trade elasticity between any two countries is constant and equal to the shape parameter. It is inconsistent with the data.
3. Under Pareto distribution firms' average sales is constant hence the model is unable to generate extensive margin.
4. Pro-competitive and variety sources of gains from trade are absent in the model. It does not allow to analyze different sources of gains from trade.

In order to fix the first problem Helpman et al. (2008) use bounded Pareto distribution combined with CES utility. Their theoretical model provides justification to two stage gravity equation, where probability

that trade happens is estimated on the first step.

Head et al. (2014), Bas et al. (2015) address the second problem and suggest to use log-normal distribution. They find that ratio of average to minimum sales is a sufficient statistics for log-normal distribution and show that trade elasticities generated on the basis of log-normal distribution approximate data better. Fernandes et al. (2015) show that assumption of log-normal distribution solves the third problem. In order to decompose the gains from trade by three different sources, Feenstra (2014) uses translog utility and bounded Pareto distribution.

Dealing with distributions I propose to use more rigorous and structured approach. First of all whenever it is possible all distributional assumptions should be justified by empirical distribution. Distributional assumption that matches some moments well, but is inconsistent with the data will generate results with the limited interpretation. For example Head et al. (2014), Bas et al. (2015) and Fernandes et al. (2015) use log-normal distribution, without trying bounded Pareto, even though it could also fix the problems the authors were solving¹. My approach is closer to Mrazova et al. (2015). They compare the performance of heterogenous firms model (in the sense of Kullback-Leibler divergence of sales or mark-up distributions observed in the data) under a variety of assumptions on utility function and under log-normal and Pareto distributions. I propose two statistical tests that allow to distinguish between most popular distributions: test of Pareto against log-normal and test of bounded against unbounded Pareto.

Another finding I make is that it is important to deal with distributions on disaggregated level. Most of the papers above assume that productivities of all the firms in the economy follow the same distribution Di Giovanni et al. (2011) do it for 24 industries and Bas et al. (2015) do it for 21 industries. Their results indicate that there is significant variation in shape parameter estimation from industry to industry. I show that this variation in the parameter estimates is even larger on the more disaggregated level. Then I perform the test of Pareto against log-normal distributions. For 87% of HS-4 industries the hypothesis it is Pareto is accepted. On the other hand under the assumption that all firms have the same shape parameter, Pareto hypothesis is rejected. Mixture of Pareto distributions with different shape parameters is not Pareto distribution and can look more like log-normal on aggregated level. Industry specific distributions resolve a problem with constant trade elasticity between the countries: it is still constant on industry level, but given that country pairs have different composition of industries in their trade flows, aggregate trade elasticity will also be varying. This approach is not a panacea: there might be some within industry variation in trade elasticities, but at least it takes care of the between industry component. Finally Levchenko and Zhang (2014) show that gains from trade estimates made on disaggregated level are different from aggregated ones even though they use the same distribution for the whole economy. As usage of Broda and Weinstein (2006) estimates of demand parameter on disaggregated level became common practice, I propose to adopt this practice to the supply side parameter. In this paper I estimate parameters of Pareto distribution on HS-4 level for more than 600 US industries under assumptions of CES and translog utility.

¹Another utility function combined with Pareto distribution might also be an answer

Besides testing distributional assumptions for the whole sample it makes sense to see whether different subsamples within an industry belong to the same distribution. Fernandes et al. (2015) show that Chaney (2008) model is unable to generate share of intensive margin consistent with the data, while the model with log-normal distribution generates better results. Di Giovanni and Levchenko (2013) on the other hand show that under Pareto assumption fixed costs do not play much role for welfare gains, while in the case of log-normal distribution they do and variable costs don't. All the papers mentioned above made an assumption that large and very productive firms draw their productivities from the same distribution as small and not very productive ones. The ratio of imports of the most and the least efficient firms can be as large as 10^{10} . It is common problem in natural sciences: Pareto distribution is well known for approximating right tails very well, while it works worse than log-normal distribution for the left tail. As a result trimming Pareto from below became a common practice. Bas et al. (2015) trim 50% of all firms and get the results different from the whole sample. Di Giovanni and Levchenko (2013) trim 90% of firms ² that represent only 7% of trade volume. Different estimates based on trimmed and not trimmed samples indicate that the left tail is not Pareto distributed. Consequently the shape parameter estimated on a trimmed sample is not informative about majority of firms. It is not a big problem if trade volumes are considered - dropped firms represent a small share of total trade. If the question of interest is extensive margin of trade, trimming is not innocuous anymore: entry and exit of firms mostly happens in the left tail we know nothing about.

Here is the dilemma: statistical test indicates that the productivities are Pareto distributed, but to get proper Pareto estimates the sample should be trimmed. This is common problem in natural sciences: Love et al. (2015) and Perline (2005) provide numerous examples. Malevergne et al. (2011) propose a solution: rather than choosing between better approximation of left or right tails they propose to use a composite distribution that is trimmed log-normal below some threshold and Pareto above it. Value of the threshold is chosen to maximize the likelihood function of this composite distribution.

2 Distinguishing Distributions

2.1 Log-Normal against Pareto

2.2 Test

Log-normal and Pareto distributions are the most popular choice to approximate right tails of distributions. It is well known that right tails of these distributions are hard to distinguish. Log-normal density can be rewritten in the following way:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\mu^2}{2\sigma^2}} x^{-1 + \frac{\mu}{\sigma^2} - \frac{\ln x}{2\sigma^2}},$$

²they follow Gabaix (1999) and choose trimming based on visual inspection. I propose more rigorous approach I called density trimming: if kernel density distribution has a hump, I drop all the observations to the left from the hump, hence allowing Pareto distribution to approximate empirical distribution with decreasing pdf

with location parameter μ and scale parameter $\sigma > 0$. Then it can be interpreted as density of Pareto distribution with varying shape parameter α :

$$\alpha(x) = \frac{1}{2\sigma^2} \ln\left(\frac{x}{e^{2\mu}}\right)$$

Then in case if σ is large, $\alpha(x)$ decreases slow and it might look like a constant. This link between log-normal and Pareto distribution leads to the debates what distribution should be used. For example Eeckhout (2004), Levy (2009) and Eeckhout (2009) disagree on what the better approximation of US city sizes is. Malevergne et al. (2011) states that the reason these authors failed to find consensus is that they tested different hypotheses and used tests that were not powerful enough. Malevergne et al. (2011) propose a the uniformly most powerful unbiased test to distinguish between log-normal and Pareto distributions. They show that after taking logarithm of variable of interest, the problem can be formulated as testing exponential against truncated normal distribution. Del Castillo and Puig (1999) show that likelihood ratio (LR) test is uniformly most powerful unbiased (UMPU) test and clipped sample coefficient of variation computed as

$$\hat{c} = \min\left\{1, \frac{s}{\bar{x}}\right\}$$

is a sufficient statistic for this test, where s is sample standard deviation computed with Bessel correction. Under hypothesis H_0 that x_1, x_2, \dots, x_n are drawn from exponential distribution. This test is possible because exponential distribution is a limiting case of truncated normal distribution: truncated normal belongs to the family of increasing failure rate (IFR) distributions. Exponential distribution corresponds to the case of constant failure rate. Now the question is how to compute critical region for \hat{c} . Del Castillo and Puig (1999) claim that \hat{c} can be approximated by 50% mixture of constant 0 and 50% of χ_1^2 distributions, but there could problems in small samples. Malevergne et al. (2011) propose to use saddle point approximation as in Gatto and Jammalamadaka (2002) or just perform Monte-Carlo simulations of \hat{c} . I apply the latter. Notice that for exponentially distributed variable x $E[x] = \sqrt{\text{Var}(x)} = \frac{1}{\lambda}$ and hence coefficient of variation does not depend on the rate parameter. As suggested by Malevergne et al. (2011) I use n is equal to sample size draws from exponential distribution with $\lambda = 1$ and number of simulations $M = 10000$ in order to compute p-value, counting share of draws when simulated values do not exceed sample coefficient of variation $pvalue = \frac{\sum_j^M (I[\hat{c} < c_j^{ML}])}{M}$.

2.3 Implementation

I do not observe productivities, but they can be backed up from sales and import shares if demand parameters are known:

$$\frac{z_i}{\underline{z}} = \left(\frac{x_i}{x(\underline{z})}\right)^{\frac{1}{\sigma-1}}$$

in CES case³, where \underline{z} is a lower bound of Pareto distribution, and $x(\underline{z})$ can be replaced by its maximum likelihood estimate x_{min} . In translog case the expression is:

$$\frac{z_i}{\underline{z}} = \left(\frac{s_i}{\gamma} + 1 \right) e^{\frac{s_i}{\gamma}}$$

where γ is a translog taste parameter. Then the data needed to construct \hat{c} and perform the test described above can be found as $\log \frac{z_i}{\underline{z}}$.

2.3.1 Data

I use PIERS data on imports. It is a dataset with imports of 50,000 largest firms to the US. Sales of each firm are provided on HS-4 level. The coverage is 1992 and 2005. In this paper I present the results for 2005 only, the results for 1992 are qualitatively similar. For demand side parameters in CES case I use the estimates of Broda and Weinstein (2006) and for translog the estimates from Feenstra and Weinstein (2010)

In CES case in order to back up productivities I need to know values of elasticity of substitution. I use Broda and Weinstein (2006) estimates of σ . The problem that they do not provide the estimates on HS-4 level, instead their estimates are on HS-10 level or in SITC Rev.3 industry classification. I find that 4-digit SITC classification is the level of disaggregation closest to HS-4 classification. I build the concordance⁴, and find elasticities Broda and Weinstein (2006) corresponding to HS-4 industries. In case when there is more than one corresponding 4-digit SITC industry, I used the median value of elasticity.

2.4 First Results

My estimates show that for CES case hypothesis H_0 that firms' distribution of productivities is Pareto distributed can be rejected for 50 HS-4 industries out of 670 (7%). For translog hypothesis H_0 can be rejected for 481 HS-4 industries out of 758 (63%). Notice that these estimates are for unbounded Pareto, while this hypothesis is rejected in the next section. Moreover, these results might be different under different trimming as suggested by Malevergne et al. (2011).

2.4.1 Pooled Test

In both CES and Translog cases under assumption that θ is the same among all industries and lower bounds can be different, the hypothesis that productivity distribution is Pareto is rejected. In other words, on a higher level of aggregation, firms' distribution of productivities look more like log-normal even though

³An interesting property: coefficient of variation does not depend on σ , it means that this test is robust to the errors in computing elasticity of substitution

⁴I use <http://unstats.un.org/unsd/trade/conversions/HS%20Correlation%20and%20Conversion%20tables.htm>

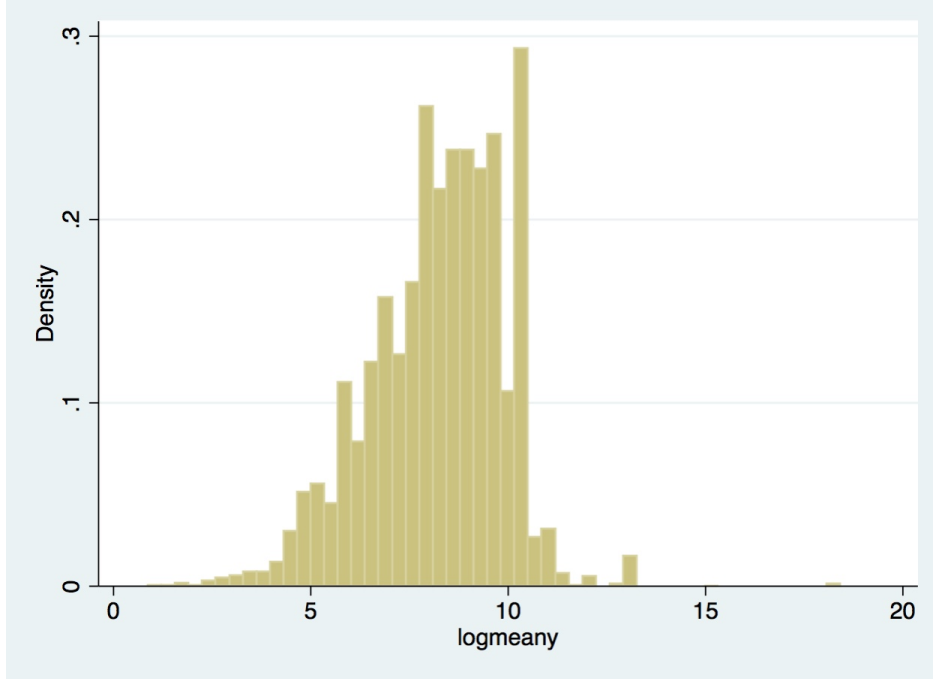


Figure 1: Distribution of Log-Linearized, Normalized Average Sales on HS-4 Level

most of these industries look more like Pareto separately. The reason is that the industries are fairly heterogeneous. Figure 1 shows the distribution of log-linearized and normalized average sales by industries.

One might think about it as a special case of the central limit theorem: a mixture of different Pareto distributions looks more like log-normal, rather than Pareto distribution.

2.5 Bounded or Unbounded Pareto

If there is unbounded Pareto distribution Z with shape parameter θ and lower bound \underline{z} , Del Castillo and Puig (1999) offer an expression for the maximum order statistics Z_{max} . They show that

$$\frac{Z_{max}}{n} \rightarrow^d Y \text{ as } n \rightarrow \infty,$$

where

$$P(Y \leq t) = e^{-Ct^{-\theta}}$$

and $C = \underline{z}^\theta$. Remember that if $Z \sim \text{Pareto}(\underline{z}, \theta)$, then $Z/\underline{z} \sim \text{Pareto}(1, \theta)$ and in this case expression from (1) simplifies:

$$\frac{Z_{max}}{\underline{z}n} \rightarrow^d Y \text{ as } n \rightarrow \infty,$$

where

$$P(Y \leq t) = e^{-t^{-\theta}}.$$

With known distribution of maximum order statistics, it is easy to write down p-value for the test on existence of the upper bound. This p-value will simply be a probability to observe z_{max} under hypothesis H_0 that Pareto distribution is not truncated if there were n draws.

$$pvalue = e^{-nZ_{max}^{-\theta}}$$

They suggest to replace θ with $\hat{\theta}$ from:

$$\hat{\theta}_{ML} = \frac{n}{\sum [\ln z_i - \ln \underline{z}]} = \frac{n}{\sum \ln \frac{z_i}{\underline{z}}},$$

z are not observed, but by definition of convergence in distribution, productivity can be expressed as inverse function of x sales or market shares.

$$\frac{Z_{max}}{\underline{z}n} \rightarrow^d Y \Rightarrow Pr \left(\frac{Z_{max}}{\underline{z}n} < z \right) = P(Y \leq x) = e^{-nx^{-\theta}}$$

As $n \rightarrow \infty$.

2.6 CES Case

In CES case normalized sales are a function of normalized productivities⁵:

$$\frac{x(z)}{x(\underline{z})} = \left(\frac{z}{\underline{z}} \right)^{\sigma-1}$$

This increasing function can be inverted as well:

$$\frac{z}{\underline{z}} = \left(\frac{x(z)}{x(\underline{z})} \right)^{\frac{1}{\sigma-1}}$$

And now

$$Pr \left(x \left(\frac{Z_{max}}{\underline{z}} \right) < \frac{z}{\underline{z}} \right) = Pr \left(\frac{1}{n} \frac{Z_{max}}{\underline{z}} < \frac{1}{n} \left(\frac{x_{max}}{x(\underline{z})} \right) \right) = Pr \left(Y < \left(\frac{x_{max}}{x(\underline{z})} \right)^{\frac{1}{\sigma-1}} \right)$$

And then p-value will be:

$$pvalue = e^{-n \left(\frac{x_{max}}{x(\underline{z})} \right)^{-\frac{\theta}{\sigma-1}}}$$

Where θ and $x(\underline{z})$ are replaced with their ML-estimates:

$$\hat{\theta} = \frac{n}{\sum_i \ln \frac{x_i}{x_{min}}} (\sigma - 1)$$

$$x(\underline{z}) = \min_i \{x_i\}$$

⁵I provide more detailed derivations of sales and market shares in CES and translog cases in part 3 and appendix of this paper

2.7 Translog Case

Good news is that market shares as a function of normalized productivities are observed:

$$s_i = \left[W \left(\frac{z_i}{z} e \right) - 1 \right] \gamma$$

This function is increasing so it can be inverted:

$$\frac{z_i}{z} = \left(\frac{s_i}{\gamma} + 1 \right) e^{\frac{s_i}{\gamma}}$$

And now

$$Pr \left(s \left(\frac{Z_{max}}{z} \right) < \frac{z}{z} \right) = Pr \left(\frac{1}{n} \frac{Z_{max}}{z} < \frac{1}{n} \left(\frac{s_{max}}{\gamma} + 1 \right) e^{\frac{s_{max}}{\gamma}} \right) = Pr \left(Y < \left(\frac{s_{max}}{\gamma} + 1 \right) e^{\frac{s_{max}}{\gamma}} \right)$$

And then p-value will be:

$$pvalue = e^{-n \left(\left(\frac{s_{max}}{\gamma} + 1 \right) e^{\frac{s_{max}}{\gamma}} \right)^{-\theta}}$$

Where θ is replaced with $\hat{\theta}$ from:

$$\hat{\theta} = \frac{n}{\sum_i \log \left(\frac{s_i}{\gamma} + 1 \right) + \frac{\sum_i s_i}{\gamma}}$$

2.8 Estimation Results

2.8.1 CES Case

Under an assumption of the same θ within HS-4 industry, but different lower bounds for different importing countries I find that out of 633 HS-4 industries for 89 industries (14%) the hypothesis that the distribution is unbounded Pareto cannot be rejected on 1% significance level. For the rest 544 industries p-value is smaller than 1% and the hypothesis that there is no upper bound is not supported by the data. As I showed in my estimation part, baseline estimates of θ are fairly small. Small θ imply very fat tails and hence for given data it becomes easier to reject the hypothesis that there is no upper bound. I compute p-values again for the robustness check with higher estimates of θ . For now I am using 50% quantile trimming and $\sigma > 2$ specification. In this case for 16 industries out of 393 H_0 is rejected, and for 376 (95%) it is not.

2.8.2 Translog Case

Only for 88 HS-4 industries out of 758 hypothesis that H_0 is unbounded Pareto cannot be rejected.

3 Estimating Pareto

In this part of the paper I estimate the shape parameter of Pareto distribution in a model with heterogeneous firms using firm level data on imports. Here I consider the case of CES utility for bounded and unbounded Pareto distribution of firms' productivities. The derivations and preliminary results on translog are provided in the appendix.

3.1 Baseline Case

Cdf of unbounded Pareto distribution with parameters (θ, \underline{z}) is:

$$1 - F_Z(z) = P(Z > z) = \left(\frac{\underline{z}}{z}\right)^\theta \quad (1)$$

Distribution of $\frac{\underline{z}}{z}$ will then be: $Pareto(\theta, 1)$. The problem is that productivities z are not observed. From Melitz (2003)

$$\frac{x(z_1)}{x(z_2)} = \left(\frac{z_1}{z_2}\right)^{\sigma-1} \quad (2)$$

where $x(z_1)$ and $x(z_2)$ are sales of firms with productivities z_1 and z_2 and σ is elasticity of substitution from CES utility function. Notice that in Melitz model due to presence of fixed costs sales of the least productive firm on the market is larger than zero $x(z^*) > 0$. Then assuming $z^* = \underline{z} \frac{x(z)}{x(\underline{z})} = \left(\frac{\underline{z}}{z}\right)^{\sigma-1}$. Inverting this expression obtain:

$$z = \underline{z} \left(\frac{x(z)}{x(\underline{z})}\right)^{\frac{1}{\sigma-1}} \quad (3)$$

and we treat $x(\underline{z})$ as a constant⁶. Notice that \underline{z} is a parameter of distribution to be estimated.

$$Pr(x(Z) > z) = Pr\left(Z > \underline{z} \left(\frac{x}{x(\underline{z})}\right)^{\frac{1}{\sigma-1}}\right)$$

So cdf of x is then $F_X(x) = 1 - \frac{x}{x(\underline{z})}^{-\frac{\theta}{\sigma-1}}$ and pdf is $\frac{\theta}{\sigma-1} x^{-\frac{\theta+(\sigma-1)}{\sigma-1}} x(\underline{z})^{\frac{\theta}{\sigma-1}}$

Log-likelihood function is then $l(\theta) = n \ln \theta - \ln(\sigma - 1) + n \frac{\theta}{\sigma-1} \ln x(\underline{z}) - \frac{\theta+(\sigma-1)}{\sigma-1} \sum_{i=1}^n \ln x_i$. In order to compute normalized shares we need to estimate $x(\underline{z})$. Log-likelihood function increases in $x(\underline{z})$ for $\forall \theta$ and then it's ML-estimate is

$$x(\underline{z})_{ML} = x_{min} \equiv \min \{x_i\}$$

⁶There is one to one mapping between normalized productivity and normalized sales, hence productivity of each firm can be backed up from it's sales. From now on I will treat productivities as if they are observed.

Notice that now when we know ML-estimate of $x(\underline{z})$, we can compute normalized sales and work with them:

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^n \ln \frac{x_i}{x_{min}}}{\sigma - 1} = 0$$

Let's introduce normalized sales: $y_i \equiv \frac{x_i}{x_{min}}$

$$\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^n \ln y_i} (\sigma - 1) \quad (4)$$

3.2 Endogenous Lower Bound

The derivations above were performed for $Pareto(\theta, \underline{z})$. The problem is that in Melitz (2003) cutoff productivity of exporting firm will be:

$$z_x^* = z^* \tau \left(\frac{f_x}{f} \right)^{\frac{1}{\sigma-1}}$$

and hence will depend on fixed costs of production and exporting and on trade costs. I do not observe firms' home sales, but observe their imports. Notice that distribution of exporters' productivities is also Pareto, but with lower bound z_x^* . From here and further I will estimate parameters of the distribution of exporters' productivities rather than distribution of domestic firms' productivities. In other words in (1) I will imply that $\underline{z} = z_x^*$.

3.3 Case of Several Subsamples

Now let's consider the case of J subsamples with different values of σ_j and \underline{z}_j , but same θ . In this case derivation of density will be similar and ML-estimates of lower bound of distribution of sales will be $x_j(\underline{z}_j)_{ML} = \min_i \{x_{ij}\}$. Log-likelihood function for normalized sales will then be:

$$l(\theta) = N \ln \theta - \sum_{j=1}^J n_j \ln(\sigma_j - 1) - \sum_{j=1}^J \left(\frac{\theta}{\sigma_j - 1} \sum_i^{n_j} \ln y_{ij} \right)$$

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{N}{\theta} - \sum_{j=1}^J \left(\frac{1}{\sigma_j - 1} \sum_i^{n_j} \ln y_{ij} \right) = 0$$

And generalized estimator (4) is:

$$\hat{\theta}_{ML} = \frac{N}{\sum_{j=1}^J \left(\frac{1}{\sigma_j - 1} \sum_i^{n_j} \ln y_{ij} \right)} \quad (5)$$

3.4 Bounded Pareto

Helpman et al. (2008) and Feenstra (2016) use bounded Pareto distribution that allows for zero trade flows and decomposition of gains from trade from different sources. Bounded Pareto might be a good alternative to lognormal distribution: trade elasticity is not constant anymore and extensive margin is different from zero. Upper bounds of Pareto distribution can vary by country of origin thus reflecting technological differences between countries. In bounded case ML estimates of upper and lower bounds:

$$\tilde{z}_j \equiv \underline{z} \hat{z}_{ML} = \left(\frac{1}{y_{max_j}} \right)^{\frac{1}{\sigma-1}}$$

And $\hat{\theta}$ is a solution to the following equation:

$$\frac{N}{\hat{\theta}} + \sum_{j=1}^J \frac{n_j (\tilde{z}_j)^{\hat{\theta}} \ln(\tilde{z}_j)}{1 - (\tilde{z}_j)^{\hat{\theta}}} = \sum_{j=1}^J \sum_{i=1}^{n_j} \left(\frac{1}{\sigma - 1} \sum_i \ln y_{ij} \right)$$

This expression does not have closed form solution, but using the results of Aban et al. (2006) I show that the probability that this equation has a solution converges to 1 as number of observations converges to ∞ , that if the solution exists, then it is unique and finally that $\hat{\theta}$ is a consistent estimator of θ .

3.4.1 Properties of Bounded Pareto Estimator

Notice that the only difference between bounded and unbounded estimators is the second term on the left hand side that corrects for different domain of bounded Pareto cdf. As $\tilde{z}_j < 1$ this term is smaller than 0 and then bounded Pareto estimate of θ is always smaller. Notice also that the correction term converges to 0 as upper bound converges to ∞ , $\tilde{z}_j \rightarrow 0$. Intuitively it means that estimating true unbounded distribution by bounded estimator on a large enough sample should not be problematic: the estimate of upper bound will be large and will push the correction term towards 0. On Figure 2 I present the results of Monte Carlo simulations. I had sample size of 1,000 and performed 10,000 simulations. Lower bound was equal to 1, true value of the shape parameter varied from 0.1 to 10. Horizontal axis and 45 degrees gray line represent the true value of the shape parameter. Yellow line is the average of 10000 estimates; red and blue line represent 95% confidence interval. The left panel corresponds to the unbounded estimator and the right panel corresponds to the bounded one.

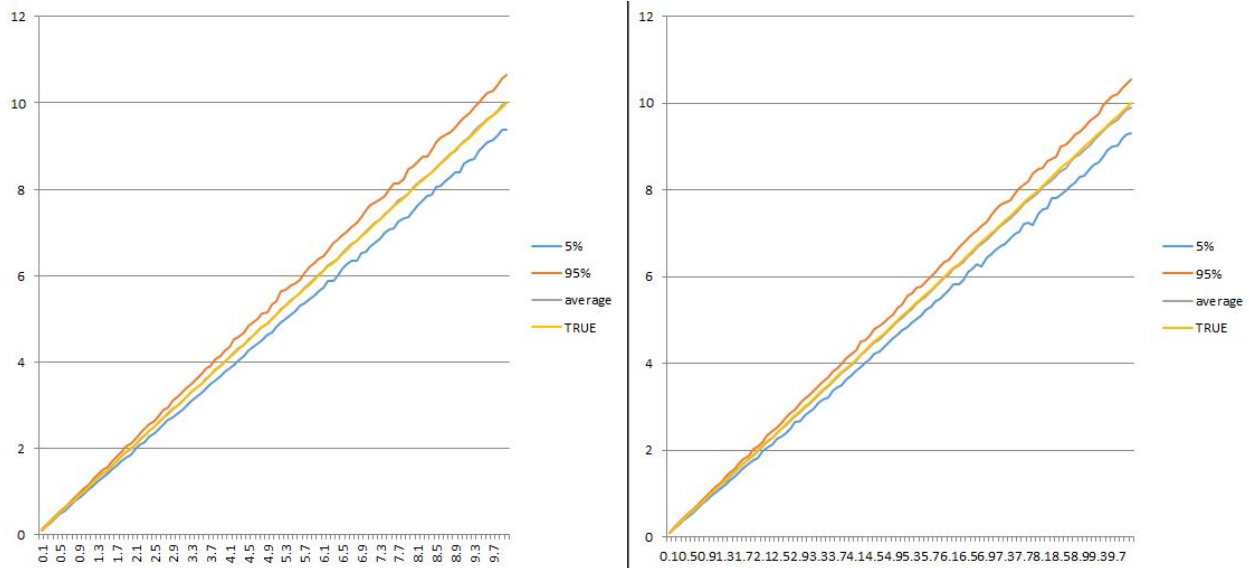


Figure 2: Unbounded and Bounded Estimates of Unbounded Pareto Distribution

Both estimators consistently estimate true parameters, moreover there is no efficiency gains for unbounded estimator: average ratio of confidence intervals is smaller than 1%. The situation is very different when true distribution is bounded Pareto. I present the results of similar Monte Carlo procedure with upper bound equal to 2 on Figure 3: Just as in the previous case bounded Pareto estimator does a good job estimating true parameter, while unbounded estimator generates significant bias that varies from 1% to 3000%. The reason why two estimators converge as true value of θ increases is that lower θ represent fatter right tail and hence increases relative importance of upper bound. As $\theta \rightarrow \infty$ for given upper bound the bounded distribution will converge to unbounded one. For fixed θ increase in the upper bound of true distribution will lead to the same outcome: bias will decrease. In other words, size of the bias depends on how large shape parameter is relative to the upper bound. My results in section 2 indicate that the upper bound is not negligible. Under assumption of Pareto distribution I recommend to use bounded Pareto estimator as it is robust to misspecification.

3.5 Application

Expression (5) can be applied to the case of import from several countries. Here we assume that σ and θ are the same for all firms i and all countries of origin j . What can be different is z_j . Given that z_j depends on fixed costs and trade costs, I allow them to vary from country to country. In particular estimates of z_j absorb trade costs.

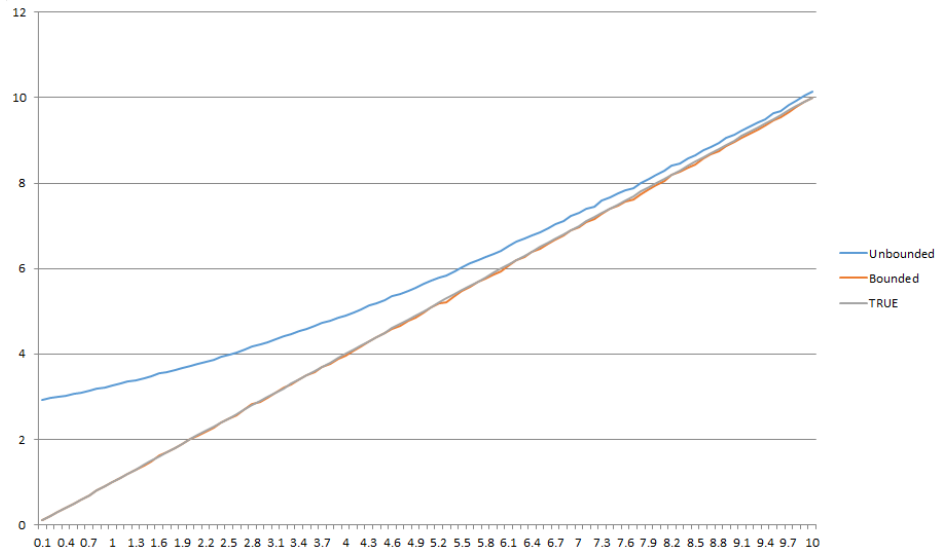


Figure 3: Unbounded and Bounded Estimates of Bounded Pareto Distribution

3.5.1 Low Values of Elasticities

From expression (3) one can see that if $\sigma < 2$, difference in sales of the firms magnifies the productivity difference. For example if $\sigma = 1.1$, and $\frac{x(z_1)}{x(z_2)} = 2$, $\frac{z_1}{z_2} = 2^{10}$. This will generate productivity distribution with very heavy tails and lead to very low estimates of θ . This is an artifact of CES utility that behaves oddly at low values of σ . As an alternative specification I report the results of the estimation for $\sigma \not\leq 2$

3.6 Estimation

3.6.1 Baseline Results

In this section I present the baseline results. Here I concentrate on the results for CES utility as the estimates can be compared with earlier work. All the derivations for the translog utility are in the appendix⁷.

For convenience I report median value of θ for each of 9 large sectors⁸. σ_k are from Broda and Weinstein (2006) on HS-4 level. I drop all industries with less than 20 firms and all industry-country pairs with less than 6 firms.

⁷Estimates of θ on a disaggregated level for translog case, and for both bounded and unbounded Pareto are available upon request

⁸Results for all 633 HS-4 sectors are available upon request

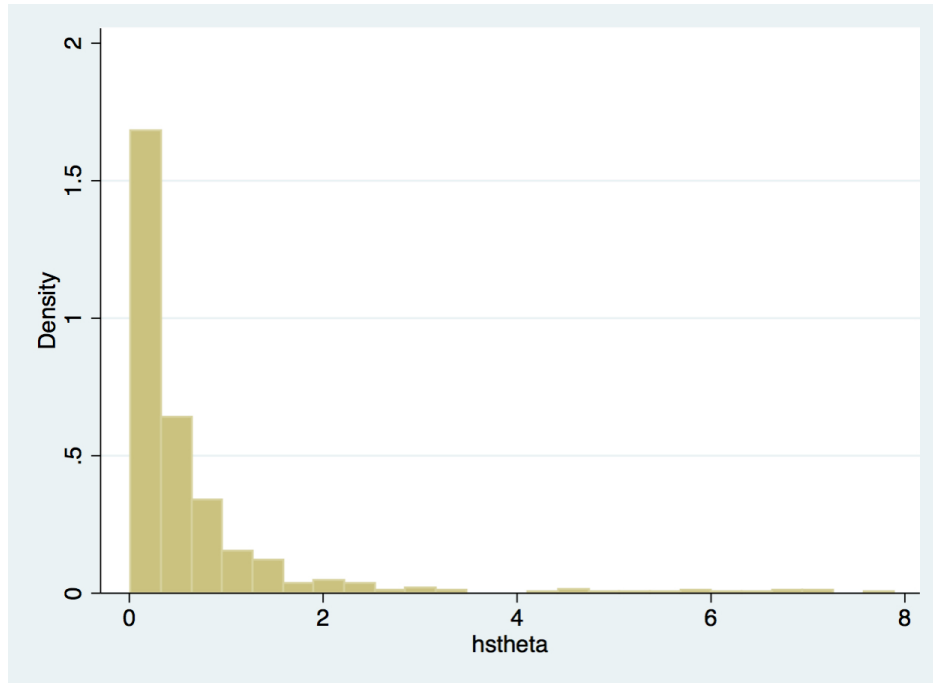


Figure 4: Baseline Estimates of Pareto Shape Parameter on HS-4 Level. $\theta > 8$ are Omitted.

Description	Baseline
Agriculture and Fishing	0.453
Food & Beverages	0.582
Textiles and Wearing Apparel	0.537
Wood and Paper	0.340
Petroleum, Chemical and Mineral Products	0.320
Metal Products	0.375
Electrical and Machinery	0.213
Transport Equipment	0.708
Other Manufacturing and Recycling	0.099

The distribution of all the estimates is presented on Figure 4.

Notice that one of the assumptions of Chaney (2008) is $\theta > \sigma - 1$. For estimates above this assumption holds for just few of HS-4 industries. To get these estimates I did not rely on $\theta > \sigma - 1$, so one possible explanation is that this assumption is not supported by the data. Three other potential explanations of these low estimates of θ : low values of σ , multi-product firms and absence of trimming. Finally assumption $\theta > \sigma - 1$ is needed so that firms' average sales are finite. In case of bounded Pareto it is not a problem and

thus the bounded Pareto estimates of θ can be larger or smaller than $\sigma - 1$.

3.6.2 Low Values of Elasticities

From expression (3) one can see that if $\sigma < 2$, difference in sales of the firms magnifies the productivity difference. For example if $\sigma = 1.1$, and $\frac{x(z_1)}{x(z_2)} = 2$, $\frac{z_1}{z_2} = 2^{10}$. This will generate productivity distribution with very heavy tails and lead to very low estimates of θ . This is an artifact of CES utility that behaves oddly at low values of σ . As an alternative specification I report the results of estimation for $\sigma \not\leq 2$

3.6.3 Multi-Product Firms

Bernard et al. (2006) show that more productive firms have larger product scope. In the baseline result we do not allow for multi-product firms and hence I miss extensive margin dimension. The data we have is on HS-4 level and hence if a firm had more than one product it will be aggregated. Given that these data are missing, I can make some additional assumptions on how productivity of a firm and its scope are linked and then take this into account. Here I use approach of Ma (2008) and assume that a consumer has CES preferences over different varieties of the same firm have elasticity of substitution η and are nested into CES preferences with the elasticity of substitution σ :

$$U = \left(\int^N q(i)^{(\sigma-1)/\sigma} di \right)^{\sigma/(\sigma-1)}, \sigma > 1.$$

Each firm i produces a continuum n_i varieties aggregated in $q(i)$:

$$q(i) = \left(\int^{n_i} q_i(j)^{(\eta-1)/\eta} dj \right)^{\eta/(\eta-1)}$$

The ratio of sales then will be⁹:

$$\frac{x(z)}{x(z^*)} = \left(\frac{z}{z^*} \right)^{\frac{(\eta-1)(\sigma-1)}{\eta-\sigma}}$$

As a result ML-estimate (5) becomes:

⁹In a single product model it is:

$$\frac{r(z)}{r(z^*)} = \left(\frac{z}{z^*} \right)^{(\sigma-1)}$$

$$\hat{\theta}_{ML} = \frac{N}{\sum_{j=1}^J \left(\frac{\eta_j - \sigma_j}{(\sigma_j - 1)(\eta_j - 1)} \sum_i^{n_j} \ln y_{ij} \right)}$$

The problem with this approach is that it requires some knowledge on within-firm elasticity of substitution η . It is impossible to identify η using maximum likelihood approach and PIERS data.

Hottman et al. (2014) find that within-firm elasticities are about 50% smaller than between-firm elasticities, so $\eta = 2\sigma$ and this ratio is very robust and does not depend on the quantile of σ in consideration¹⁰.

Under assumption that $\eta = 2\sigma$ the estimate becomes:

$$\hat{\theta}_{ML} = \frac{N}{\sum_{j=1}^J \left(\frac{\sigma_j}{(\sigma_j - 1)(2\sigma_j - 1)} \sum_i^{n_j} \ln y_{ij} \right)}$$

Given that $\frac{\sigma}{(\sigma-1)(2\sigma-1)} < \frac{1}{\sigma-1}$, final estimates of θ will be larger:

3.7 Trimming

Trimming smallest 50% of firms (less than 5% of total sales) significantly increases the estimates of θ . Trimming more than 50% of firms does not affect the estimates much. It supports the hypothesis that right tail of firms' productivity distribution follows Pareto distribution, while the whole distribution does not. This finding does not contradict to what I found in section 2: the test indicated, that the distribution is Pareto rather than log-normal, but it did not test Pareto against other alternatives. Trimming Pareto from below is common practice in estimating shape parameter, however there is no standard procedure of how to choose a threshold level. Most authors just rely on visual inspection. As I showed in section 2, it is important to deal with disaggregated data. With more than 600 sectors relying on visual inspection becomes problematic, so there is need in some standardized procedure. I performed a set of standard procedures: dropping n smallest firms, dropping firms below quantile q , dropping firms that represent share of total imports less than s . Finally I propose and perform two new procedures in this paper. The first one I called density trimming and it is described in the next section. The second one is described in part 4 of this paper.

3.7.1 Density Trimming

Estimation results are sensitive to trimming, besides some industries might require more trimming. I propose a new procedure I called density trimming. If distribution has a hump, it is not necessarily a

¹⁰Their results, however, are based on Nielsen data (very disaggregated), besides they consider mostly food products. So usage of their findings might be questionable for broader set of industries.

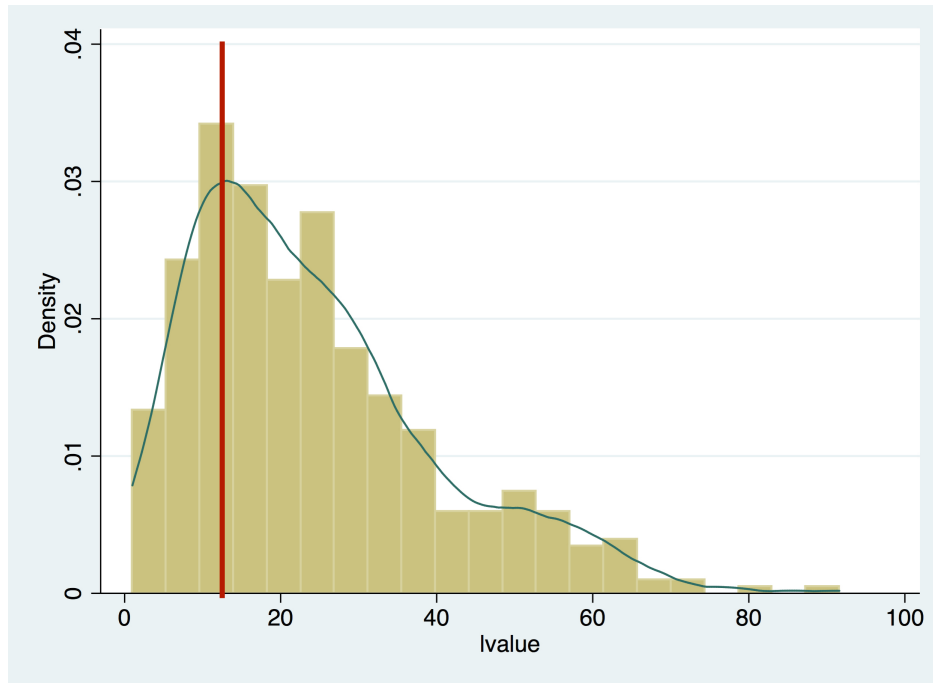


Figure 5: Density Trimming Procedure. HS-5407 Industry

problem: small left tail represents small share of trade, but can bias estimates of the shape parameter a lot. I propose to first perform kernel density estimation of firms' productivity distribution. I will keep then only the right tail (e.g. drop all observations with values on an increasing part of kernel density curve). The graph represents distribution of Chinese firms' sales in HS-5407 (Woven fabrics of synthetic filament yarn). I drop observations to the left from the red line.

3.8 Discussion

These are the results with the alternative specifications. Here I provide them separately, but the assumptions can be combined in any order.

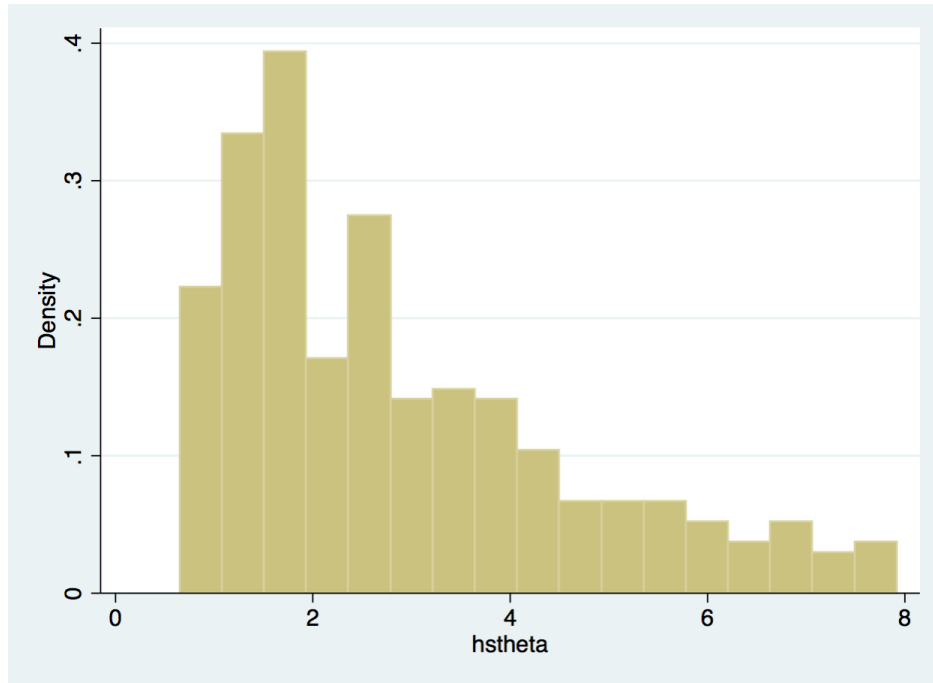


Figure 6: Main Estimates of Pareto Shape Parameter on HS-4 Level. $\sigma > 2$, 50% of Firms are Trimmed. $\theta > 8$ are Omitted.

Table 2. Shape Parameter Under CES Utility. Alternative Specifications					
Description	Baseline	$\sigma \geq 2$, 50% Drop	$\sigma \geq 2$	Multi-product	50% Drop
Agriculture and Fishing	0.453	6.530	0.970	1.338	2.951
Food & Beverages	0.582	3.40	0.775	1.351	2.524
Textiles and Wearing Apparel	0.537	3.656	0.629	1.215	3.160
Wood and Paper	0.340	1.872	0.443	0.684	1.568
Petroleum, Chemical and Mineral Products	0.320	2.684	0.553	0.694	1.613
Metal Products	0.375	3.905	0.799	0.858	1.801
Electrical and Machinery	0.213	1.781	0.392	0.389	0.985
Transport Equipment	0.708	5.574	1.093	2.170	3.479
Other Manufacturing and Recycling	0.099	2.359	0.466	0.148	0.408

I use the case with $\sigma \geq 2$ and 50% trimming as the main specification. In this case for 67% of HS-4 industries $\theta > \sigma - 1$. This is the histogram of the estimates with $\theta > 8$ dropped for convenience:

These estimates are comparable with the results other researchers obtained when they were estimating shape parameter of Pareto distribution on micro-level data:

Paper		$\hat{\theta}$
EKK (2011)		1.75
Fernandez et al. (2015)		2.3
Head et al. (2014)	Baseline	1.3
	Trim 50%	2.16

3.9 Extensions

My method estimates lower bound of exporters' productivities for each sector and each country. Remember that these lower bounds are endogenous and depend on trade and fixed costs: $z_x^* = z^* \tau \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}}$. Then the ratio of these lower bounds for different countries i and j , but for the same industry k will be:

$$\frac{z_{xik}^*}{z_{xjk}^*} = \frac{z_{ik}^*}{z_{jk}^*} \frac{\tau_{ik}}{\tau_{jk}} \left(\frac{f_{xik}f_{jk}}{f_{xjk}f_{ik}} \right)^{\frac{1}{\sigma-1}} \quad (7)$$

Left hand side is estimated above.

Under the assumption of the same technology at home: $z_{xik}^* = z_{xjk}^*$ expression (7) becomes:

$$\frac{z_{xik}^*}{z_{xjk}^*} = \frac{\tau_{ik}}{\tau_{jk}} \left(\frac{f_{xik}f_{jk}}{f_{xjk}f_{ik}} \right)^{\frac{1}{\sigma-1}}$$

where left-hand side is known and right-hand side represents relative access to the foreign market. Under stronger assumption $\frac{f_{xik}f_{jk}}{f_{xjk}f_{ik}} = 1$, that can be interpreted as that costs of exporting relative to the costs of home production are similar across countries, expression (7) allows to identify relative trade costs.

4 Composite Distribution

Statistical test by Malevergne et al. (2011) rejects the hypothesis that the data is log-normally distributed. On the other hand, trimming smallest firms affects estimates of shape parameter of Pareto distribution, that is impossible under assumption of true Pareto distribution. After dropping 50% of firms, that represents less than 3% of total sales estimates of shape parameter do not depend on trimming anymore. This evidence suggests that the left side of firms' distribution can be described as log-normal distribution while the right tail as a Pareto distribution. In the context of Melitz (2003) it will mean that the impact of trade liberalization on intensive and extensive margins will follow different rules: inefficient firms that will leave the market will be lognormally distributed while the most productive firms that will enjoy the reallocated resources will have fat-tailed Pareto distribution. Malevergne et al. (2011) suggests to use this approach to

approximate the distribution of US cities and Ioannides and Skouras (2013) implement it. They accept the hypothesis that top 1000 US cities are distributed according to Pareto distribution. Then they claim that small US cities follow log-normal distribution. The trick is to determine the trimming level. Malevergne et al. (2011) suggest to choose threshold parameter u separating the log-normal from Pareto distribution by maximizing likelihood function¹¹. Below I follow the derivations of Ioannides and Skouras (2013)

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad u > x > 0,$$

where (μ, σ) are the mean and standard deviation of $\ln x$ and u is the threshold parameter. The kernel of Pareto density is:

$$g(x; \theta, u) = \frac{1}{x^{\theta+1}}, \quad x \geq u, \theta > 0,$$

where θ is the shape parameter. The density function of composite distribution is then:

$$h(x; \mu, \sigma, u, \theta) = \left\{ \begin{array}{l} b(u, \mu, \sigma, \theta) f(x; \mu, \sigma), \quad u > x > 0 \\ a(u, \mu, \sigma, \theta) b(u, \mu, \sigma, \theta) g(x; \theta, u), \quad x \geq u \end{array} \right\},$$

where $a(u, \mu, \sigma, \theta)$ is a normalizing constant that scales $g(x; \theta, u)$ so that $h(x; \mu, \sigma, u, \theta)$ is continuous at the switching point u :

$$a(u, \mu, \sigma, \theta) = \frac{f(u; \mu, \sigma)}{g(u; \theta, u)} = f(u; \mu, \sigma, u) u^{1+\theta}$$

and $b(u, \mu, \sigma, \theta)$ is a normalizing constant that ensures that density $h(x; \mu, \sigma, u, \theta)$ integrates to 1. It can be expressed as:

$$b(u, \mu, \sigma, \theta) \equiv \frac{1}{\Phi(u; \mu, \sigma) + f(u; \mu, \sigma, u) \frac{u}{\theta}};$$

where $\Phi(u; \mu, \sigma)$ is cdf of log-normal distribution with parameters (μ, σ) .

$h(x; \mu, \sigma, u, \theta)$ satisfies the regularity conditions that are necessary for maximum likelihood estimation and inference. Notice that $h(x; \mu, \sigma, u, \theta)$ nests both log-normal ($u \leq 1$) and Pareto distribution ($u \geq x_{max}$). It means that the composite distribution approximates the data at least as well as Pareto and log-normal

Maximizing of $h(x; \mu, \sigma, u, \theta)$ is a two step procedure: first for given u find (μ, σ, θ) . Then choose u such that maximizes value of the likelihood function. The first step is tricky here: normalizing constants are functions of the parameters of the distributions and should be taken into account.

¹¹This is another procedure that allows to trim data. I, however, do not recommend to use this approach without making assumptions on the distribution of left tail: in this case the procedure will keep just few observations that perfectly fit data

4.1 Composite Distribution with Bounded Pareto

The derivations above can be easily adapted for the case of bounded Pareto. The new kernel of density is:

$$g(x; \theta, u, v) = \frac{1}{x^{\theta+1}}, v \geq x \geq u, \theta > 0,$$

where v is the upper bound of bounded Pareto distribution. $a(u, \mu, \sigma, \theta)$ remains the same, while normalizing constant $b(u, \mu, \sigma, \theta, v)$ is now equal to:

$$b(u, \mu, \sigma, \theta, v) \equiv \frac{1}{\Phi(u; \mu, \sigma) + f(u; \mu, \sigma, u) \frac{u}{\theta} \left(1 - \left(\frac{u}{v}\right)^\theta\right)}.$$

5 Conclusion

In their famous methodological dispute Milton Friedman and Paul Samuelson had opposite views on how assumptions in theoretical models should be made. Friedman insisted that the main criterion of quality of theory should be its predictive power, not how realistic assumptions are. Samuelson criticized this approach (he called it the F-twist) and claimed that theory should describe real world and rely on assumptions that make sense.

Both instrumentalism and descriptivism have caveats. Changing assumptions to improve performance of the model, there is a risk to fix what is not broken and change correct assumption for the wrong one; in this case mistakes can cancel out and lead to better predictive power of the model. The problem is that the model with two wrong assumptions is even less reliable than a model with just one.

I believe it is important to know how realistic the assumptions researchers use are. I provide a set of tools that allow to choose between different distributional assumptions, that is quite debatable question in international trade. Moreover I propose a new alternative: a distribution that combines strengths of two most popular distributions. Finally I follow extensive literature that proves that welfare and policy analysis in international trade should be performed on disaggregated level. Up to now there was no estimates of supply side parameters on highly disaggregated level. I propose a new way to estimate shape parameters of Pareto distribution that does not depend on measurement error of trade costs and provide the estimates under a number of alternative specifications and trimming procedures. Finally I introduce a composite distribution that combines strengths of two most popular distributions in the literature: Pareto and log-normal.

References

- 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, 270–277.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2010): "Gains from trade under monopolistic competition: A simple example with translog expenditure functions and pareto distributions of firm-level productivity," *Manuscript, available at <http://economics.mit.edu/files/5876>*.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.
- BAS, M., T. MAYER, AND M. THOENIG (2015): "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity." .
- BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2006): "Multi-product firms and trade liberalization," Tech. rep., National Bureau of Economic Research.
- BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains From Variety," *The Quarterly Journal of Economics*, 121, 541–585.
- CHANEY, T. (2008): "Distorted gravity: the intensive and extensive margins of international trade," *The American Economic Review*, 98, 1707–1721.
- DEL CASTILLO, J. AND P. PUIG (1999): "The best test of exponentiality against singly truncated normal alternatives," *Journal of the American Statistical Association*, 94, 529–532.
- DI GIOVANNI, J. AND A. A. LEVCHENKO (2013): "Firm entry, trade, and welfare in Zipf's world," *Journal of International Economics*, 89, 283–296.
- 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, 42–52.
- EECKHOUT, J. (2004): "Gibrat's law for (all) cities," *American Economic Review*, 1429–1451.
- (2009): "Gibrat's law for (all) cities: Reply," *The American Economic Review*, 99, 1676–1683.
- FEENSTRA, R. C. (2014): "Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity," Tech. rep., National Bureau of Economic Research.
- (2016): "Gains from Trade Under Monopolistic Competition," *Pacific Economic Review*, 21, 35–44.
- FEENSTRA, R. C. AND D. E. WEINSTEIN (2010): *Globalization, markups, and the US price level*, w15749.
- FERNANDES, A. M., P. J. KLENOW, S. MELESHCHUK, M. D. PIEROLA, AND A. RODRIGUEZ-CLARE (2015): "The Intensive Margin in Trade: Moving Beyond Pareto," .
- GABAIX, X. (1999): "Zipf's law for cities: an explanation," *Quarterly journal of Economics*, 739–767.
- GATTO, R. AND S. R. JAMMALAMADAKA (2002): "A saddlepoint approximation for testing exponentiality against some increasing failure rate alternatives," *Statistics & probability letters*, 58, 71–81.
- HEAD, K., T. MAYER, AND M. THOENIG (2014): "Welfare and trade without pareto," .
- HELPMAN, E., M. J. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating trade flows: trading partners and trading volumes," *Quarterly Journal of Economics*, 123, 441–487.

- HOTTMAN, C., S. J. REDDING, AND D. E. WEINSTEIN (2014): "Quantifying the Sources of Firm Heterogeneity," Tech. rep., National Bureau of Economic Research.
- IOANNIDES, Y. AND S. SKOURAS (2013): "US city size distribution: Robustly Pareto, but only in the tail," *Journal of Urban Economics*, 73, 18–29.
- LEVCHENKO, A. A. AND J. ZHANG (2014): "Ricardian productivity differences and the gains from trade," *European Economic Review*, 65, 45–65.
- LEVY, M. (2009): "Gibrat's law for (all) cities: Comment," *The American Economic Review*, 99, 1672–1675.
- LOVE, J. J., E. J. RIGLER, A. PULKKINEN, AND P. RILEY (2015): "On the lognormality of historical magnetic storm intensity statistics: Implications for extreme-event probabilities," *Geophysical Research Letters*, 42, 6544–6553.
- MA, H. (2008): "Product Rationalization and Trade Liberalization," *Unpublished manuscript*.
- MALEVERGNE, Y., V. PISARENKO, AND D. SORNETTE (2011): "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities," *Physical Review E*, 83, 036111.
- MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.
- MRAZOVA, M., J. P. NEARY, AND M. PARENTI (2015): "Technology, Demand and the Size Distribution of Firms," Tech. rep., Working Paper, University of Oxford.
- PERLINE, R. (2005): "Strong, weak and false inverse power laws," *Statistical Science*, 68–88.

Appendix

Nested Translog Utility

In this part I show that a consumer that has translog preferences for each of K groups of goods and spends constant share of his income on each group can be represented as translog preferences over all available goods.

Here I use a translog expenditure function over N goods from Feenstra and Weinstein (2015). In case of K groups of goods his expenditure function will be:

$$\ln e = \sum_{k=1}^K \beta_k \left(\alpha_0 + \sum_{i=1}^{n_k} \alpha_{ik} \ln p_{ik} + \frac{1}{2} \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} \gamma_{ijk} \ln p_{ik} \ln p_{jk} \right)$$

Where p_i is a price of good i , and γ_{ij} are taste parameters. I consider symmetric case: $\gamma_{iik} = -\gamma \left(1 - \frac{1}{n_k}\right) \frac{N}{n_k} < 0$, and $\gamma_{ijk} = \frac{\gamma}{N} > 0$ for $i \neq j$.

Where β_k is the expenditure share on group of goods k and $\sum_{k=1}^K \beta_k = 1$. Then (2) can be re-expressed as:

$$\ln e = \alpha_0 + \sum_{k=1}^K \sum_{i=1}^{n_k} \beta_k \alpha_{ik} \ln p_{ik} + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} \beta_k \gamma_{ijk} \ln p_{ik} \ln p_{jk}$$

Using symmetry:

$$\ln e = \alpha_0 + \sum_{k=1}^K \sum_{i=1}^{n_k} \beta_k \alpha_{ik} \ln p_{ik} - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \beta_k \gamma \left(1 - \frac{1}{n_k}\right) \frac{N}{n_k} \ln p_{ik} \ln p_{ik} + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j \neq i}^{n_k} \beta_k \frac{\gamma}{N} \ln p_{ik} \ln p_{jk}$$

Now assume that number of goods from each group is proportional to its share of consumption, or $\gamma_k = \frac{n_k}{N}$ for $\forall k$. Then we obtain:

$$\ln e = \alpha_0 + \sum_{k=1}^K \sum_{i=1}^{n_k} \beta_k \alpha_{ik} \ln p_{ik} - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \gamma \ln p_{ik} \ln p_{ik} + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} \frac{\gamma}{N} \ln p_{ik} \ln p_{jk}$$

Then it is translog expenditure function on aggregated level with parameters $\alpha_i = \beta_k \alpha_{ik}$, $\gamma_{iik} = -\frac{\gamma(N-1)}{N}$, $\gamma_{ijk} = \frac{\gamma}{N} > 0$ for $i \neq j$ and i and j from the same group and $\gamma_{ijk} = 0$ for i and j from different groups.

Bounded and Unbounded Pareto Distributions

I derive shape parameter for both unbounded and bound Pareto distribution cases. In order to save space, I provide derivations for the bounded case only and then analyze the results for finite and infinitely large upper bound of the distribution.

Cdf of bounded Pareto with parameters $(\theta, \underline{z}, \bar{z})$ is:

$$1 - F_Z(z) = P(Z > z) = \frac{\underline{z}^\theta (z^{-\theta} - \bar{z}^{-\theta})}{1 - (\underline{z}/\bar{z})^\theta} \quad (1)$$

That can be rewritten as $1 - \frac{z^{-\theta} - (\underline{z}/\bar{z})^{-\theta}}{1 - (\underline{z}/\bar{z})^\theta}$. The point of this paper is to estimate $(\theta, \underline{z}, \bar{z})$.

Translog Utility Function

Here I estimate shape parameter for translog utility function. This function is often used in international trade literature. Arkolakis et al. (2010) show the following link between firms' sales and their productivities:

$$x_{ij}(z) = \left[W \left(\frac{z}{\underline{z}_{ij}} e \right) - 1 \right] \gamma Y$$

Where β is a parameter of translog expenditure function, $\sum_i x_{ij} = Y_j$ and $W(y)$ is the Lambert function defined as the unique solution to $xe^x = y$. From the definition of the Lambert function, the inverse Lambert function $W^{-1}(x)$ is then just $W^{-1}(x) = xe^x$. I define market shares of the firms' as:

$$s_{ij} = \frac{x_{ij}(z)}{\sum_i x_{ij}} = \frac{x_{ij}(z)}{Y_j} = \left[W \left(\frac{z}{\underline{z}_{ij}} e \right) - 1 \right] \gamma \quad (3)$$

Starting from here I stop using country of origin and destination subscripts i and j .

Now given that $\frac{z}{\underline{z}} \sim \text{Pareto}(\theta, 1, \bar{z}/\underline{z})$ we have to derive the distribution of market shares s . First, re-express productivity through market shares:

$$\frac{z}{\underline{z}} = \frac{1}{e} W^{-1} \left(\frac{s}{\gamma} + 1 \right) = \left(\frac{s}{\gamma} + 1 \right) e^{\frac{s}{\gamma}} \quad (4)$$

Now combine (1) and (3):

$$1 - F_Z(z) = P\left(s \left(\frac{z}{\underline{z}} \right) > \frac{z}{\underline{z}}\right)$$

and together with (4) it becomes:

$$1 - F_S(s) = P\left(s\left(\frac{Z}{\underline{Z}}\right) > \frac{z}{\underline{z}}\right) = P\left(\frac{Z}{\underline{Z}} > \left(\frac{s}{\gamma} + 1\right)e^{\frac{s}{\gamma}}\right) = \frac{\left(\left(\frac{s}{\gamma} + 1\right)e^{\frac{s}{\gamma}}\right)^{-\theta} - (\underline{z}/\underline{z})^{-\theta}}{1 - (\underline{z}/\underline{z})^\theta}$$

Rearranging the terms obtain:

$$F_S(s) = 1 - \left[\frac{\left(\frac{s}{\gamma} + 1\right)^{-\theta} e^{-\frac{s\theta}{\gamma}}}{1 - (\underline{z}/\underline{z})^\theta} - \frac{(\underline{z}/\underline{z})^{-\theta}}{1 - (\underline{z}/\underline{z})^\theta} \right] \quad (5)$$

Taking derivative of (5) obtain the expression of pdf of market shares:

$$f_S(s) = \frac{\theta\left(\frac{s}{\gamma} + 1\right)^{-\theta} e^{-\frac{s\theta}{\gamma}} \left[\left(\frac{s}{\gamma} + 1\right)^{-1} + 1 \right]}{1 - (\underline{z}/\underline{z})^\theta} \frac{1}{\gamma} \quad (6)$$

Log-likelihood function is then:

$$l(\theta, \underline{z}, \bar{z}) = A + n \ln \theta - \theta \sum_i \ln\left(\frac{s_i}{\gamma} + 1\right) - \sum_i \left(\frac{s_i \theta}{\gamma}\right) - n \ln\left(1 - (\underline{z}/\bar{z})^\theta\right) \quad (7)$$

Where $A = \ln\left(n \frac{\left[\left(\frac{s}{\gamma} + 1\right)^{-1} + 1\right]}{\gamma}\right)$.

Upper and lower bounds enter only the last term of (7) and for any given value of θ the value of log-likelihood function is decreasing in \underline{z}/\bar{z} and, hence, ML-estimates of these parameters should be such that \underline{z}/\bar{z} is as small as possible. Notice that here \underline{z}/\bar{z} is the value of upper bound of Pareto distribution with parameters $(\theta, 1, \underline{z}/\bar{z})$. The smallest possible value of this upper bound should be largest observed draw from this productivity. As we observe shares rather than productivities, smallest possible value of the upper bound is determined by the largest observed market share. So, using (4) obtain¹²:

$$\underline{z}/\bar{z}_{ML} = \left(\max_i \left\{ \left(\frac{s_i}{\gamma} + 1\right)e^{\frac{s_i}{\gamma}} \right\} \right)^{-1}$$

and given that the inverse Lambert function is increasing

$$\bar{z} \equiv \hat{\underline{z}}_{ML} = \left(\left(\frac{s_{max}}{\gamma} + 1\right)e^{\frac{s_{max}}{\gamma}} \right)^{-1} \quad (8)$$

where s_{max} is the largest market share observed.

Put (8) in (7) and take the derivative of log-likelihood function with respect to θ :

¹²Here I have assumed that lower bound of the distribution is exogenous parameter of the distribution or is observed. In the model this lower bound is a complex endogenous object that depends on the shape parameter of the distribution.

$$\frac{\partial l(\theta, \underline{z}, \bar{z})}{\partial \theta} = \frac{n}{\theta} + \frac{n(\bar{z})^\theta \ln(\bar{z})}{1 - (\bar{z})^\theta} - \sum_i \ln\left(\frac{s_i}{\gamma} + 1\right) - \frac{\sum_i s_i}{\gamma} \quad (9)$$

And then $\hat{\theta}_{ML}$ is the solution to the following equation:

$$\frac{n}{\hat{\theta}} + \frac{n(\bar{z})^{\hat{\theta}} \ln(\bar{z})}{1 - (\bar{z})^{\hat{\theta}}} = \sum_i \ln\left(\frac{s_i}{\gamma} + 1\right) + \frac{1}{\gamma} \quad (10)$$

Now let's show that this solution exists and has asymptotic normal distribution.

Substitute (3) into (10) to get the expression of the estimate through productivities:

$$\frac{n}{\hat{\theta}} + \frac{n(\bar{z})^{\hat{\theta}} \ln(\bar{z})}{1 - (\bar{z})^{\hat{\theta}}} = \sum_i \ln\left(W\left(\frac{z}{\bar{z}}\right)\right) + \sum_i W\left(\frac{z}{\bar{z}}\right) + n \quad (11)$$

And now use the property of the Lambert function that $\ln W(x) = \ln(x) - W(x)$ and definition of \bar{z} to obtain:

$$\frac{n}{\hat{\theta}} + \frac{n(\underline{z}/\bar{z}_{ML})^{\hat{\theta}} \ln(\underline{z}/\bar{z}_{ML})}{1 - (\underline{z}/\bar{z}_{ML})^{\hat{\theta}}} = \sum_i \ln \frac{z}{\bar{z}} - n \quad (12)$$

Aban et al. (2006) shows that the estimate of θ similar to one in (12) has an asymptotic normal distribution with asymptotic mean θ . Moreover they show that the probability that a solution to the equation (12) exists converges to 1 as $n \rightarrow \infty$, and if it exists, then it is unique.

ML-Estimation of few bounded Pareto distributions

Here we assume there is more than one sample from bounded Pareto distribution with the same shape parameter, but with potentially different upper and lower bounds¹³. Likelihood function will be then a product of likelihood functions for each separate sample. From (7) log-likelihood function is equal to:

$$l(\theta, \underline{z}, \bar{z}) = \sum_j \left[A + n_j \ln \theta - \theta \sum_i \ln\left(\frac{s_{ij}}{\gamma} + 1\right) - \sum_i \left(\frac{s_{ij}\theta}{\gamma}\right) - n \ln\left(1 - (\underline{z}_j/\bar{z}_j)^\theta\right) \right]$$

Where $j = \{1, 2, \dots, J\}$ denotes country. Rearranging terms obtain:

$$l(\theta, \underline{z}, \bar{z}) = AJ + N \ln \theta - \theta \sum_j \sum_i \ln\left(\frac{s_{ij}}{\gamma} + 1\right) - \sum_j \sum_i \left(\frac{s_{ij}\theta}{\gamma}\right) - \sum_j n \ln\left(1 - (\underline{z}_j/\bar{z}_j)^\theta\right) \quad (13)$$

¹³Lower bound of exporters' productivities distribution is endogenous in this model, hence this assumption allows to take into account differences in wages and trade costs between the countries

Where N is a total number of firms. Notice that for any given values of θ and z_j/\bar{z}_j , $j \neq v$, minimum value of z_v/\bar{z}_v will maximize likelihood function. Thus similarly to (8) $z_v/\bar{z}_v \hat{z}_{ML} = \left(\frac{\max_i(s_{ij})}{\gamma} + 1\right)e^{\frac{\max_i(s_{ij})}{\gamma}}$. This is true for any v and hence:

$$z_j/\hat{z}_{ML} \equiv \tilde{z}_j = \left(\left(\frac{\max_i(s_{ij})}{\gamma} + 1\right)e^{\frac{\max_i(s_{ij})}{\gamma}}\right)^{-1} \text{ for } j = \{1, \dots, J\} \quad (14)$$

Taking derivative of (13) with respect to θ and combining this expression with (14) we can obtain the expression similar to (9):

$$\frac{\partial l(\theta, \underline{z}, \bar{z})}{\partial \theta} = \frac{N}{\hat{\theta}} + \sum_j \frac{n_j (\tilde{z}_j)^\theta \ln(\tilde{z}_j)}{1 - (\tilde{z}_j)^\theta} - \sum_j \sum_i^{n_j} \ln\left(\frac{s_{ij}}{\gamma} + 1\right) - \frac{\sum_j \sum_i^{n_j} s_{ij}}{\gamma}$$

Where $N = \sum_j n_j$. $\hat{\theta}$ is then a solution to the following equation:

$$\frac{N}{\hat{\theta}} + \sum_j \frac{n_j (\tilde{z}_j)^{\hat{\theta}} \ln(\tilde{z}_j)}{1 - (\tilde{z}_j)^{\hat{\theta}}} = \sum_j \sum_i \ln\left(\frac{s_{ij}}{\gamma} + 1\right) + \frac{J}{\gamma} \quad (15)$$

Expression (15) is different from expression (10) in two different ways. First, second term is weighed by a number of firms. Second, second term includes different \tilde{z}_j for different subsamples. Notice that if $n_j = n_i$ and $\tilde{z}_j = \tilde{z}_i$ for $\forall i, j$, then expressions (10) and (15) are similar and give the same solution for $\hat{\theta}$. The same would happen if $\tilde{z}_j = \tilde{z}_i$: it is similar to the case when a sample from one distribution is divided by few unequal subsamples, but all the observations are used to estimate θ . Let's consider the case when $n_j = n_i$ and $\tilde{z}_j \neq \tilde{z}_i$ for $\forall i, j$. In this case the only term that differs between different subsamples is a second term on the left hand side: $N \sum_j \frac{(\tilde{z}_j)^{\hat{\theta}} \ln(\tilde{z}_j)}{1 - (\tilde{z}_j)^{\hat{\theta}}}$. This term appears in the equation as a derivative of a denominator in expression (6): a term that scales density for given support of the distribution \tilde{z} . In other words, this term takes into account that different subsamples have different support and hence their contribution to likelihood function is different.

ML-Estimation unbounded Pareto distribution

ML-estimate of unbounded Pareto will be solution to the following equation:

$$\frac{n}{\theta} = \sum_i \log\left(\frac{s_i}{\gamma} + 1\right) + \sum_i \left(\frac{s_i}{\gamma}\right)$$

This expression is similar to expression (10), but without second term on the left hand side that becomes 0 as $\bar{z} \rightarrow \infty$. This expression has analytical solution:

$$\hat{\theta} = \frac{n}{\sum_i \log\left(\frac{s_i}{\gamma} + 1\right) + \frac{1}{\gamma}}$$

Expression for multiple sub-samples similar to (15) will be then:

$$\frac{N}{\hat{\theta}} = \sum_j \sum_i \ln\left(\frac{s_{ij}}{\gamma} + 1\right) + \frac{J}{\gamma} \quad (16)$$

And

$$\hat{\theta} = \frac{N}{\sum_i \log\left(\frac{s_i}{\gamma} + 1\right) + \frac{1}{\gamma}}$$

The only difference between expressions (15) and (16) is the second term on the left hand side of (15). As $\tilde{z}_j < 1$ for $\forall j$, this term is negative and increases in $\hat{\theta}$. It means that estimates of the shape parameter in the bounded case will be smaller than in the unbounded case. The intuition here is the following: we do not observe extremely large firms in the data. In case of unbounded Pareto it can be explained by thin tails of the distribution (high values of the shape parameter), while in bounded case there could be another reason why extremely large firms are not observed: productivity distribution can be bounded from above¹⁴.

¹⁴This assumption can be explained by technological limitations, a firm with infinitely large productivity cannot exist in real world