The Log of Heavy-Tailed Gravity

Ruben Dewitte, and Samuel Standaert*

September, 2022

Abstract

The gravity equation has been the workhorse model for analyzing the determinants of bilateral trade flows for over 50 years. Predominantly, this gravity equation is estimated using Poisson Pseudo-Maximum-Likelihood (P-PML). This paper demonstrates that P-PML yields imprecise estimates if the underlying data is heavy-tailed, which is the case for trade data. Consequently, we argue that the gravity equation should be estimated using a heteroskedasticity-robust log-linearized version of the gravity equation. Log-linearization of the data reduces the influence of the heavy tail on the coefficient estimates. We demonstrate the superiority of such heteroskedasticity-robust log-linear estimators over P-PML through a Monte Carlo exercise and an application to bilateral trade data.

Keywords: Gravity, International Trade, Heavy-tailed Data, Pseudo-Maximum-Likelihood

JEL Codes: C13, C18, F14

*Dewitte (corresponding author): Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium, rubenl.dewitte@ugent.be; Standaert: United Nations University and Ghent University, Potterierei 72, 8000 Brugge, sstandaert@cris.unu.edu. Ruben Dewitte gracefully acknowledges financial support from the Research Foundation Flanders (FWO) grant 12B8822N.

1 Introduction

The gravity equation provides a parsimonious and tractable representation of economic interaction in a many country world. It has been the workhorse model for analyzing the determinants of bilateral trade flows for over 50 years. This popularity exists despite the challenges bilateral trade flows pose for the applied researcher. The distribution of bilateral trade data is non-negative with many (non-random) zero values, heteroskedastic, positively skewed and heavy-tailed. The seminal paper of Silva and Tenreyro (2006) led to a predominant reliance on the Poisson Pseudo-Maximum-Likelihood (P-PML) estimator to obtain efficient and consistent gravity estimates from non-negative, heteroskedastic trade flows with many zero values that are not economically determined (see Yotov (2022) for an overview of gravity applications). Manning and Mullahy (2001) argue, however, that the P-PML estimator can yield very imprecise estimates if, additionally, the underlying data is heavy-tailed.[1]

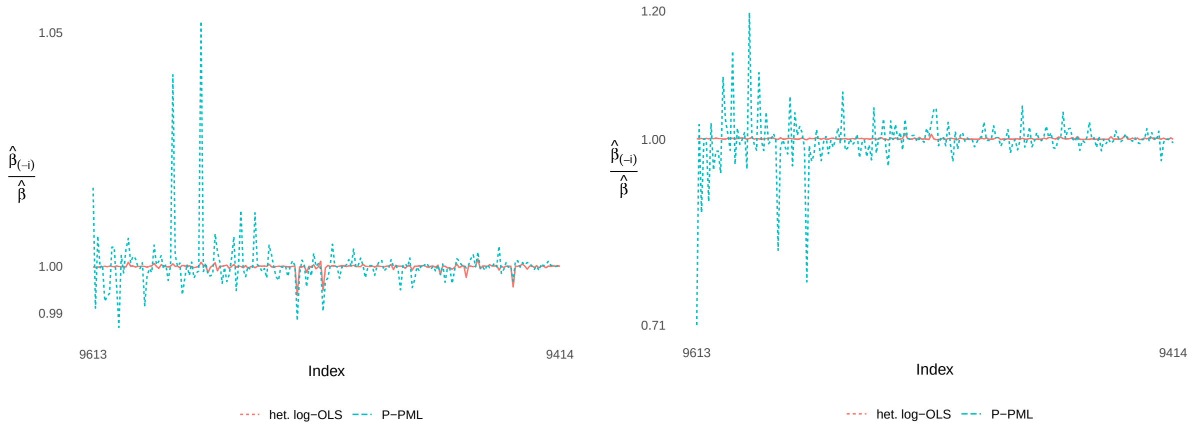
This paper evaluates the performance of gravity estimation methodologies in non-negative trade data with many (economically determined) zero values that is heteroskedastic, positively skewed and heavy-tailed. We confirm the finding of Manning and Mullahy (2001) that pseudo-maximum-likelihood estimators such as P-PML yield very imprecise estimates if trade data is heavy-tailed. Consequently, if zero trade values are not economically determined, we argue that the gravity equation should be estimated with heteroskedasticity-robust log-linear method (Manning, 1998). Heteroskedasticity-robust log-linear estimators remove the heavy-tail influence through a transformation of the data. They account for heteroskedasticity, explicitly modeling the impact of the variance function on coefficient estimates. If zero-trade values are economically determined, we argue that both the log-linear and PML estimators need to model these zero-trade values accordingly. We propose a heteroskedasticity-robust log-linear Tobit I-model as the most apt estimator for non-negative trade data with economically determined zero values that is heteroskedastic, positively skewed, and heavy-tailed.

We are not the first to expose the vulnerability of PML estimators to heavy-tailed data. Manning and Mullahy (2001) demonstrate the inefficiency of PML estimators when the dependent variable is heavy-tailed. In heavy-tailed data, large values take on a lot more weight relative to small values than in light-tailed data. The disproportionate size of large values in heavy-tailed data greatly influences the coefficient estimates of PML estimators, which do not log-linearize the data. Log-linear estimators do not suffer this problem, as transformation reduces the disproportional weight of large observations in the estimator. For instance, a heavy-tailed Pareto error distribution reduces to an exponential distribution after taking logs. Probability distributions that have thinner tails than an exponential distribution considered light-tailed distributions.

We intuitively demonstrate the differential effect of heavy-tailed gravity data on coefficient estimates of the P-PML (Silva and Tenreyro, 2006) and heteroskedasticity-robust log-linear OLS

[1] This paper refers heavy-tailed distributions as distributions of which the tail probabilities decay more slowly than those of any exponential distribution. Related definitions exist, such as distributions with the property that some of their moments are infinite. Also, a distribution can be defined as heavy-tailed if observations are i.i.d. according to a distribution in the domain of attraction of an α-stable law with index $\alpha < 2$.

(Manning and Mullahy, 2001) estimators in Figure 1. We replicate the naive gravity regression of Silva and Tenreyro (2006) applied to a non-zero cross-sectional bilateral trade database for 136 countries in 1990. We then estimate the same regression specification on a sample with one of the 200 largest observations (out of 9613 total observations) removed. We evaluate the variability of the Leave-One-Out (LOO) coefficient estimates for log distance and a Free Trade Agreement (FTA) dummy in Figure panels 1a and 1b respectively. It can be observed that the heteroskedasticity-robust log-linear coefficient estimates of the complete sample are almost equivalent to the LOO coefficient estimates. In the case of the P-PML estimator, on the other hand, it can be observed that LOO estimates can increase relative to the total sample estimates by more than 5% for the continuous distance variable. For the discrete FTA variable, we observe that solely omitting the largest observation reduces the LOO FTA coefficient by almost 30% relative to the total sample estimate. This behavior is unwarranted for a widely-used estimator in the trade literature and beyond.



(a) Log distance

(b) FTA

Figure 1: Ratio of Leave-One-Out coefficient estimate over total sample estimate for 200 largest observations.

Note: Results obtained from a replication of the Naive gravity regression applied to a non-zero cross-sectional bilateral trade database for 136 countries in 1990 with 9613 observations (Silva and Tenreyro, 2006). The database is sorted from small to large so that the largest trade observation in the dataset is indexed by the value 9613.

In the following sections, we uncover the roots of the vulnerability of P-PML to heavy-tailed data and propose alternative robust estimators. In section 2, we situate the paper relative to existing literature. We specify existing and novel gravity estimators section 3 and discuss their distinct properties. Section 4 showcases the (in)ability of these estimators to accurately recover the 'true' coefficients in a Monte Carlo exercise. We demonstrate the economic impact of the proposed alternative estimators on real-life gravity data in section 5. Section 6 concludes.

2 Background and Related Literature

We are not the first to scrutinize the performance of the P-PML estimator in the trade literature. The most contested feature of the P-PML estimator is its ability to handle zero trade values. Although P-PML allows estimating data with zero trade values, it does not explicitly model them. This makes P-PML susceptible to misspecification and sample selection bias, particularly if zero trade values are correlated with gravity equation regressors (Mnasri and Nechi, 2021). Martínez-Zarzoso (2013) extend the Monte Carlo simulation of (Silva and Tenreyro, 2006) to allow for many zero values and challenge the P-PML as the go-to gravity estimator, but the conclusions of this paper were refuted by sil (2008). Burger et al. (2009) propose to extend the Poisson framework to account for zeroes using a zero-inflated Poisson estimator. Santos Silva and Tenreyro (2011), however, demonstrate that the P-PML estimator can adequately handle a large number of zero values if these values are not economically determined.

If the number of zero values is economically determined, on the other hand, the P-PML estimator is shown to deliver biased estimates Head and Mayer (2014), Martin and Pham (2020), Mnasri and Nechi (2021). Alternative estimation procedures that account for economically determined zero trade values rely on a two-stage procedure. Helpman et al. (2008) derive a Heckman two-step estimator based on a log-linear bilateral trade relationship. The proposed methodology does not account for the possible existence of heteroskedasticity. Alternative two-step estimators that account for heteroskedasticity and economically determined zero trade values are the TS-MM moment estimator (Xiong and Chen, 2014), the heteroskedasticity-robust Tobit estimator (Martin, 2020), and the two-step generalization of existing PML estimators (Sukanuntathum, 2012).

The feature that received somewhat less attention in the trade literature is the efficiency of the P-PML estimator, especially in relation to heavy-tailed data. Martínez-Zarzoso (2013) proposed to consider the Feasible Generalized Least Squares (FGLS) as an efficient alternative to P-PML in light-tailed data. Burger et al. (2009) proposed to rely on the Negative Binomial PML (NB-PML) estimator as an efficient alternative for P-PML estimator in the presence of overdispersion.[2] While the traditional NB-PML estimator suffers from scale dependence (Head and Mayer, 2014), Bosquet and Boulhol (2014) provide a methodology that allows overcoming the scale dependence of the NB-PML estimator. We demonstrate below that neither estimator is sufficiently accurate in heavy-tailed data.

Manning and Mullahy (2001) discuss the trade-offs between PML estimators and log-linear estimators for healthcare expenditure data, which holds similar features to bilateral trade data. They focus on the robustness to heteroskedasticity and precision loss when the data is heavy-tailed, concluding that PML estimators can yield very imprecise estimates in heavy-tailed data. (Manning et al., 2005) propose to rely on a three-parameter generalized Gamma estimator as an alternative to PML estimators. They consider PML estimators to be more sensitive to data problems than (log-)OLS, as robust estimators typically sacrifice precision in favor of robustness Manning et al. (2005). Alternatively, Basu and Rathouz (2005) propose an extension of the PML estimators to simultaneously determine parameters of the link function and variance structure with the regression coefficients. Kurz (2017), on the other hand, propose the Tweedie (compound Poisson-Gamma) distribution as a viable alternative for the PML estimator, which allows for a large number of not-economically determined zeros. (Jones et al., 2016) discuss the

[2] We refer to overdispersion as the presence of greater variability in the data than what would be expected based on the statistical model, in this case, the Poisson model. Overdispersion does not necessarily imply heavy-tailed data.

varying performance and/or applicability of these alternative estimators.

3 Gravity estimators

Imagine bilateral export flows Y^* from country $i \in \{1, \ldots, I\}$ to country $j \in \{1, \ldots, J\}$ at time $t \in \{1, \ldots, T\}$ as a function of B (discrete and continuous) explanatory variables, gathered in the matrix $\boldsymbol{x}_{ijt}$ of dimension $IJT \times B$, and a disturbance term ϵ_{ijt}:

$$Y^*_{ijt} = e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}+\epsilon_{ijt}}. \tag{1}$$

As a researcher, we do not always observe these bilateral export flows. Some of the export flows can be unobserved or latent, resulting in zero trade flows in the dataset. We model the presence of zero trade flows using binary variable W, which takes the value of zero when we do not observe the bilateral trade flow at time t:

$$Y_{ijt} = W_{ijt}Y^*_{ijt}. \tag{2}$$

The binary variable W allows us to differentiate between observed trade flows Y and latent trade flows Y^*.

The estimation of equation (1) is burdened by the non-linear nature of the specification and the distribution of bilateral trade being non-negative with many (non-random) zero values, heteroskedastic, positively skewed, and heavy-tailed. We start by describing estimators that can handle such data without accounting for zero-trade values below. We categorize the proposed methodologies Carroll and Ruppert (1983) into two categories: methods based on data transformations and methods based on weighting. We discuss the ability of these transformation and/or weighting methodologies to deal with (non-random) zero values (see equation (2)) in the final subsection.

3.1 Data Transformation

One can transform the data so that one or multiple difficulties in estimating bilateral trade flows are avoided. We gather the possibilities of transformation (including the possibility of no transformation) into a single model

$$g(Y_{ijt}, \lambda) = g(e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}+\epsilon_{ijt}}, \lambda). \tag{3}$$

where the function $g(v, \lambda)$ is the power transformation family:

$$g(v, \lambda) = \begin{cases} (v^\lambda - 1)/\lambda & \text{if } \lambda \neq 0; \\ ln(v) & \text{if } \lambda = 0. \end{cases} \tag{4}$$

Choosing $\lambda = 0$ implies a log-transformation, and $\lambda = 1$ implies no transformation at all (Carroll and Ruppert, 1983). While alternative transformations are possible (such as the square-roots transformation ($\lambda = 0.5$)), they are less commonly relied on in applied work (Jones et al., 2016). They have, to our knowledge, not been applied in the gravity literature.

The log transformation linearizes the estimation problem and reduces the skewness and heavy-tailedness of the data Manning and Mullahy (2001). It can, however, not account for zero trade values and delivers inconsistent estimates under heteroskedasticity if not accounted for (Manning, 1998, Silva and Tenreyro, 2006). Manning (1998) indicate that there are important trade-offs between transforming or not in terms of precision and bias when estimating the equations, as commonly-used estimators on non-transformed heavy-tailed data can yield very imprecise estimates. Below, we discuss common estimators that specify both the mean and the variance of the estimation function and, as such, aim to increase efficiency of the estimator.

3.2 Data Weighting

Considering that the conditional variance might not be constant (i.e., heteroskedastic), one can specify the variance along with the mean and reweigh the estimation procedure to, for instance, reduce the influence of more variable observations. This course of action commonly relies on Generalized Linear Models (GLM) Nelder and Wedderburn (1972). Below, we mention three consistent estimator types for the gravity equation that differ in the treatment of the conditional variance function: the Pseudo-Maximum Likelihood estimators (PML), the Quasi-Generalized PML (QGPML) estimators, and the Quadratic Exponential PML (QEPML) estimators.

3.2.1 Pseudo-Maximum Likelihood (PML)

It can then be shown that if the transformed dependent variable $g(Y_{ijt}, \lambda)$ follows a distribution from the linear exponential family, consistent parameter estimates ($\hat{\beta}$) can be obtained using a Pseudo-Maximum Likelihood (PML) estimator. The PML estimator consists of a First-Order Condition (FOC) based only on the first, $E\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]$, and second, $V\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]$, moments of the underlying distributional assumption (McCullagh, 1989, Gourieroux et al., 1984a,b):

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \frac{g(Y_{ijt}, \lambda) - E\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]}{V\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]} \frac{\partial E\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]}{\partial \boldsymbol{\beta}} = 0. \tag{5}$$

It can be observed from this FOC that the weighting of the observations by the variance function ($V\left[\cdot\right]$) depends on the underlying distributional assumption. However, all that is needed for a consistent estimator based on the above FOC is the correct specification of the conditional mean. As such, the specification of the mean function is separated from that of the variance function. Buntin and Zaslavsky (2004) argues that if the mean function is correctly specified, the choice of the variance function is solely a question of efficiency. If the mean function is misspecified, the model does not fit the data equally well across the entire range of predicted values. An optimal fit in one part of the range implies at least slightly worse fit in another. In that case, the variance function affects the relative weighting of the goodness of fit in different parts of the data range. The variance function affects both the efficiency of estimation and the criterion of fit, and the choice of this function might reflect a compromise between the two (Buntin and Zaslavsky, 2004).[3]

We list the conditional variance and the resulting PML estimator linked to the Normal, Poisson, Gamma, Negative Binomial, and Lognormal distribution in Table 1. The PML estimator from the Normal family tends to put more weight on noisier observations, where $e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$ is large (Silva and Tenreyro, 2006), compared to the Poisson PML estimator. The PML estimator for the Gamma family, on the other hand, down weights these larger observations and gives too much weight to observations of lesser quality, where $e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$ is small (Silva and Tenreyro, 2006). Silva and Tenreyro (2006) argue that the P-PML estimator, which weighs all observations equally, is an adequate middle ground and proves to be the most robust estimator in practice. The weighting of the negative binomial estimator depends on the parameter a. As $a \to 0$, the estimator takes the form of the Poisson PML estimator, while for positive values of $a \to 1$, the estimator is very similar to the Gamma PML estimator. Despite this flexibility, the negative binomial estimator is seldom relied on. Specifying a consistent estimator for the dispersion parameter a is difficult due to the scale dependence of such estimator Head and Mayer (2014). As we discuss below, however, it is possible to retrieve a scale-independent consistent estimator for a (Bosquet and Boulhol, 2014).

Lastly, the Lognormal estimator relies on a log transformation of the data, resulting in a FOC that down weights larger observations. In the absence of heteroskedasticity, the interpretation of the coefficients of a lognormal regression can be compared straightforwardly with those of non-transformed estimators. In the presence of heteroskedasticity, however, the coefficient estimates of the lognormal estimator need to be adapted accordingly. The adaptation follows straightforwardly from the specification of the conditional mean in Table 1 (Manning, 1998):

$$\frac{\partial E\left[Y_{ijt} \mid \boldsymbol{x}_{ijt}\right]}{\partial \boldsymbol{\beta}} = E[\widehat{Y_{ijt} \mid \boldsymbol{x}}_{ijt}]\left[\hat{\boldsymbol{\beta}} + 0.5\frac{\partial \hat{\sigma}^2}{\partial \boldsymbol{\beta}}\right]. \tag{6}$$

[3] In case of varying trade elasticities across markets, for instance, it has been argued that the P-PML estimator provides a weighted average of these elasticities, which is more preferable from an economic point of view than the unweighted average provided by NLLS (Breinlich et al., 2022).

Table 1: Specification of the PML estimator for different linear exponential families.

λ	**Family**	$E\left[Y_{ijt}\vert\boldsymbol{x}_{ijt}\right]$	$V\left[Y_{ijt}\vert\boldsymbol{x}_{ijt}\right]$	$E\left[g(Y_{ijt},\lambda)\vert\boldsymbol{x}_{ijt}\right]$	$V\left[g(Y_{ijt},\lambda)\vert\boldsymbol{x}_{ijt}\right]$	FOC
1	Normal	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	σ^2	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	σ^2	$\frac{\boldsymbol{x}_{ijt}^T e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{\sigma^2}$
1	Poisson	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\boldsymbol{x}_{ijt}^T\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)$
1	Gamma	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$ke^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$ke^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{\boldsymbol{x}_{ijt}^T e^{-\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{k}$
1	Negative Binomial type 2	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}+ae^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}+ae^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{-\boldsymbol{x}_{ijt}^T\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{1+ae^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}$
0	Log-normal	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}+\frac{\sigma^2}{2}}$	$\left(e^{\sigma^2}-1\right)e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}+\sigma^2}$	$\boldsymbol{x}_{ijt}\boldsymbol{\beta}$	σ^2	$\frac{\boldsymbol{x}_{ijt}^T(lnY_{ijt}-\boldsymbol{x}_{ijt}\boldsymbol{\beta})}{\sigma^2}$

3.2.2 Quasi-Generalized Pseudo-Maximum Likelihood (QGPML)

None of the previous untransformed PML estimators is uniformly better than the others, i.e., with a smaller asymptotic covariance matrix for all possible distributions of the disturbance term (Gourieroux et al., 1984b). If $e^{\epsilon_{ijt}}$ equals one, the PML estimator based on the Poisson family is asymptotically efficient and better than the other estimators. If $e^{\epsilon_{ijt}}$ has a log-gamma distribution, the PML estimator based on the negative binomial family is asymptotically efficient and better than the others.

However, it is possible to build uniformly better estimators than the previously specified PML estimators (Gourieroux et al., 1984a,b). Quasi-Generalized PML estimators exploit the second moment of the conditional distribution to construct more efficient estimators than the PML estimators. Gourieroux et al. (1984a) proposed to rely on the variance specification of Poisson models, $V\left[Y_{ijt}|\boldsymbol{x}_{ijt}\right] = e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta_1 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$, where the variance of the disturbance term is defined here as $V\left[\epsilon_{ijt}\right] \equiv \theta_1$ to build three QGPML estimators.[4]

These QGPML estimators require a consistent estimator $\hat{\eta}^2$ of η^2. As Bosquet and Boulhol (2014) note, the original estimator for η^2 proposed by (Gourieroux et al., 1984a) suffered from scale-dependence. Therefore, we follow Bosquet and Boulhol (2014) in the two-step approach to obtain a scale-independent estimate of η^2. In the first step, we rely on any previously mentioned PML estimator to obtain a consistent estimate for $\boldsymbol{\beta}$. In a second step, we estimate $\hat{\eta}^2 = \frac{\hat{b}}{\hat{a}}$ from the following regression:

$$\frac{\left(Y_{ijt} - e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}\right)^2}{e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}} = a + b e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}} + \varepsilon_{ijt}. \tag{7}$$

The FOCs related to each specific family of QGPML estimators based on the consistent estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\eta}^2$ are provided in Table 2 in the column titled QGPML. Notice that as $\hat{\eta}^2 \to 0$, the respective FOC of the QGPML estimators converge to the P-PML estimator. Notice also how the QGPML estimators of the Normal, Gamma, and Negative binomial family are almost equivalent. We, therefore, do not expect large performance differentials between the QGPML estimators of different families.

[4] By the law of total variance, $V\left[Y_{ijt}|\boldsymbol{x}_{ijt}\right] = E\left[V\left[Y_{ijt}|\boldsymbol{x}_{ijt}, \epsilon_{ijt}\right]\right] + V\left[E\left[Y_{ijt}|\boldsymbol{x}_{ijt}, \epsilon_{ijt}\right]\right]$, where it is assumed there is a constant term in $\boldsymbol{x}_{ijt}\boldsymbol{\beta}$ such that $E\left[e^{\epsilon_{ijt}}\right] = 1$.

Table 2: Specification of the QGPML estimator for different linear exponential families.

λ	**Family**	$E\left[Y_{ijt} \vert \boldsymbol{x}_{ijt}\right]$	$V\left[Y_{ijt} \vert \boldsymbol{x}_{ijt}\right]$	$E\left[g(Y_{ijt},\lambda) \vert \boldsymbol{x}_{ijt}\right]$	$V\left[g(\widehat{Y_{ijt},\lambda}) \vert \boldsymbol{x}_{ijt}\right]$	FOC
1	Normal	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \eta^2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}} + \hat{\eta}^2 e^{2\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}$	$\frac{\boldsymbol{x}_{ijt}^T e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}(1+\hat{\eta}^2 e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}})}$
1	Gamma	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \eta^2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$(e^{-\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}} + \hat{\eta}^2) e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{\boldsymbol{x}_{ijt}^T e^{-\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{e^{-\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}(1+\hat{\eta}^2 e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}})}$
1	Negative Binomial Type 2	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \eta^2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \hat{\eta}^2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{\boldsymbol{x}_{ijt}^T\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{1+\hat{\eta}^2 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}$

Notes: $\hat{\boldsymbol{\beta}}$ and $\hat{\eta}$ indicate consistent estimates of the parameter vector $\boldsymbol{\beta}$ and the disturbance's standard deviation η^2 obtained in a preceding estimation stage.

3.2.3 Quadratic Exponential Pseudo-Maximum Likelihood (QEPML)

The parameters of the mean ($\boldsymbol{\beta}$) and the variance function can be estimated simultaneously by using the PML approach based on quadratic exponential families. The QEPML estimator relies on a FOC of the form (Gourieroux et al., 1984a):

$$\sum_i \sum_j \sum_t \left[-\frac{1}{2} log \left(V\left[g(Y_{ijt},\lambda) | \boldsymbol{x}_{ijt} \right] \right) - \frac{1}{2} \frac{\left(g(Y_{ijt},\lambda) - E\left[g(Y_{ijt},\lambda) | \boldsymbol{x}_{ijt} \right] \right)^2}{V\left[g(Y_{ijt},\lambda) | \boldsymbol{x}_{ijt} \right]} \right] = 0, \quad (8)$$

where the variance function can be specified freely, for instance, as a constant variance (CV), a power function of the mean (PV), and a quadratic function of the mean (QV) (Basu and Rathouz, 2005). We list the specification of the mean and variance function for a QEPML estimator under both the normal and lognormal distributional assumption in Table 3.

Table 3: Specification of the PML estimator for quadratic exponential families.

λ	**Family**	$E\left[Y_{ijt} \vert \boldsymbol{x}_{ijt}\right]$	$V\left[Y_{ijt} \vert \boldsymbol{x}_{ijt}\right]$	$E\left[g(Y_{ijt},\lambda) \vert \boldsymbol{x}_{ijt}\right]$	$V\left[g(Y_{ijt},\lambda) \vert \boldsymbol{x}_{ijt}\right]$
1	N	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta 2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta 2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$
1	N	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\theta 2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\theta 2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$
0	LN	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	σ^2	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \vert \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{\sigma^2}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}+1\right)$
0	LN	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta 2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \vert \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{V[Y_{ijt} \vert \boldsymbol{x}_{ijt}]}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}+1\right)$
0	LN	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta 1 e^{\theta 2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \vert \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{V[Y_{ijt} \vert \boldsymbol{x}_{ijt}]}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}+1\right)$

3.3 Economically determined zero trade flows

In our discussion on the gravity estimators above, we did not yet touch on their ability to handle zero values. The robustness of gravity estimators to zero-trade flows has already been identified as an important topic in the literature (see, for instance, Martin and Pham (2020), Martin (2020), Mnasri and Nechi (2021)). This robustness will depend on the origin of the existence of zero-valued data. Following Head and Mayer (2014), we focus on the appearance of zero-values due to the left-censoring of the data.

Left-censoring of the data entails that the binary variable W takes the value one if the trade flows are larger than a certain cutoff, which is defined country-pair specific: $W_{ijt} = \mathbb{I}\left[Y^*_{ijt} > Y^L_{ij}\right]$. This sample selection specification can be motivated by, for instance, the presence of fixed costs to export. When zero-trade costs are present, log-linearization results in a truncated sample and biased estimates. PML and QGPML on non-transformed data can accommodate zero trade values. When these zero trade values are economically determined, i.e., the result of sample selection, non-transformed PML, and QGPML estimators will be biased if they do not explicitly model the selection specification. In this paper, provided the performance of QEPML estimators in the presence of heavy tails, we focus on the specification of sample selection for (log-linear) QEPML estimators. We model the left-censoring with a Tobit I-specification, resulting in a log-likelihood specification of the following form:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \sigma) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{T}&\left(\frac{1}{\sigma}\varphi\left(\frac{g(Y_{ijt}, \lambda) - E\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right]}{\sigma}\right)\right)^{W_{ijt}} \\ &\times\left(1 - \Phi\left(\frac{E\left[g(Y_{ijt}, \lambda)|\boldsymbol{x}_{ijt}\right] - Y^L_{ij}}{\sigma}\right)\right)^{1-W_{ijt}}\end{aligned} \tag{9}$$

The specification of the conditional mean and variance, then, remains equivalent to the specifications provided in section 3.2.3.

4 Monte Carlo Evidence

Our Monte Carlo evidence relies on a replication and extension of the original Monte Carlo exercise of Silva and Tenreyro (2006). We simulate the multiplicative model

$$E\left[Y_i \mid x\right] = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right), \qquad i = 1, \ldots, 1000, \tag{10}$$

where x_{1i} is drawn from a normal distribution with mean zero and standard deviation equal to one, and x_2 is a binary dummy variable that equals 1 with a probability of 0.4 The two covariates are independent, and a new set of observations of all variables is generated in each replication using $\beta_0 = 0, \beta_1 = \beta_2 = 1$. Data on y are generated as

$$Y_i = \mathbb{I}\left[e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\eta_i \geq Q_p(e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\eta_i)\right] e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\eta_i, \tag{11}$$

where η_i is a log normal random variable with mean 1 and variance σ_i^2. $\mathbb{I}$ is the indicator function, and $Q_p(\cdot)$ is the quantile function evaluated at the $p - th$ percentile.

To assess the performance of the estimators under different patterns of heteroskedasticity, we follow Silva and Tenreyro (2006) in considering the four following specifications of σ_i^2:

Case 1: $\sigma_i^2 = e^{-2\boldsymbol{x}_i\boldsymbol{\beta}}; V\left[Y_i \mid x\right] = 1;$

Case 2: $\sigma_i^2 = e^{-\boldsymbol{x}_i\boldsymbol{\beta}}; V\left[Y_i \mid x\right] = e^{-\boldsymbol{x}_i\boldsymbol{\beta}};$

Case 3: $\sigma_i^2 = 1; V\left[Y_i \mid x\right] = e^{2\boldsymbol{x}_i\boldsymbol{\beta}};$

Case 4: $\sigma_i^2 = e^{-\boldsymbol{x}_i\boldsymbol{\beta}} + \exp(x_{2i})\,; V[Y_i \mid x] = e^{\boldsymbol{x}_i\boldsymbol{\beta}} + \exp(x_{2i})\, e^{2\boldsymbol{x}_i\boldsymbol{\beta}}$.

We perform 1.000 Monte Carlo simulations and estimate for each simulation the model using estimators that differ in the underlying distributional assumption (Normal (N), Poisson (P), Gamma (G), Negative Binomial type 2 ($NB2$), and Log-normal (LN)), in the estimator type (PML, QGPML, and QEPML), and the variance function specification (constant variance (CV), power variance (PV), and quadratic variance (QV)). For space constraints, we limit our results to the QGPML estimator under the Gamma distribution, as results of the QGPML-estimator for the Normal and Negative Binomial type-2 distribution are very similar.

4.1 No sample selection

We start with a narrow MC replication of Silva and Tenreyro (2006). This entails no selection equation, evaluating the quantile function at the zeroth percentile ($Q_0(\cdot)$) in equation (11). The narrow MC replication exercise results are shown in Table 4. The results of Silva and Tenreyro (2006) hold, that the P-PML estimator showcases good performance in all cases. However, the addition of new alternative estimators demonstrates that the performance of the P-PML estimator can be matched or even improved upon. The performance of the heteroskedasticity-robust OLS estimator is close to the performance of the P-PML estimator in low-variance cases (Case 1 and 2). It improves the performance of the P-PML estimator in high variance cases (Case 3 and 4). Similarly, the G-QGPML estimator matches the performance of the P-PML estimator in low-variance cases (Case 1 and 2) and improves on the performance of the P-PML estimator in high-variance cases (Case 3 and 4). A power variance specification of the suggested QEPML estimators delivers the best results for both the Normal and Lognormal assumption, with better performance for the Lognormal distribution. Compared to P-PML, the LN-QEPML-PV estimator delivers consistently lower bias and lower variability of the coefficient estimates in Cases 1, 2, and 3. However, the QEPML estimators are not robust to a misspecification of the variance function, as shown by the results in Case 4 of the Monte Carlo exercise.

To evaluate the performance of the estimators in the presence of heavy tails, we extend the Monte Carlo to the case where an explanatory variable or the error term has a heavy tail. First, we assume that the continuous variable x_{1i} is drawn from a multiplicative Normal-Exponential distribution with mean zero and standard deviation equal to one and an exponential rate parameter of 1.1. As such, the explanatory and dependent variables have a heavy right tail. The results of this wide MC replication exercise are shown in Table 5. It can be observed that, relative to the narrow MC exercise, the good performance of the P-PML estimator remains. For the cases with larger variance, Cases 3 and 4, it can be observed that the P-PML estimator becomes imprecisely estimated relative to the G-QGPML and OLS-het. alternatives. These results are in line with both (Manning and Mullahy, 2001) and Head and Mayer (2014). (Head and Mayer, 2014) find sizable biases for the P-PML estimator due to the high coefficient of variation in their simulation, which is calibrated on real data.

Second, we extend Case 3 of the original Monte Carlo exercise to the case where the error term is heavy-tailed. We assume the error term η_i follows a Multiplicative Lognormal-Pareto distribution with mean zero and standard deviation equal to one and a Pareto shape parameter of

1.1. The results of this exercise are available in Table 6, demonstrating the impreciseness of the P-PML estimator when the data becomes heavy-tailed. The G-QPML performs relatively better. The performance of the OLS-het. estimator is remarkable as its distributional assumption for the heteroskedasticity correction is not satisfied.

4.2 Sample selection

We move on to a MC replication of Silva and Tenreyro (2006) with sample selection. We specify a selection equation that results in censoring 50% of the data observations, evaluating the quantile function at the zeroth percentile ($Q_0(\cdot)$) in equation (11). To ensure a comparable sample size to the MC without sample selection, we increase the number of observations to 2,000. In a first replication exercise, we generate the data sample assuming the continuous variable x_{1i} is drawn from a multiplicative Normal-Exponential distribution with mean zero and standard deviation equal to one and an exponential rate parameter of 1.1. The results are displayed in Table 7. We observe that economically determined trade values bias the PML and GQPML estimators that do not model the zero trade values explicitly. The heteroskedasticity-robust Tobit QEPML estimators, on the other hand, deliver consistent estimates as long as the variance function is correctly specified.

In Table 8, we display the results of a MC exercise where, instead of an explanatory variable, the error term has a heavy right tail originating from a right-Pareto Lognormal distribution with shape parameter equal to 1.1. Similar to the MC exercise without economically determined zeroes, we observe that the P-PPML estimator becomes very inefficient. On top of that, the estimator is now also biased due to the existence of economically determined zero values. Also here, the log-linear heteroskedasticity-robust Tobit QEPML estimators, delivers consistent estimates despite the presence of a heavy-tailed error term.

Table 4: Narrow replication results Log of Gravity Table 1.

	β_1		β_2		θ_1^2		θ_2^2	
	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
				Case 1: $V[y_i\|x] = 1$				
OLS	0.38430	0.040	0.34507	0.051				
OLS-het.	-0.00446	0.056	-0.00491	0.065				
P-PML	-0.00025	0.016	-0.00495	0.029				
N-PML	-0.00026	0.007	-0.00183	0.018				
G-PML	0.01666	0.063	0.00078	0.076				
G-QGPML	-0.00025	0.016	-0.00495	0.029	0.00000	0.000		
N-QEPML-PV	0.00095	0.008	-0.00127	0.020	0.12931	0.017	0.14849	0.233
N-QEPML-QV	-0.04672	0.066	-0.05363	0.113	0.30872	0.123	-0.01488	0.007
LN-QEPML-CV	-0.00032	0.007	-0.00107	0.016	0.35550	0.018		
LN-QEPML-PV	-0.00033	0.007	-0.00105	0.016	0.12647	0.013	-0.00403	0.052
LN-QEPML-QV	0.06303	0.022	0.07026	0.034	0.22945	0.036	-0.00754	0.006
				Case 2: $V[y_i\|x] = \mu(x_i\beta)$				
OLS	0.20828	0.028	0.18997	0.048				
OLS-het.	0.00352	0.037	-0.00637	0.058				
P-PML	0.00066	0.019	-0.00656	0.042				
N-PML	-0.00170	0.034	-0.00308	0.063				
G-PML	0.00765	0.039	-0.00632	0.060				
G-QGPML	0.00145	0.019	-0.00617	0.043	0.00937	0.011		
N-QEPML-PV	0.00535	0.023	-0.01581	0.060	0.35829	0.041	1.06264	0.161
N-QEPML-QV	0.00733	0.032	-0.01060	0.066	0.34065	0.065	0.00976	0.024
LN-QEPML-CV	-0.14600	0.032	-0.16607	0.041	0.88034	0.062		
LN-QEPML-PV	0.00022	0.018	-0.00279	0.036	0.35387	0.027	0.99901	0.057
LN-QEPML-QV	-0.00153	0.019	-0.00578	0.037	0.35843	0.034	-0.00350	0.012
				Case 3: $V[y_i\|x] = \mu(x_i\beta)^2$				
OLS	0.00083	0.024	-0.00838	0.057				
OLS-het.	0.00232	0.028	-0.01188	0.063				
P-PML	-0.00125	0.069	-0.01267	0.110				
N-PML	0.04807	0.400	0.05020	0.562				
G-PML	0.00280	0.028	-0.01282	0.063				
G-QGPML	0.01096	0.051	-0.00732	0.084	1.90766	4.903		
N-QEPML-PV	0.00498	0.035	-0.02609	0.106	0.96999	0.135	2.00217	0.101
N-QEPML-QV	0.00977	0.049	-0.01786	0.114	0.00571	0.036	0.95344	0.141
LN-QEPML-CV	-0.51302	0.025	-0.51191	0.038	1.73170	0.124		
LN-QEPML-PV	0.00136	0.027	-0.00781	0.058	0.99115	0.071	2.00200	0.053
LN-QEPML-QV	0.00103	0.028	-0.00941	0.059	0.00030	0.012	0.98836	0.078
				Case 4: $V[y_i\|x] = \mu(x_i\beta) + exp(x_{2i})\mu(x_i\beta)^2$				
OLS	0.13227	0.037	-0.13681	0.078				
OLS-het.	0.00272	0.050	-0.01669	0.096				
P-PML	-0.00319	0.093	-0.02165	0.147				
N-PML	0.15821	0.998	0.10634	1.162				
G-PML	0.00974	0.052	-0.02016	0.100				
G-QGPML	0.01837	0.072	-0.01463	0.123	7.29868	60.539		
N-QEPML-PV	-0.04397	0.073	0.14763	0.186	2.12409	0.439	1.86182	0.167
N-QEPML-QV	-0.02125	0.102	0.17637	0.227	0.20497	0.294	1.81743	0.581
LN-QEPML-CV	-0.47125	0.025	-0.67644	0.045	2.25548	0.167		
LN-QEPML-PV	0.04277	0.044	-0.23312	0.084	2.28934	0.215	1.77819	0.078
LN-QEPML-QV	0.03598	0.043	-0.22763	0.080	0.27527	0.121	1.86095	0.244

Table 5: Wide replication results Log of Gravity Table 1 with heavy-tailed independent variable.

	β_1		β_2		θ_1^2		θ_2^2	
	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
Case 1: $V[y_i\|x] = 1$								
OLS	0.16960	0.023	0.18865	0.041				
OLS-het.	-0.01238	0.029	-0.02150	0.048				
P-PML	0.00000	0.002	-0.00001	0.006				
N-PML	0.49327	4.933	-4.34125	43.415				
G-PML	0.00447	0.032	0.00197	0.058				
G-QGPML	0.00000	0.002	-0.00001	0.006	0.00000	0.000		
N-QEPML-PV	-0.00003	0.001	0.00033	0.002	0.00398	0.002	0.05541	0.119
N-QEPML-QV	0.00169	0.010	-0.01215	0.037	0.03306	0.021	-0.00027	0.000
LN-QEPML-CV	-0.05861	0.188	-0.03701	0.126	0.37944	1.087		
LN-QEPML-PV	-0.00005	0.001	0.00028	0.002	0.00380	0.002	-0.00654	0.038
LN-QEPML-QV	-0.00229	0.008	-0.00971	0.039	0.02776	0.010	-0.00025	0.000
Case 2: $V[y_i\|x] = \mu(x_i\beta)$								
OLS	0.11690	0.017	0.12672	0.041				
OLS-het.	0.00172	0.021	-0.00433	0.046				
P-PML	-0.00028	0.004	0.00108	0.018				
N-PML	-0.00112	0.009	-0.35055	3.534				
G-PML	0.00262	0.022	-0.00071	0.050				
G-QGPML	-0.00014	0.004	0.00116	0.018	0.00019	0.001		
N-QEPML-PV	0.00019	0.004	0.00273	0.022	0.06049	0.020	1.03082	0.103
N-QEPML-QV	0.00095	0.006	-0.00104	0.031	0.05790	0.019	0.00049	0.002
LN-QEPML-CV	-0.04627	0.142	-0.03593	0.093	0.40984	0.684		
LN-QEPML-PV	-0.00018	0.004	0.00049	0.017	0.05906	0.019	0.99701	0.037
LN-QEPML-QV	-0.00260	0.007	-0.01194	0.032	0.06149	0.019	-0.00006	0.001
Case 3: $V[y_i\|x] = \mu(x_i\beta)^2$								
OLS	0.00194	0.019	-0.00434	0.053				
OLS-het.	0.00285	0.023	-0.00267	0.067				
P-PML	-0.02400	0.083	0.01006	0.296				
N-PML	6.14853	26.242	-27.12316	121.872				
G-PML	0.00295	0.024	-0.00259	0.068				
G-QGPML	0.00354	0.047	0.00230	0.141	0.16937	0.763		
N-QEPML-PV	0.00174	0.028	-0.00194	0.127	0.97523	0.176	2.00627	0.079
N-QEPML-QV	0.01212	0.045	0.01063	0.136	0.00311	0.013	0.93266	0.131
LN-QEPML-CV	-0.59758	0.094	-0.54688	0.066	3.30863	1.909		
LN-QEPML-PV	0.00297	0.024	-0.00301	0.057	1.00635	0.089	2.00362	0.045
LN-QEPML-QV	0.00225	0.022	-0.00525	0.059	-0.00002	0.004	1.00157	0.074
Case 4: $V[y_i\|x] = \mu(x_i\beta) + exp(x_{2i})\mu(x_i\beta)^2$								
OLS	0.07008	0.025	-0.20394	0.071				
OLS-het.	0.00356	0.035	-0.00483	0.101				
P-PML	-0.03362	0.104	-0.00583	0.382				
N-PML	3.28311	19.795	-15.06773	99.419				
G-PML	0.00548	0.038	-0.00456	0.104				
G-QGPML	0.00369	0.063	-0.01341	0.171	0.16744	0.610		
N-QEPML-PV	-0.03253	0.049	0.23177	0.197	1.72458	0.464	1.93133	0.122
N-QEPML-QV	-0.00479	0.075	0.27416	0.221	0.04219	0.068	1.56718	0.374
LN-QEPML-CV	-0.55589	0.084	-0.70973	0.066	3.99727	1.983		
LN-QEPML-PV	0.02583	0.036	-0.25458	0.077	1.83121	0.225	1.88586	0.059
LN-QEPML-QV	0.02058	0.031	-0.26340	0.074	0.05102	0.031	1.73957	0.195

Table 6: Wide replication results Log of Gravity Table 1 with heavy-tailed error term.

	β_1		β_2		θ_1^2		θ_2^2	
	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
				Case 3: $V[y_i\|x] = \mu(x_i\beta)^2$				
OLS	0.00174	0.042	0.00168	0.076				
OLS-het.	-0.00126	0.080	-0.00619	0.142				
P-PML	-0.08209	0.301	-0.09947	0.630				
N-PML	1.22025	9.462	-0.48490	6.880				
G-PML	-0.03072	0.227	-0.07784	0.492				
G-QGPML	-0.04459	0.318	-0.08214	0.501	0.05765	0.107		
N-QEPML-PV	-0.12667	0.293	-0.16729	1.037	351.47806	2285.368	2.24643	1.049
N-QEPML-QV	0.08444	0.557	-0.08505	1.001	-417.03342	3933.090	666.05699	5044.765
LN-QEPML-CV	-0.55308	0.025	-0.55340	0.040	1.90808	0.446		
LN-QEPML-PV	-0.00391	0.071	-0.00361	0.094	5.18590	1.083	1.98124	0.111
LN-QEPML-QV	-0.00410	0.058	-0.00630	0.083	0.03805	0.097	5.01343	0.764

Table 7: Wide replication results Log of Gravity Table 1 with heavy-tailed independent variable and economically determined zero values.

	β_1		β_2		θ_1^2		θ_2^2	
	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
				Case 1: $V[y_i\|x] = 1$				
OLS	-0.09497	0.014	-0.12204	0.022				
OLS-het.	-0.13367	0.022	-0.17335	0.033				
P-PML	0.03620	0.014	0.07690	0.028				
N-PML	27.63249	206.863	-240.28046	1803.953				
G-PML	0.54232	0.185	0.58874	0.227				
G-QGPML	0.06000	0.025	0.10886	0.031	0.00179	0.002		
N-QEPML-PV	-0.01827	0.050	-0.02708	0.121	1.54977	1.052	0.03337	0.534
N-QEPML-QV	5.37159	36.370	52.80498	567.192	NaN		NaN	
LN-QEPML-CV	-0.09564	0.247	-0.02516	0.104	24.83569	84.077		
LN-QEPML-PV	0.05364	0.161	0.03826	0.229	0.62719	0.952	0.38191	0.865
LN-QEPML-QV	-0.28559	0.620	-0.43625	0.877	NaN		NaN	
				Case 2: $V[y_i\|x] = \mu(x_i\beta)$				
OLS	-0.12992	0.013	-0.16611	0.028				
OLS-het.	-0.16120	0.016	-0.20600	0.031				
P-PML	0.03217	0.012	0.06846	0.026				
N-PML	4.60059	19.018	-37.33633	137.243				
G-PML	0.60034	0.141	0.64785	0.169				
G-QGPML	0.04817	0.020	0.09088	0.029	0.00118	0.001		
N-QEPML-PV	0.00382	0.004	0.00685	0.013	2.22507	0.484	0.76528	0.069
N-QEPML-QV	7.08151	56.390	15.48976	88.620	NaN		NaN	
LN-QEPML-CV	-0.10165	0.248	-0.03592	0.114	30.03869	88.773		
LN-QEPML-PV	0.01981	0.092	0.02186	0.176	0.61062	0.845	1.15297	0.446
LN-QEPML-QV	-0.19953	0.284	-1.35558	1.606	NaN		NaN	
				Case 3: $V[y_i\|x] = \mu(x_i\beta)^2$				
OLS	-0.25290	0.024	-0.29659	0.058				
OLS-het.	-0.17795	0.028	-0.22278	0.067				
P-PML	0.01742	0.073	0.03438	0.283				
N-PML	15.48084	54.966	-87.29489	302.338				
G-PML	0.51370	0.064	0.55450	0.112				
G-QGPML	0.09912	0.062	0.13250	0.144	0.06365	0.171		
N-QEPML-PV	0.12568	0.038	0.14515	0.142	6.47325	1.958	1.54611	0.082
N-QEPML-QV	0.16006	0.077	0.21565	0.165	7.14477	3.341	0.89710	0.796
LN-QEPML-CV	-0.39671	0.137	-0.27433	0.118	188.98813	156.361		
LN-QEPML-PV	0.01422	0.055	0.02365	0.101	0.83318	0.682	2.04154	0.191
LN-QEPML-QV	0.00408	0.032	0.00818	0.098	-0.09728	0.866	1.02019	0.160
				Case 4: $V[y_i\|x] = \mu(x_i\beta) + exp(x_{2i})\mu(x_i\beta)^2$				
OLS	-0.32044	0.028	-0.38245	0.069				
OLS-het.	-0.22149	0.035	-0.16905	0.085				
P-PML	0.00451	0.092	0.00329	0.365				
N-PML	11.99932	36.908	-88.24480	258.467				
G-PML	0.26847	0.073	0.33268	0.123				
G-QGPML	0.06736	0.058	0.09449	0.168	0.08661	0.295		
N-QEPML-PV	0.17730	0.063	0.64962	0.355	34.54487	25.675	1.37929	0.107
N-QEPML-QV	0.21356	0.185	0.73276	0.342	25.94088	18.878	4.15090	9.056
LN-QEPML-CV	-0.41571	0.117	-0.53519	0.104	218.20895	208.926		
LN-QEPML-PV	0.00548	0.040	-0.18950	0.086	2.00259	0.487	1.98045	0.069
LN-QEPML-QV	0.01209	0.024	-0.18634	0.062	0.54741	0.949	1.80881	0.185

Table 8: Wide replication results Log of Gravity Table 1 with heavy-tailed error term and ecnomically determined zero values

	β_1		β_2		θ_1^2		θ_2^2	
	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
					Case 3: $V[y_i\|x] = \mu(x_i\beta)^2$			
OLS	-0.56646	0.039	-0.57390	0.071				
OLS-het.	-0.43722	0.071	-0.44803	0.136				
P-PML	0.00147	0.301	-0.08302	0.728				
N-PML	8.50882	48.903	-9.83361	87.519				
G-PML	0.13669	0.227	0.11726	0.518				
G-QGPML	0.08627	0.314	0.02801	0.591	0.02800	0.034		
N-QEPML-PV	-0.87995	0.491	-0.89222	0.432	1038370.189	8704347.602	1758.149	1484.791
N-QEPML-QV	-0.12267	0.667	-0.11240	1.708	8918093.198	63643681.162	37484087.758	128430459.403
LN-QEPML-CV	-0.40329	0.027	-0.39099	0.046	91.353	16.299		
LN-QEPML-PV	-0.03595	0.054	-0.00082	0.335	24.968	12.099	1.659	0.152
LN-QEPML-QV	-0.04669	0.052	-0.04530	0.076	59.595	28.310	4.614	1.330

5 Real-life Application

The favorable performance of the heteroskedasticity-robust log-linear models in the MC exercise impels us to evaluate the performance of these estimators on real-life data. We rely on the cross-sectional bilateral trade database for 136 countries in 1990 provided by Silva and Tenreyro (2006).

In the first exercise, we replicate the naive gravity analysis of Silva and Tenreyro (2006) without zero trade values, of which the results are displayed in Table 9. We observe large difference in point estimates between the different estimators, contrary to the expectations formed during the MC exercise. The point estimate for the Free-trade agreement dummy, for instance, attains a value of 0.8 for the OLS-het estimator, 0.18 for the P-PML estimator, , -0.001 for the G-QGPML estimator, and 0.13 for the LN-QEPML-PV estimator. The size of the θ_1 parameter estimated by the LN-QEPML-PV estimator indicates the presence of a large variability in the data. Based on the MC-exercise, this means the values reported by the P-PML estimator are likely to be imprecisely estimated. The large difference between the heteroskedasticity robust log-linear estimation (OLS-het.) and the QEPML-estimators which specify a variance function points to a possible misspecifiation of the variance function for the QEPML estimators.

Augmenting our estimation procedure with country Fixed Effects, we estimate a structural gravity equation. The results of this specification are displayed in Table 10. Also in this case, we find large differences in the point estimates between different estimators. We can also observe that the Fixed effects capture a lot of variation, as the estimated θ_1 parameter drops from 4.5 to 1.5 for the LN-QEPML-PV estimator.

The results on data with zero-trade values barely differs from results estimated on data with the zero trade values included. We present the results in Table 11 for the naive and Table 12 for the structural gravity equation. Both the PML estimators, which do not model the zero trade flows, and the QEPML estimators, which explicitly model the zero trade flows, we observe no significant differences in coefficient estimates from the non-zero trade sample. Therefore, zero trade flows are not likely economically determined in this dataset.

The results from the naive and structural gravity equation for data with both zero and non-zero trade values point to data features in line with Cases 2, 3, and 4 of the MC exercise with heavy-tailed variation in the dependent variable. The large difference between the heteroskedasticity robust log-linear estimation (OLS-het.) and the QEPML-estimators which specify a variance function points to a possible misspecifiation of the variance function for the QEPML estimators, in line with MC case 4. In this case, we demonstrated the favorable performance of the OLS-het. estimator relative to the P-PML estimator, especially so in heavy-tailed data.

Table 9: Narrow replication results Log of Gravity Table 3 without zero trade values.

Dependent Variable	OLS-het	P-PML	G-QGPML	N-QEPML-PV	N-QEPML-QV	LN-QEPML-PV	LN-QEPML-QV
Constant	-25.044	-31.530	-27.987	-41.924	-36.176	-32.199	-31.070
Log exporter's GDP	0.623	0.721	0.728	0.609	0.594	0.726	0.698
Log importer's GDP	0.673	0.732	0.735	0.668	0.671	0.615	0.591
Log exporter's GDP per capita	0.243	0.154	0.189	0.219	0.208	0.137	0.135
Log importer's GDP per capita	0.097	0.133	0.168	0.270	0.246	0.083	0.083
Log distance	-0.940	-0.776	-0.816	-0.926	-0.818	-0.834	-0.801
Contiguity dummy	0.265	0.202	0.180	0.526	0.397	0.003	0.101
Common-language dummy	0.644	0.751	0.888	0.815	0.892	0.433	0.448
Colonial-tie dummy	0.226	0.020	-0.098	0.083	0.011	0.327	0.267
Landlocked-exporter dummy	-0.500	-0.872	-0.626	-0.755	-0.725	-0.098	-0.047
Landlocked-importer dummy	-0.816	-0.703	-0.700	-0.637	-0.620	-0.432	-0.401
Exporter's remoteness	0.932	0.647	0.453	0.696	0.243	0.307	0.299
Importer's remoteness	0.134	0.549	0.303	0.717	0.490	-0.019	-0.017
Free-trade agreement dummy	0.801	0.179	-0.001	0.413	0.176	0.127	0.183
Openess	-0.456	-0.139	-0.411	-0.538	-0.647	-0.023	-0.023
Nr. observations	9613	9613	9613	9613	9613	9613	9613
θ_1			6.4e-07	1.8e+00	8.1e-02	4.5e+00	3.6e+00
θ_2				1.6e+00	2.2e+00	1.2e+00	1.9e-01

Table 10: Narrow replication results Log of Gravity Table 4 without zero trade values.

Dependent Variable	OLS-het	P-PML	G-QGPML	N-QEPML-PV	N-QEPML-QV	LN-QEPML-PV	LN-QEPML-QV
Log distance	-1.166	-0.770	-0.934	-1.120	-1.072	-0.917	-0.970
Contiguity dummy	0.347	0.352	0.332	0.411	0.404	-0.019	0.107
Common-language dummy	0.328	0.418	0.279	0.579	0.556	0.269	0.281
Colonial-tie dummy	0.627	0.038	0.341	0.363	0.331	0.510	0.440
Free-trade agreement dummy	1.191	0.374	0.131	0.399	0.423	-0.111	0.018
Nr. observations	9613	9613	9613	9613	9613	9613	9613
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
θ_1			6.5e-07	1.1e+00	1.8e-04	1.5e+00	6.1e-01
θ_2				1.8e+00	1.8e+00	1.3e+00	3.6e-01

Table 11: Narrow replication results Log of Gravity Table 3 with zero trade values.

Dependent Variable	P-PML	G-QGPML	N-QEPML-PV	N-QEPML-QV	LN-QEPML-PV	LN-QEPML-QV
Constant	-32.326	-29.511	-41.277	-35.528	-31.552	-30.423
Log exporter's GDP	0.732	0.758	0.609	0.594	0.726	0.698
Log importer's GDP	0.741	0.757	0.668	0.671	0.615	0.591
Log exporter's GDP per capita	0.157	0.199	0.219	0.208	0.137	0.135
Log importer's GDP per capita	0.135	0.176	0.270	0.246	0.083	0.083
Log distance	-0.784	-0.827	-0.926	-0.818	-0.834	-0.801
Contiguity dummy	0.193	0.191	0.526	0.397	0.003	0.101
Common-language dummy	0.746	0.870	0.815	0.892	0.433	0.448
Colonial-tie dummy	0.025	-0.070	0.083	0.011	0.327	0.267
Landlocked-exporter dummy	-0.863	-0.565	-0.755	-0.725	-0.098	-0.047
Landlocked-importer dummy	-0.696	-0.678	-0.637	-0.620	-0.432	-0.401
Exporter's remoteness	0.660	0.469	0.696	0.243	0.307	0.299
Importer's remoteness	0.561	0.292	0.717	0.490	-0.019	-0.017
Free-trade agreement dummy	0.181	0.004	0.413	0.176	0.127	0.183
Openess	-0.107	-0.380	-0.538	-0.647	-0.023	-0.023
Nr. observations	18360	18360	18360	18360	18360	18360
θ_1		1.1e-06	2.3e+00	1.5e-01	7.6e+00	6.9e+00
θ_2			1.6e+00	2.2e+00	1.2e+00	1.9e-01

Table 12: Narrow replication results Log of Gravity Table 4 with zero trade values.

Dependent Variable	P-PML	G-QGPML	N-QEPML-PV	N-QEPML-QV	LN-QEPML-PV	LN-QEPML-QV
Log distance	-0.750	-0.910	-1.120	-1.072	-0.917	-0.970
Contiguity dummy	0.370	0.381	0.411	0.404	-0.019	0.107
Common-language dummy	0.383	0.213	0.579	0.556	0.269	0.281
Colonial-tie dummy	0.079	0.440	0.363	0.331	0.510	0.440
Free-trade agreement dummy	0.376	0.113	0.399	0.423	-0.111	0.018
Nr. observations	18360	18360	18360	18360	18360	18360
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
θ_1		1.1e-06	1.2e+00	3.4e-04	2.4e+00	1.2e+00
θ_2			1.8e+00	1.8e+00	1.3e+00	3.6e-01

6 Conclusion

This paper evaluates the performance of gravity estimation methodologies in non-negative trade data with many (economically determined) zero values that is heteroskedastic, positively skewed and heavy-tailed. We demonstrate that pseudo-maximum-likelihood estimators such as P-PML yield very imprecise estimates in heavy-tailed trade data. Consequently, if zero trade values are not economically determined, we argue that the gravity equation should be estimated with a heteroskedasticity-robust log-linear method. Heteroskedasticity-robust log-linear estimators remove the heavy-tail influence through a transformation of the data and account for heteroskedasticity, explicitly modeling the impact of the variance function on coefficient estimates. If zero-trade values are economically determined, we argue that both the log-linear and PML estimators need to model these zero-trade values accordingly. We propose a heteroskedasticity-robust log-linear Tobit I-model as the most apt estimator for non-negative trade data with economically determined zero values that is heteroskedastic, positively skewed, and heavy-tailed. Future work should focus on the generalizability of the variance function specification and the extension to easily account for high-dimensional fixed effects.

References

(2008). Comments on "The Log of Gravity revisited".

Basu, A. and P. J. Rathouz (2005, January). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics 6*(1), 93–109.

Bosquet, C. and H. Boulhol (2014, October). Applying the GLM Variance Assumption to Overcome the Scale-Dependence of the Negative Binomial QGPML Estimator. *Econometric Reviews 33*(7), 772–784. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07474938.2013.806102.

Breinlich, H., D. Novy, and J. Santos Silva (2022, July). Trade, Gravity and Aggregation. *The Review of Economics and Statistics*, 1–29.

Buntin, M. B. and A. M. Zaslavsky (2004, May). Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. *Journal of Health Economics 23*(3), 525–542.

Burger, M., F. van Oort, and G.-J. Linders (2009, June). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis 4*(2), 167–190. Publisher: Routledge _eprint: https://doi.org/10.1080/17421770902834327.

Carroll, R. J. and D. Ruppert (1983). *Transformation and Weighting in Regression.* Routledge.

Gourieroux, C., A. Monfort, and A. Trognon (1984a). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica 52*(3), 701–720. Publisher: [Wiley, Econometric Society].

Gourieroux, C., A. Monfort, and A. Trognon (1984b). Pseudo Maximum Likelihood Methods: Theory. *Econometrica 52*(3), 681–700. Publisher: [Wiley, Econometric Society].

Head, K. and T. Mayer (2014, January). Chapter 3 - Gravity Equations: Workhorse,Toolkit, and Cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 131–195. Elsevier.

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

Jones, A. M., J. Lomas, P. T. Moore, and N. Rice (2016, October). A quasi-Monte-Carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. *Journal of the Royal Statistical Society. Series A, (Statistics in Society) 179*(4), 951–974.

Kurz, C. F. (2017, December). Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Medical Research Methodology 17*(1), 171.

Manning, W. G. (1998, June). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics 17*(3), 283–295.

Manning, W. G., A. Basu, and J. Mullahy (2005, May). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics 24*(3), 465–488.

Manning, W. G. and J. Mullahy (2001, July). Estimating log models: to transform or not to transform? *Journal of Health Economics 20*(4), 461–494.

Martin, W. (2020, September). Making Gravity Great Again. Working Paper, World Bank, Washington, DC. Accepted: 2020-09-17T20:04:34Z.

Martin, W. and C. S. Pham (2020, June). Estimating the gravity model when zero trade flows are frequent and economically determined. *Applied Economics 52*(26), 2766–2779. Publisher: Routledge _eprint: https://doi.org/10.1080/00036846.2019.1687838.

Martínez-Zarzoso, I. (2013, January). The log of gravity revisited. *Applied Economics 45*(3), 311–327. Publisher: Routledge _eprint: https://doi.org/10.1080/00036846.2011.599786.

McCullagh, P. (1989). *Generalized Linear Models* (2nd edition ed.). Routledge. Google-Books-ID: SPSDDwAAQBAJ.

Mnasri, A. and S. Nechi (2021, February). New nonlinear estimators of the gravity equation. *Economic Modelling 95*, 192–202.

Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General) 135*(3), 370–384. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/2344614.

Santos Silva, J. M. C. and S. Tenreyro (2011, August). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters 112*(2), 220–222.

Silva, J. M. C. S. and S. Tenreyro (2006, November). The Log of Gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Sukanuntathum, A. (2012, January). Robust Estimation of Gravity Models under Heteroskedasticity and Data Censoring. *Procedia - Social and Behavioral Sciences 40*, 731–735.

Xiong, B. and S. Chen (2014, August). Estimating gravity equation models in the presence of sample selection and heteroscedasticity. *Applied Economics 46*(24), 2993–3003.

Yotov, Y. (2022). Gravity at Sixty: The Workhorse Model of Trade. *SSRN Electronic Journal*.