The Gravity Equation in International Trade: A Note

Ruben Dewitte*

12th October 2021

Chaney (2018) presents an intuitive, dynamic model of firm-to-firm trade that defines the functional relationship between aggregate trade φ and distance x under three conditions: if

(i) firm size $K \sim Pareto(K_{min}, \lambda)$ with shape parameter $\lambda \geq 1$ for all firms,

(ii) the average squared distance of exports is an increasing power function of firm size with power μ,

(iii) $\lambda < 1 + \mu$,

then the elasticity of aggregate trade with respect to distance is asymptotically constant:[1]

$$\varphi(x) \underset{x \to \infty}{\propto} \frac{1}{x^{1+2(\lambda-1)/\mu}}. \tag{1}$$

Moreover, if $\lambda \approx 1$ (firm exporter size approximates Zipf's law), the model provides an explanation for the distance coefficient in gravity equations often being close to one: $\varphi(x) \approx 1/x$.

This note tries and fails to replicate the empirical motivation for conditions (i)-(iii) of Chaney (2018). It uncovers multiple independent inconsistencies between the described methodology and its actual implementation. These discrepancies result in outcomes more closely aligned with conditions (i)-(iii) than those obtained from an exact implementation of the described methodology. Furthermore, this note demonstrates that both the parameter restrictions imposed by conditions (i) and (iii), as well as the functional form requirements set out in conditions (i)-(ii), are sensitive to sample size and are not simultaneously satisfied in the data. As a result, it becomes difficult to pin down 'the' asymptotic predicted distance elasticity. Therefore, the aim of this note is to incite dialogue on possible generalizations of Chaney (2018)'s model specification that can broaden its empirical support. In this regard, I kindly refer the reader to the reply from the original author to this article.

1 Replication

A replication of the empirical motivation for conditions (i)-(iii) of Chaney (2018) uncovered the following inconsistencies between the described methodology and its actual implementation:[2,3]

*Dewitte: Ghent University, Department of Economics, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be. I thank Harald Uhlig (the editor), Thomas Chaney, three anonymous referees, Bram De Lange, Gerdie Everaert, Gert Peersman, Glenn Rayp, Ilse Ruyssen, Angelos Theodorakopoulos, Milan van den Heuvel, and the Department of Economics at Ghent University. I acknowledge financial support from Research Foundation Flanders (FWO) grant 12B8822N. Data are provided as supplementary material online.

[1] Given the stability of the firm size distribution over time, Chaney (2018) interprets the theoretical model as an explanation for the distance puzzle. See Yotov (2012) for an alternative explanation of the distance puzzle relying on the differential effects of distance on inter- and intra-national trade.

[2] See online Appendix A for a detailed overview of the replication exercise with annotated original code.

[3] All reported standard errors in this note are robust standard errors. See the authors' reply for a description of the inconsistency regarding reported standard errors in the original paper.

1. (Chaney, 2018, Table 1, p. 162) reports a fitted Pareto shape parameter ($\lambda = 1.005$, $s.e. = .0288$) "using data on all 27,903 French firms that export more than 1 million French francs in 1992 (≈\$200,000)" (Chaney, 2018, p. 162). The actual implementation, however, additionally omits 7,477 firms. An exact implementation, relying on the complete dataset, produces a Pareto shape parameter below one ($\lambda = 0.971$, $s.e. = .0289$). This inconsistency has two implications for the reported results in Table 1. First, a point estimate of λ below one results in a degenerate equilibrium of the theoretical model where old firms become "too" large (Chaney, 2018, p.167). Second, the reported μ parameter in Table 1 is obtained under the sample described in the original contribution (firms with exports > 1 million French francs in 1992). As a result, five out of the eight rows in Table 1 conflate parameter estimates obtained from different sample sizes.
2. (Chaney, 2018, p.160) states that "As proposition 1 is an asymptotic result, I consider two specifications of (6): Either with all distances or with only distances above 2,000 kilometers (approximately 1,200 miles, about 80 percent of the sample)." The author's code reveals that the distance elasticity $\zeta_{long} = 1.090$ ($s.e. = .210$) is obtained from using a cutoff at 1,900 kilometers. Applying the reported cutoff of 2,000 kilometers results in a distance elasticity of $\zeta_{long} = 1.185$ ($s.e. = 0.250$).
3. Chaney (2018) visually evaluates the functional form requirements of conditions (i)-(ii), claiming that "The straight lines in the top and bottom panels of fig. 1 correspond to fitted regression lines from estimating eqq. (4) and (5) respectively" (Chaney, 2018, footnote 9, p. 162). This claim is not correct.
   - 3.1. The upper panel of (Chaney, 2018, fig. 1) displays a regression line fitted on restricted data (see item 1), vertically shifted, and omitting a fitted observation. The difference between this omitted observation and the (normalized) empirical CDF is 0.53. An updated visualization reveals that observed deviations from the data are too large to defend Pareto, and therefore condition (i), providing a good fit to the specified data sample.
   - 3.2. Similar to the upper panel of (Chaney, 2018, fig. 1), the regression line displayed in the lower panel is fitted on restricted data and is vertically shifted. An updated visualization indicates that a simple log-linear specification between firm size and average squared distance of exports, and therefore condition (ii), does not suffice.
4. When performing the Wald test for *all* distances, Chaney (2018) relied on the actual distance elasticity for *long* distances in the numerator of the Wald test statistic. Correcting for this inconsistency lowers the p-value from the reported 99.3% (Chaney, 2018, p.162) to 54.9%.

Overall, of the two figures and eight lines of results presented in (Chaney, 2018, p.161-162) to motivate the theoretical model assumptions, only two results can be replicated using the methodology as stated in the paper. The discrepancy between the stated methodology and the results reported is due to multiple deviations in the STATA code. These discrepancies result in outcomes more closely aligned with conditions (i)-(iii) than those obtained from an exact implementation of the described methodology.

Provided the replication failure, this note aims to provide the outcomes obtained from an exact implementation of the described methodology and evaluate the sensitivity of these outcomes with respect to sample size. Figure 1 provides an overview of this replication exercise, displaying the evolution of (a) the estimated Pareto shape parameter (λ) and the power exponent of the specified firm size - distance relation, increased by one ($\mu + 1$), (b) the predicted distance elasticity ($1 + 2(\lambda - 1)/\mu$) and (c) the estimated distance elasticity (ζ), as a function of their respective minimum bounds used to define the sample.

Parameter estimates obtained from an exact implementation of the described methodology are indicated by white squares in Figure 1, while white diamonds indicate parameter estimates as reported by Chaney (2018). The estimates based on an exact implementation are obtained from the available full sample of firms and are therefore located on the utter left side of the x-axis. As described above, this exact implementation results in a point estimate of the Pareto shape parameter below one, violating condition (i). Still, naively calculating the predicted distance elasticity provides an elasticity of 0.48 (see panel (c) of Figure 1). Its accompanying confidence interval indicates that this elasticity is not significantly different from one, neither is this naive estimate significantly different from the estimated distance elasticity for all ($\zeta_{all} = 0.77$) or long ($\zeta_{long} = 1.19$) distances.[4] However, the success of these statistical tests cannot be viewed as a validation of the theoretical model. The parameter restriction of condition (i) is not satisfied when the Pareto shape parameter falls below unity, and the functional form requirements set out in conditions (i)-(ii) are not satisfied for this specific data sample (see item 3.1 and 3.2 above).

Therefore, this note discusses the sensitivity of the theoretical parameter restrictions and functional form requirements to alternative lower bounds that define the sample. First, panel (a) of Figure 1 demonstrates that the Pareto shape parameter (λ) ranges between 0.97 and 1.74 for alternative export cutoff values. Meanwhile, the power exponent of the specified firm size - distance relation increased by one ($\mu + 1$) takes values between 1.03 and 1.25. As a result, condition (iii) only holds for cutoff values of exports smaller than 8.63 million FF. Whereas the author states that "in practice, condition (iii) is unlikely to be violated." (Chaney, 2018, p. 156), the range of export cutoff values for which condition (iii) is violated accounts for 94.4% of total exports above 1 million FF. Additionally imposing the parameter restriction of condition (i) restricts the sample range where all parameter restrictions hold to samples with cutoff values spanning between 1.91 to 8.63 million FF, as indicated by the white area in panels (a) and (b) of figure 1.[5] This range accounts for 4.6% of total exports above 1 million FF (between 99.0% and 94.4% of total exports above 1 million FF), or equivalently, for 9.6% of all exporting firms (between 16.9% and 7.3%).[6]

As for the functional form requirements, these are not satisfied within this sample range. The Pareto assumption for firm size, for instance, is more likely to hold for samples that consider only the largest 5% (Freund and Pierola, 2015) and/or 1% Head et al. (2014); Bas et al. (2017) of exporting firms,[7] which falls outside the sample range where parameter restrictions are satisfied. Overall, both parameter restrictions and functional form requirements are sensitive to sample size and cannot simultaneously be satisfied according to the data.

The sensitivity of model conditions (i)-(iii) to sample size renders theoretical predictions volatile (see panel (b) of Figure 1). As such, it becomes hard to pin down 'the' asymptotic *predicted* distance elasticity. Within the theoretically acceptable parameter range, predicted distance elasticities increase from 1.09 to 2.69. In comparison, point estimates for the distance elasticity of trade initially oscillate around the value one in function of distance (see panel (c) of Figure

[4] A Wald test on the equality between the actual and naively predicted distance elasticities of trade attains a p-value of 59.4% for all distances and 22.7% for long distances.

[5] The white areas in panel 1a and 1b indicate the sample range that satisfies the parameter restrictions of conditions (i) the Pareto parameter is larger than one ($\lambda \geq 1$) to avoid a degenerate equilibrium of the theoretical model where old firms become "too" large (Chaney, 2018, p.167), and (iii) the Pareto shape parameter is small enough such that the entry of new firms does not exceed the sum of the growth rates of individual firms and the gross creation of new contacts ($\lambda < 1 + \mu$).

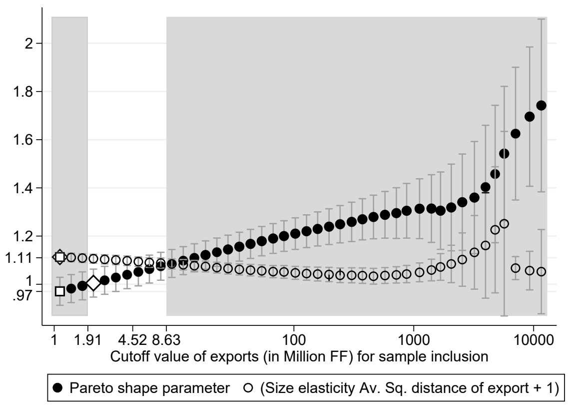
[6] This range is further restricted to cutoff values spanning between 1.91 and 4.52 million FF if one requires the point estimate ($\mu + 1$) to be significantly larger than (λ). The resulting range of cutoff values accounts for 2.1% of total exports above 1 million FF (between 99.0% and 96.9% of total exports above 1 million FF), or equivalently, for 6.2% of all exporting firms (between 16.9% and 10.7%).

[7] See also online Appendices B and C.

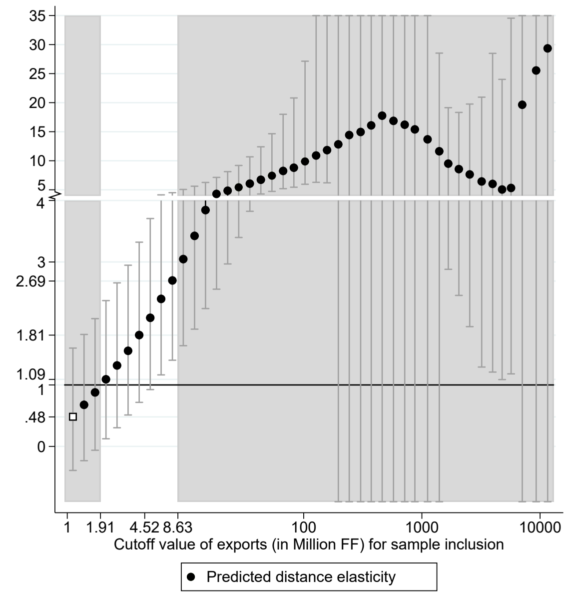
1) before diverging for long distances. The reported point estimate of the 'long' actual distance elasticity (for distances above 1,900 kilometers) by Chaney (2018) transpires to be a snapshot of reality. Consequently, it also becomes hard to pin down 'the' asymptotic *actual* distance elasticity as a benchmark to compare predicted distance elasticities with.

2 Conclusion

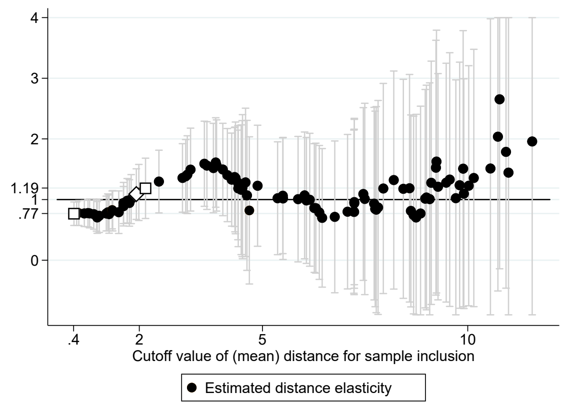
In conclusion, the replication exercise reveals that Chaney (2018)'s theoretical predictions do not hold up as a general result. The revealed tension between parameter restrictions and functional form requirements can guide further dialogue on alternative model specifications. For instance, the relied upon Steindl (1965) growth process is, in the context of city size distributions, perceived as a problematic explanation for Zipf's law due to its empirically implausible parametric restrictions (Gabaix, 1999). Also, one could consider more general functional form specifications for the firm size distribution, bringing two advantages. First, the point estimates of the distance elasticity reported in panel (c) of Figure 1, similar to distance elasticities reported by the author in online Appendix Table 4 (Chaney, 2018, Online Appendix p.22), seem to point at a richer evolution of the distance elasticity than a gradual decrease towards a constant value (of one). In line with the literature on the variability of the aggregate trade elasticity (the response of trade flows to a change in trade costs) across country pairs (Melitz and Redding, 2015; Bas et al., 2017), one could consider alternative firm size distributions that help to explain this evolution. Second, the Pareto/Zipf distribution only covers the right tail of the firm size distribution, without indication of the length of this right tail. Alternative distributions could spur research on the functional relation between distance and aggregate trade for both small and large firms. Given the widespread reliance on the gravity equation for all firm sizes, it is all the more important to fully understand the implications of this equation.



(a) Estimated parameters as a function of the lower bound of sales (K_{min}) for inclusion in the sample[a,c,d]



(b) Predicted distance elasticity as a function of the lower bound of sales (K_{min}) for sample inclusion[a,b,c,d]



(c) Estimated distance elasticity as a function of the lower bound distance for inclusion in the sample[a,b,d]

Figure 1: Evolution of the parameter estimates and model outcomes as a function of their respective minimum bounds used to define the sample.

Note: All results obtained from samples of at least 5 observations. [a]Light-grey capped spikes (I-beams) delineate the 95% confidence intervals in panel (a) and (c) and 95% bootstrapped confidence intervals for panel (b). [b]For visual purposes, confidence intervals are restricted to take values between -0.9 and 35 in panel (b) and between -0.9 and 4 in panel (c). [c]White areas indicate the theoretically acceptable parameter point estimates (see also footnote 8). [d]White squares indicate the parameter estimates and model outcomes obtained from an exact implementation of the described methodology of Chaney (2018). White diamonds indicate parameter estimates as reported by Chaney (2018). Note that the predicted elasticity by Chaney (2018) can not be displayed in panel (b) as it depends on parameters obtained under different cutoff values of exports for sample inclusion.

References

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1 – 19.

Chaney, T. (2014). The network structure of international trade. *The American Economic Review 104*(11), 3600–3634.

Chaney, T. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy 126*(1), 150–177.

Freund, C. and M. D. Pierola (2015). Export Superstars. *The Review of Economics and Statistics 97*(5), 1023–1032.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics 114*(3), 739–767.

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

Melitz, M. J. and S. J. Redding (2015). New trade models, new welfare implications. *American Economic Review 105*(3), 1105–46.

Steindl, J. (1965). *Random processes and the growth of firms: a study of the Pareto Law.* London: Charles Griffin.

Yotov, Y. V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters 117*(3), 794 – 798.

Online Appendix to "The Gravity Equation in International Trade: A Note"

Ruben Dewitte*

12th October 2021

*Dewitte: Ghent University, Department of Economics, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be. I thank Harald Uhlig (the editor), Thomas Chaney, three anonymous referees, Bram De Lange, Gerdie Everaert, Gert Peersman, Glenn Rayp, Ilse Ruyssen, Angelos Theodorakopoulos, Milan van den Heuvel, and the Department of Economics at Ghent University. I acknowledge financial support from Research Foundation Flanders (FWO) grant 12B8822N. Data are provided as supplementary material online.

Appendix A Implementation inconsistencies

This appendix details the inconsistencies between the described methodology in Chaney (2018) and the actual implementation. It relies on the code and data provided in the supplemental materials by Chaney (2018), accessible at doi.org/10.1086/694292. Specifically, it relies on the STATA code in the location 'ReplicationFiles/Analysis/Code/Analysis_A_redux.do'. The reported code line indicators correspond to the code lines in the original file, while the code highlighted in red pinpoints the origin of the reported inconsistencies.

1. The minimum firm size, K_{min}, is of critical importance to the empirical exercise in Chaney (2018). It determines the size of the data sample and is a parameter of the Pareto distribution. Hence, the lower bound influences the ensuing estimation of the Pareto shape parameter λ, featuring a positive correlation (see, for instance, Head et al. (2014)). Chaney (2018) chooses to set an arbitrary lower bound at 1 million francs, resulting in a dataset of 27,968 firms out of a total of 121,581 French exporters in 1992 (Chaney, 2014, p.3605). However, it appears from the STATA code that a higher cutoff has been used only for fitting the Pareto distribution:

   ```
   89  capture quietly reg ln_rank_n ln_exp_n if bin>2
   ```

   Omitting the first three bins ("if bin>2" in the STATA code) resets the lower bound to approximately 1.91 million francs, resulting in a dataset of 20,491 firms ($\pm$ 73% of the original dataset).

   This inconsistency has two implications for the reported results in Table 1. First, relying on the complete dataset results in a Pareto shape parameter λ below one ($\lambda = 0.971$, $s.e. = .0289$). A λ below one, though not significantly so, results in a degenerate equilibrium of the theoretical model where old firms become "too" large (Chaney, 2018, p.167). Second, the reported μ parameter in Table 1 is obtained under a different sample, specifically the sample as described in the original contribution (firms with exports $>$ 1 million French francs in 1992). As a result, five out of the eight rows in Table 1 rely on parameter estimates obtained from different sample sizes.

2. (Chaney, 2018, p.160) claims, "As proposition 1 is an asymptotic result, I consider two specifications of (6): Either with all distances or with only distances above 2,000 kilometers (approximately 1,200 miles, about 80 percent of the sample).". The code of the author reveals that the distance elasticity $\zeta_{long} = -1.090$ ($s.e. = .210$) is obtained from using a cutoff at 1,900 kilometers:

   ```
   225  capture quietly reg ln_exports ln_distance ln_imports if dist>1.9
   ```

   Using the cutoff at 2,000 kilometers results in a distance elasticity of $\zeta_{long} = -1.185$ ($s.e. = 0.250$) as 2 extra observations (of the 84) are dropped.

3. Chaney (2018) claims that "The straight lines in the top and bottom panels of fig. 1 correspond to fitted regression lines from estimating eqq. (4) and (5) respectively" (Chaney, 2018, footnote 9, p.162). This claim is not correct.

   3.1. The upper panel of (Chaney, 2018, fig. 1) displays a straight line that is fitted on restricted data ("if bin> 2"), omits a fitted observation ("bin==3"), and is vertically shifted ("local anchor=45").

   ```
   458  *fit rank-size regression line
   459  capture quietly reg ln_rank_n ln_exp_n if bin>2
   ```

```
capture mat temp=e(b)
local lambda=-temp[1,1]
local anchor=45
local rank_anchor=rank_n[`anchor']
local exp_anchor=exp_n[`anchor']
gen
    rank_n_hat=`rank_anchor'*`exp_anchor'^(`lambda')*exp_n^(-`lambda')

*generate rank-size plot on a log-log scale
twoway (scatter rank_n exp_n ) ///
  (line rank_n_hat exp_n if bin>3,sort), ...
```

Figure 1 demonstrates the implications of the deviations between the authors' claims and actions. The replication reveals that the omitted observation from the fit amounts to 1.19. This value deviates largely from the empirical CDF with a difference of 0.53, as can be deduced from the right panel of Figure 1. It is hard to interpret these differences in terms of distribution fit, however, for two reasons.

First, we are comparing a properly normalized empirical CDF with a non-normalized power-law distribution.[1] To represent the Complementary CDF of a properly normalized Pareto distribution, the displayed straight line should pass through the value one at its lower bound: $1 - Pareto(K_{min}) = \left(\frac{K_{min}}{K_{min}}\right)^{\lambda} = 1$. The replication updates the fit to the complete dataset and displays the resulting Complementary CDF[2] as the dashed orange line in the left panel of Figure 1. We notice that this updated CDF differs largely from the empirical CDF over a broad range of the data: 13 out of the 50 bins display a gap greater than 0.1 with a maximum deviation of 0.21, as can also be observed from the right panel.

Second, a visual inspection of the deviations between a fitted Pareto CDF and the empirical CDF requires a benchmark relative to which these distances can be evaluated. I propose to rely on the Kolmogorov-Smirnov test to acquire such a benchmark.[3] This test provides us with the intuition that the maximum deviation between the empirical and Pareto CDF, under the null hypothesis that the data follows a Pareto distribution, would at most attain the value of 0.122 in 95% of the cases. This leads us to conclude that the observed differences, with a maximum deviation of 0.21, represent too large an error to defend Pareto being a good fit, at least to this range of the data.[4]

[1] Chaney (2018) estimates the Pareto shape parameter performing an OLS log-log rank-size regression. This exercise is designed solely to estimate the shape parameter of the Pareto distribution, it is not designed to estimate a properly normalized Pareto CDF with values ranging between 0 and 1. The claim that "Figure 1 shows that conditions (i)–(iii) in proposition 1 are approximately satisfied. The top panel shows that the distribution of firm sizes is well approximated by Zipf's law ..." (Chaney, 2018, pp. 161-162) was therefore unfounded: no Pareto distribution was displayed in the original article.

[2] $1 - Pareto(K) = \left(\frac{1}{K}\right)^{0.971}$.

[3] See online Appendix B for an overview of the performed Kolmogorov-Smirnov test.

[4] The large deviation of the fitted CDF from the empirical values stands in stark contrast with the high R^2 value reported by the author in evidence of Pareto: "The Pareto distribution in condition (i) offers a precise approximation of the distribution of firm sizes. The R2 from estimating equation (4) is 98.1 percent." (Chaney, 2018, p.162). This is the consequence of the argument made in footnote 1: an OLS regression fit does not represent a proper CDF. The R^2 obtained from a simple OLS is then not informative on the distribution fit.

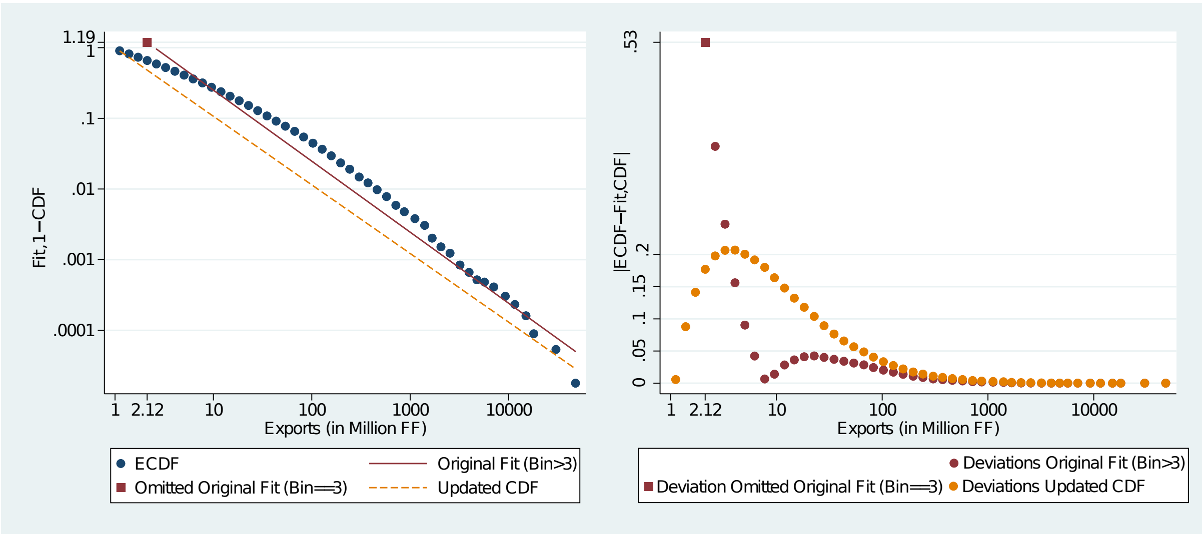


Figure 1: The original fitted regression line in the **left panel** does not represent a proper CDF, as becomes clear when displaying the omitted fit with a value of 1.19. The Updated CDF obtained from fitting to all data showcases a large divergence over a broader range of the data. This can be deduced from the **right panel**, showcasing the absolute difference between empirical and fitted values.

3.2. Similar to the upper panel of (Chaney, 2018, fig. 1), the straight line displayed in the lower panel is fitted on restricted data ("if bin<37") and is vertically shifted ("local anchor=8"):

```
capture quietly reg ln_Delta_n ln_exp_n if bin<37
capture mat temp=e(b)
local mu=temp[1,1]
local anchor=8
local Delta_anchor=Delta_n[‘anchor’]
local exp_anchor=exp_n[‘anchor’]
gen Delta_n_hat = ‘Delta_anchor’*‘exp_anchor’^(-‘mu’)*exp_n^(‘mu’)
```

The implications of the deviations between the authors' claims and actions for the lower panel are displayed in Figure 2. The replicated line has a slope of $\mu = 0.113$ ($s.e. = 0.008$) and an accompanying R2 of 81.7% for the full sample. This replicated line is vertically shifted upwards and has a less steep slope relative to the originally displayed straight line, which had a slope of $\mu = 0.152$ ($s.e. = 0.005$) and an accompanying R2 of 96.5% for the relied upon restricted sample (bin<37).[5] Moreover, a log-quadratic fit to the complete sample increases the R2 to 89.4 percent. This indicates that a simple log-linear specification between firm size and average squared distance of exports might not suffice.[6]

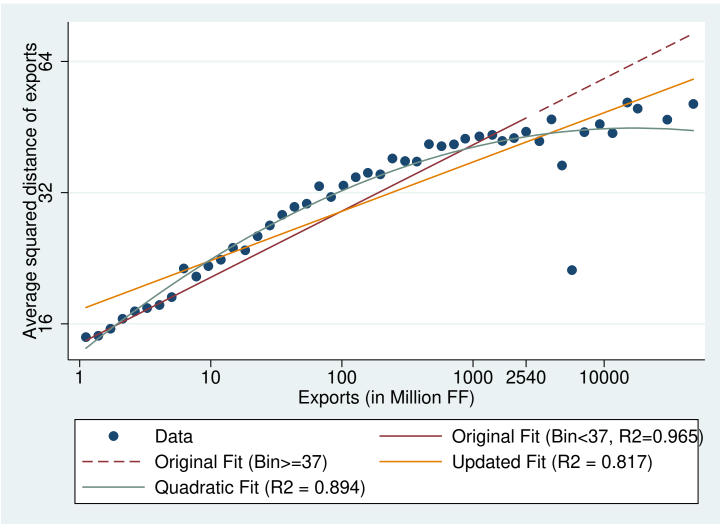


Figure 2: Original straight line over its fitted (bin<37, full line) and excluded (bin≥37, dashed line) data range, as well as the replicated log-linear fitted line and a log-quadratic fitted line.

4. Chaney (2018) relies on a Wald test to test the equality between the actual and predicted distance elasticities of trade. When performing the Wald test for *all* distances, the author uses the difference between the predicted distance elasticity and the actual distance elasticity for *long* distances ('theta' rather than 'theta_all') as the numerator of the test statistic:

```
*Wald test for proposition 1 (all distances)
```

[5]Note that the claim "Figure 1 shows that conditions (i)–(iii) in proposition 1 are approximately satisfied. ... The bottom panel shows that the average squared distance of exports is approximately a power function of firm size, $\Delta(K) \propto K^{\mu}$ with $\mu = 0.113$" (Chaney, 2018, pp. 161-162) was therefore unfounded: the displayed straight line had a slope of $\mu = 0.152$.

[6]See online Appendix C for a formal statistical test which confirms this observation.

```
gen theta_all=(zeta_all-1-2*(lambda-1)/mu)
gen theta2_all=theta^2
```

Correcting for this inconsistency, the reported p-value attains a value of 54.9 percent rather than the reported 99.3% in the original paper (Chaney, 2018, p.162).

Appendix B Kolmogorov-Smirnov test

Chaney (2018) relies on both a visual inspection of the upper panel of Figure 1 on p.161 and the reported R^2 in Table 1 on p.162 to evaluate whether the Pareto distribution is a sufficiently good approximation of the actual firm size distribution, as condition (i) requires:

- "Figure 1 shows that conditions (i)–(iii) in proposition 1 are approximately satisfied. The top panel shows that the distribution of firm sizes is well approximated by Zipf's law" (Chaney, 2018, pp. 161-162)
- "Table 1 presents formal statistical tests for conditions (i)–(iii) and proposition 1. The Pareto distribution in condition (i) offers a precise approximation of the distribution of firm sizes. The R2 from estimating equation (4) is 98.1 percent." (Chaney, 2018, p.162)

However, a visual inspection of the deviations between a fitted Pareto CDF and the empirical CDF requires a benchmark relative to which these distances can be evaluated. Moreover, the reported R^2 is obtained from an estimation procedure solely designed to estimate the shape parameter of the Pareto distribution, not to estimate a properly normalized Pareto CDF. The obtained R^2 can then not be informative on whether the Pareto distribution offers a precise approximation of the firm size distribution observed in the data (which is properly normalized).

To resolve these two shortcomings, I propose to rely on the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic quantifies the distance between the empirical distribution function ($F(K_b)$)[7] and the cumulative distribution function of the Pareto distribution ($Pareto(K_b)$) as follows:

$$KS = \sup_{K_b} \left| F(K_b) - Pareto(K_b | K_{min}, \lambda) \right| . \tag{1}$$

There exists, to my knowledge, no critical values for this test when the parameters of the Pareto distribution are determined from binned data. Therefore, we need to rely on a parametric bootstrap:

1. $H_0 : F(\cdot) = Pareto(\cdot | K_{min}, \hat{\lambda})$;
2. Draw B bootstrap samples of size N from $Pareto(\cdot | K_{min}, \hat{\lambda})$ and construct bins of equal log width as Chaney (2018), with the number of bins equal to the number of bins in the original data;
3. Obtain an estimate of the shape parameter of the Pareto distribution for each bootstrap sample of firm sizes, $\hat{\lambda}^*$;
4. For each sample of the parametric distribution, calculate the KS statistic

$$KS^* = \sup_{K_b^*} \left| F(K_b^*) - Pareto(K_b^* | K_{min}, \lambda^*) \right| ; \tag{2}$$

[7] Note that (Chaney, 2018, p.159) specifies this empirical distribution function as $F(K_b) = 1 - \frac{\sum_{b'=b}^{50} \sum_i \mathbf{1}[i \in b']}{\sum_{b''=1}^{50} \sum_i \mathbf{1}[i \in b'']}$. In the actual implementation, however, the author uses another specification $F(K_b) = 1 - \frac{\sum_{b'=b}^{50} \sum_i \mathbf{1}[i \in b'] - 0.5}{\sum_{b''=1}^{50} \sum_i \mathbf{1}[i \in b'']}$, as becomes obvious from the STATA code (See codefile Analysis_A.do, line 123). As we only have access to the binned dataset and as the influence of this deviation on the KS-test statistic is small, we continue using the second specification as used by the author for his implementation.

5. One can then retrieve specific percentiles of the set of bootstrapped KS-statistics as well as the p-value, which is defined as

$$\hat{p} = \frac{1}{B+1}\left[\sum_{B} \mathbf{1}\left(KS^{*} \geq KS\right) + 1\right]. \tag{3}$$

This bootstrap procedure provides us with an indication of the likelihood of observing a deviation between the empirical and Pareto CDF as large as KS under the null hypothesis. As such, we attain a benchmark relative to which visual deviations between an empirical CDF and a properly normalized Pareto CDF can be interpreted, as well as an alternative formal statistical test for the functional form requirement set out in condition (i).

We perform the bootstrap test over the first 15 bins as lower bounds for inclusion in the sample. Each bootstrap test is performed with $B = 10,000$ replications and sample size equal to the remaining original sample size after setting the respective bins as lower bounds. The results of the tests are reported in Table 1. The test statistics are large for lower bins as sample cutoffs, with errors of at least 0.2 on a CDF range between 0 and 1. These large errors result in a statistically significant rejection of the null hypothesis, as can be deduced from the reported p-values. When the full sample is used, the reported percentiles of the set of bootstrapped KS-statistics provide us with the intuition that the maximum deviation between the empirical and Pareto CDF under the null hypothesis would at most take a value of 0.122 in 95% of the cases. This leads us to conclude that the observed differences represent too large an error to defend Pareto being a good fit, at least for these values of sample cutoffs. The KS-statistics decrease as the lower bounds increase, however. When considering only 7.29% of 121,581 exporting firms in the sample, we obtain a p-value of 0.094 so that we cannot reject the assumption that the observed data originates from a Pareto distribution. This result is in line with earlier literature that considers only the largest 5% (Freund and Pierola, 2015) and/or 1% (Head et al., 2014; Bas et al., 2017) of exporting firms to fit a Pareto distribution.[8]

[8] Please note that the presented formal statistical test relies on binned data and it is therefore unclear how powerful the test is.

Table 1: Kolmogorov-Smirnov test over different bins as lower bounds for inclusion in the sample.

Selected firms	**Perc. firms**[a]	**KS**	**P-Value**	$P_{90}{}^b$	$P_{95}{}^b$	$P_{99}{}^b$
Firms above 1.00 million FF (bin 0 and above)	23.00	0.207	0.000	0.114	0.122	0.141
Firms above 1.24 million FF (bin 1 and above)	20.81	0.200	0.000	0.116	0.124	0.144
Firms above 1.54 million FF (bin 2 and above)	18.77	0.193	0.000	0.117	0.126	0.146
Firms above 1.91 million FF (bin 3 and above)	16.85	0.188	0.000	0.119	0.128	0.148
Firms above 2.37 million FF (bin 4 and above)	15.14	0.182	0.001	0.120	0.129	0.150
Firms above 2.94 million FF (bin 5 and above)	13.53	0.172	0.003	0.121	0.131	0.154
Firms above 3.64 million FF (bin 6 and above)	12.04	0.164	0.006	0.123	0.133	0.155
Firms above 4.52 million FF (bin 7 and above)	10.66	0.155	0.011	0.125	0.135	0.157
Firms above 5.61 million FF (bin 8 and above)	9.40	0.150	0.020	0.126	0.137	0.159
Firms above 6.96 million FF (bin 9 and above)	8.31	0.141	0.041	0.128	0.138	0.162
Firms above 8.63 million FF (bin 10 and above)	7.29	0.131	0.094	0.130	0.140	0.165
Firms above 10.71 million FF (bin 11 and above)	6.34	0.121	0.196	0.132	0.142	0.170
Firms above 13.28 million FF (bin 12 and above)	5.48	0.115	0.294	0.133	0.144	0.173
Firms above 16.48 million FF (bin 13 and above)	4.73	0.112	0.358	0.135	0.147	0.177
Firms above 20.44 million FF (bin 14 and above)	4.08	0.109	0.441	0.138	0.150	0.182
Firms above 25.36 million FF (bin 15 and above)	3.50	0.103	0.596	0.141	0.153	0.182

Note: Rows colored in gray indicate the sample ranges for which the parameter restrictions imposed by conditions (i) and (iii) do not hold. [a]Perc. firms indicates the number of selected firms as a percentage of all 121,581 exporting firms. [b]P_x indicates the xth percentile of the set of bootstrapped KS-statistics.

Appendix C Ramsey (1969)-RESET test

Chaney (2018) relies on both a visual inspection of the lower panel of Figure 1 on p.161 and the reported R^2 in Table 1 on p.162 to evaluate whether a log-linear specification is sufficient to model the relationship between the average squared distance of exports ($\Delta(K_b)$) and firm size (K_b), as condition (ii) requires:

- "The bottom panel shows that the average squared distance of exports is approximately a power function of firm size, $\Delta(K) \propto K^{\mu}$ with $\mu = 0.113$, as in condition (ii)." (Chaney, 2018, p.162)
- "The relationship between firm size and the average squared distance of exports is close to the log-linear relation of condition (ii). The R2 from estimating equation (5) is 81.7 percent" (Chaney, 2018, p.162)

However, the displayed straight line in the bottom panel of Figure 1 is obtained from fitting a restricted data sample. This resulted in a displayed straight line with an actual slope of $\mu = 0.152$ and an R2 of 96.5% for this restricted sample. Moreover, a log-quadratic fit to the complete sample can increase the R2 to 89.4%. This raises the question whether a log-linear specification is sufficient to model the relationship between the average squared distance of exports and firm size, as in (Chaney, 2018, eq. (8) p. 160):

$$ln\Delta(K_b) = a + \mu ln(K_b) + \epsilon_b.$$

We test whether this is the case relying on the Ramsey (1969)-RESET test. Specifically, we re-estimate the previous regression with up to four powers ($\alpha = 2, \ldots, 4$) of the fit to the dependent variable $\left(\widehat{ln\Delta(K_b)}\right)^{\alpha}$:

$$ln\Delta(K_b) = a + \mu ln(K_b) + t_2\left(\widehat{ln\Delta(K_b)}\right)^2 + \ldots + t_4\left(\widehat{ln\Delta(K_b)}\right)^4 + \epsilon_b,$$

and perform an F-test under the null hypothesis that the coefficients on the added covariates are jointly zero:[9]

$$H_0: \quad 0 = t_2 = t_3 = t_4. \tag{4}$$

If this null hypothesis is rejected, then the regression is misspecified, and the log-linear relationship between the average squared distance of exports and firm size is deemed insufficient.

We conduct the Ramsey (1969)-RESET test under the null hypothesis that the average squared distance of exports is a power function of firm size over different lower bounds of bin membership for inclusion in the sample. The results are reported in Table 2. One can observe that the test statistics are large, resulting in a statistically significant rejection of the hypothesis that a simple log-linear specification between firm size and average squared distance of exports suffices. The test statistics are decreasing with increasing lower bounds, however. When we only consider 4.08% of 121,581 exporting firms in the sample, we obtain a p-value of 0.076 so that we cannot reject the assumption of a log-linear relation between the average squared distance of exports and firm size.

[9] This test is performed in stata using the postestimation command "estat ovtest".

Table 2: Ramsey (1969)-RESET test over different bins as lower bounds for inclusion in the sample.

Selected firms	Perc. firms[a]	RESET	P-Value
Firms above 1.00 million FF (bin 0 and above)	23.00	15.98	0.000
Firms above 1.24 million FF (bin 1 and above)	20.81	14.97	0.000
Firms above 1.54 million FF (bin 2 and above)	18.77	13.77	0.000
Firms above 1.91 million FF (bin 3 and above)	16.85	12.57	0.000
Firms above 2.37 million FF (bin 4 and above)	15.14	11.49	0.000
Firms above 2.94 million FF (bin 5 and above)	13.53	10.44	0.000
Firms above 3.64 million FF (bin 6 and above)	12.04	9.25	0.000
Firms above 4.52 million FF (bin 7 and above)	10.66	7.94	0.000
Firms above 5.61 million FF (bin 8 and above)	9.40	6.75	0.001
Firms above 6.96 million FF (bin 9 and above)	8.31	6.36	0.002
Firms above 8.63 million FF (bin 10 and above)	7.29	5.41	0.004
Firms above 10.71 million FF (bin 11 and above)	6.34	4.64	0.008
Firms above 13.28 million FF (bin 12 and above)	5.48	3.88	0.018
Firms above 16.48 million FF (bin 13 and above)	4.73	3.35	0.032
Firms above 20.44 million FF (bin 14 and above)	4.08	2.54	0.076
Firms above 25.36 million FF (bin 15 and above)	3.50	2.00	0.136

Note: Rows colored in gray indicate the sample ranges for which the parameter restrictions imposed by conditions (i) and (iii) do not hold. [a]Perc. firms indicates the number of selected firms as a percentage of all 121,581 exporting firms.

Appendix References

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1 – 19.

Chaney, T. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy 126*(1), 150–177.

Freund, C. and M. D. Pierola (2015). Export Superstars. *The Review of Economics and Statistics 97*(5), 1023–1032.

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

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological) 31*(2), 350–371.