The Log of Heavy-Tailed Gravity

Ruben Dewitte[*]

September, 2023

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 evaluates the consistency and efficiency of the P-PML relative to alternative estimators when data is heavy-tailed, using 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, heteroscedastic, 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, heteroscedastic trade flows (see Yotov (2022) for an overview of gravity applications). The general consistency of the P-PML estimator in high-dimensional fixed effects settings reinforced this use (Fernández-Val and Weidner, 2016, Weidner and Zylkin, 2021). Manning and Mullahy (2001) argue, however, that the P-PML estimator can yield 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 that is heteroscedastic, positively skewed, and heavy-tailed. To this aim, we introduce several novel estimation methodologies to the trade literature. Using a Monte Carlo analysis, we demonstrate that no gravity estimation methodology, including P-PML, performs better in terms of efficiency than all other estimation methodologies. Rather, the efficiency of a gravity estimator depends on the specification of the variance function of the underlying model and on the number of fixed effects relied on. When taking the models to the data, we find large discrepancies between the MC results and real-life results. We attribute these discrepancies to model misspecification and discuss the possible influence of model misspecification on gravity estimators.

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 researched 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 explanatory variables in the gravity equation (Mnasri and Nechi, 2021). Martínez-Zarzoso (2013) extend the Monte Carlo simulation of (Silva and

[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 $\alpha < 2$.

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 Silva and Tenreyro (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, even when these zero values are correlated with explanatory variables.

If the number of zero values is not random, the P-PML estimator has been 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 heteroscedasticity. Alternative two-step estimators that account for heteroscedasticity and economically determined zero trade values are the TS-MM moment estimator (Xiong and Chen, 2014), the heteroscedasticity-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 cross-sectional heavy-tailed data. Recently, Kwon et al. (2022) introduced a generalized P-PPML estimator as a more efficient alternative to P-PML. In line with the generalized PML estimator proposed by Gourieroux et al. (1984a) considered below, they propose a two-step estimator where the variance structure is estimated in the first stage, and this variance is then relied on in a second stage to improve the precision of the estimates. Kwon et al. (2022) demonstrate the consistency of the estimator in two-way fixed effect settings (exporter-time and importer-time fixed effects), leaving its consistency in three-way fixed effect settings (exporter-time, importer-time, and exporter-importer fixed effects) for future research.

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 heteroscedasticity and precision loss when the data is heavy-tailed, concluding that PML estimators can yield 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. Alternatively, Basu and Rathouz (2005) propose an extension of the PML estimators to simultaneously determine parameters of the link function and variance

[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.

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 these alternative estimators' varying performance and/or applicability.

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}:

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

We start discussing the properties of gravity estimators in the absence of fixed effects, and generalize this discussion to a fixed-effects setting in section 3.3. 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, heteroscedastic, 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.

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

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

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 heteroscedasticity 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 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 the efficiency of the estimator.

3.2 Data Weighting

Considering that the conditional variance might not be constant (i.e., heteroscedastic), 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 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, only a correct conditional mean specification is needed for a consistent estimator based on the above FOC. 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 a 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, downweights 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 discussed below, retrieving a scale-independent consistent estimator for a (Bosquet and Boulhol, 2014) is possible.

Lastly, the Lognormal estimator relies on a log transformation of the data, resulting in an FOC that reduces the influence of larger observations. Interpreting the coefficients as an elasticity requires accounting for the explanatory variable's impact on the dependent variable's variance. Under the distributional assumption that the error term follows a normal distribution, $ln(\varepsilon_{ijt}) \sim \mathcal{N}\left(0, V\left[ln(\varepsilon_{ijt})\right]\right)$, we can rely on the specification of the mean of a lognormal variable (see also 1) to obtain (Manning, 1998):

$$\tilde{\beta}_x \equiv \frac{\partial lnE\left[Y_{ijt}|\boldsymbol{x}_{ijt}\right]}{\partial x_{ijt}} = \beta_x + 0.5\frac{\partial V\left[ln(\varepsilon_{ijt})|\boldsymbol{x_{ijt}}\right]}{\partial x_{ijt}}, \tag{6}$$

where $\beta_x \equiv \frac{\partial E[lnY_{ijt}|\boldsymbol{x}_{ijt}]}{\partial lnx_{ijt}}$. The equation above suggests a two-stage procedure to obtain elasticity under heteroskedasticity. The first stage consists of estimating a log-level regression, which provides us with a coefficient estimate β_x and an estimate of the residual variance. A second equation with this variance as a dependent variable gives us an estimated heteroscedasticity correction term that allows us to obtain the $\tilde{\beta}_x$ (Manning, 1998).

[3]Note that if the model is misspecified, the variance function can influence the interpretation of the coefficients. See, for instance, Breinlich et al. (2022a).

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

λ	**Family**	$E\left[Y_{ijt}\mid\boldsymbol{x}_{ijt}\right]$	$V\left[Y_{ijt}\mid\boldsymbol{x}_{ijt}\right]$	$E\left[g(Y_{ijt},\lambda)\mid\boldsymbol{x}_{ijt}\right]$	$V\left[g(Y_{ijt},\lambda)\mid\boldsymbol{x}_{ijt}\right]$	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 ε_{ijt} equals one, the PML estimator based on the Poisson family is asymptotically efficient and better than the other estimators. If ε_{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 $V\left[ln(\varepsilon_{ijt})\right] \equiv \theta_1$ to build three QGPML estimators.[4]

These QGPML estimators require a consistent estimator $\hat{\theta}_1$ of θ_1. As Bosquet and Boulhol (2014) note, the original estimator for θ_1 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 θ_1. 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 $\theta_1 = \frac{\hat{b}}{\hat{a}}$ from the following regression:

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

The FOCs related to each specific family of QGPML estimators based on the consistent estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}_1$ are provided in Table 2 in the column titled QGPML. Notice that as $\hat{\theta}_1 \to 0$, 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 families are almost equivalent. We, therefore, do not expect large performance differentials between the QGPML estimators of different families.

Notice the similarities and differences between the QGPML approach and the correction for heteroscedasticity of a log-normal estimator (see Eq. 6) in the previous section. Both approaches rely on the consistent estimation of the conditional variance in a first stage. The QGPML approach, then, relies on a parametrization of this conditional variance in the second stage to recover more efficient estimates. The correction for heteroscedasticity of the log-normal estimator, on the other hand, relies on a distributional assumption of the error term to recover the consistent coefficient estimates.

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

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	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta_1 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}} + \hat{\theta}_1 e^{2\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}$	$\frac{\boldsymbol{x}_{ijt}^T e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}(1+\hat{\theta}_1 e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}})}$
1	Gamma	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta_1 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$ke^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$(e^{-\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}} + \hat{\theta}_1) e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{\boldsymbol{x}_{ijt}^T e^{-\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{e^{-\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}}(1+\hat{\theta}_1 e^{\boldsymbol{x}_{ijt}\hat{\boldsymbol{\beta}}})}$
1	Negative Binomial Type 2	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta_1 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \hat{\theta}_1 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\frac{\boldsymbol{x}_{ijt}^T \left(Y_{ijt}-e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}\right)}{1+\hat{\theta}_1 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}}$

Notes: $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}_1$ indicate consistent estimates of the parameter vector $\boldsymbol{\beta}$ and the disturbance's variance θ_1 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):

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

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**	$E\left[Y_{ijt} \| \boldsymbol{x}_{ijt}\right]$	$V\left[Y_{ijt} \| \boldsymbol{x}_{ijt}\right]$	$E\left[g(Y_{ijt}, \lambda) \| \boldsymbol{x}_{ijt}\right]$	$V\left[g(Y_{ijt}, \lambda) \| \boldsymbol{x}_{ijt}\right]$
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	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \| \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{\sigma^2}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}} + 1\right)$
0	LN	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta_1 e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}} + \theta_2 e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \| \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{V[Y_{ijt} \| \boldsymbol{x}_{ijt}]}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}} + 1\right)$
0	LN	$e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\theta_1 e^{\theta_2 \boldsymbol{x}_{ijt}\boldsymbol{\beta}}$	$\boldsymbol{x}_{ijt}\boldsymbol{\beta} - \frac{V[lnY_{ijt} \| \boldsymbol{x}_{ijt}]}{2}$	$ln\left(\frac{V[Y_{ijt} \| \boldsymbol{x}_{ijt}]}{e^{2\boldsymbol{x}_{ijt}\boldsymbol{\beta}}} + 1\right)$

3.3 Extension to Fixed Effects Specifications

In line with the applied literature, we can extend the original specification (see Eq. 1) to a three-way fixed effect setting with exporter-time, importer-time, and exporter-importer fixed effects (respectively, $\theta_{it}, \gamma_{jt},$ and η_{ij}):

$$Y^*_{ijt} = e^{\boldsymbol{x}_{ijt}\boldsymbol{\beta} + \theta_{it} + \gamma_{jt} + \eta_{ij}} + \varepsilon_{ijt}. \qquad (9)$$

The corresponding FOC for PML estimators are, then:

$$
\begin{aligned}
\boldsymbol{\beta} &: \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \frac{g(Y_{ijt},\lambda) - E\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]}{V\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]} \frac{\partial E\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]}{\partial \boldsymbol{\beta}} = 0; \\
\theta_{it} &: \sum_{j=1}^{J} \frac{g(Y_{ijt},\lambda) - E\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]}{V\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]} = 0; \\
\gamma_{jt} &: \sum_{i=1}^{I} \frac{g(Y_{ijt},\lambda) - E\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]}{V\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]} = 0; \\
\eta_{ij} &: \sum_{t=1}^{T} \frac{g(Y_{ijt},\lambda) - E\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]}{V\left[g(Y_{ijt},\lambda)|\boldsymbol{x}_{ijt}\right]} = 0.
\end{aligned}
\tag{10}
$$

Claims regarding the consistency of the PML estimators carry over to the fixed effects setting, with one caveat: the incidental-parameters problem (IPP). The IPP arises when estimation noise from estimates of fixed effects and other "incidental parameters" contaminates the scores of the main parameters of interest, inducing bias. Fernandes et al. (2018) demonstrate the asymptotic consistency of the above-mentioned PML estimators in two-way fixed effect settings (i.e. with exporter-time and importer-time but without exporter-importer fixed effects). (Weidner and Zylkin, 2021) demonstrate that the P-PML estimator is also asymptotically consistent (but standard errors are asymptotically biased) in a three-way gravity setting but is the only estimator in the set of above-mentioned untransformed PML estimators. The G-PML estimator, for instance, is only consistent when the variance specification is correct. The log-transformed PML estimator, on the other hand, is also expected to be free from the IPP in a three-way gravity setting.

Research considering the consistency of QGPML estimators in fixed effects settings is scarce. Building on Fernandes et al. (2018), it appears a PML estimator will be asymptotically consistent in a two-way setting if the variance relied on is consistently estimated. Kwon et al. (2022) demonstrate for their specific QGPML estimator that if the parameters from the variance specification are consistently estimated in the first stage of a QGPML estimator, then the QGPML estimator will be asymptotically consistent in a two-way setting. Building on the results of Weidner and Zylkin (2021), it appears that a PML estimator will also be asymptotically consistent (with asymptotically biased standard errors) in a three-way setting if the variance is correctly specified or equal to one. Kwon et al. (2022) use a similar argument to argue that their results for their specific QGPML estimator likely spill over to a three-way setting.

Extension of the QEPML estimators to a fixed effects setting is difficult in practice due to the nonlinear nature of the score function. The closest approximation observed in the literature is the iterative QGPML estimator proposed by Mnasri and Nechi (2021), where the first-stage estimates of the variance function are continuously updated using the second-stage outcomes until convergence is achieved. There is no knowledge of research examining the behavior of such estimators relating the IPP, although it can be expected that, in line with the one-shot QGPML estimator, an iterative QGPML estimator will be asymptotically consistent if the variance is consistently estimated (two-way setting) or correctly specified or equal to one (three-way gravity setting).

4 Monte Carlo Evidence

We compare the efficiency of the above-specified gravity estimators in a naive gravity, two-way, and three-way gravity setting using an extension of the original Monte Carlo exercise of Silva and Tenreyro (2006) to a panel setting, in line with Weidner and Zylkin (2021). We simulate the multiplicative model

$$\begin{aligned} E\left[Y_{ijt}|RTA_{ijt}, tariff_{ijt}, \theta_{it}, \gamma_{jt}, \eta_{ij} \equiv \hat{Y}_{ijt} = \right. \\ \exp\left(\beta_{rta}RTA_{ijt} + \beta_{tariff}ln(tariff_{ijt}) + \theta_{it} + \gamma_{jt} + \eta_{ij}\right), \end{aligned} \tag{11}$$

with $N = 50$ and $T = 10$. We simulate a continuous variable that mimics a tariff variable in a gravity model: $log(tariff_{ijt}) = log(tariff_{ijt-1})/2 + +\theta_{it} + \gamma_{jt} + \eta_{ij} + \nu_{ijt}$, where $\nu_{ijt} \sim \mathcal{N}(0, 0.5)$, and normalize the variable such that $log(tariff_{ijt}) \sim \mathcal{N}(1.7, 1)$. RTA_{ijt} is a binary variable that is simulated based on a continuous random normal variable with a standard deviation of 1 and a correlation with $log(tariff)_{ijt}$ equal to 0.7, which is the discretized such that the smallest 30% of the variable takes on the value one. As such, this variable could capture the existence of a Regional Trade Agreement or Non-Tariff Barriers to Trade, correlated with the tariff variable. A new set of observations of all variables is generated in each replication using $\beta_{RTA} = 0.5, \beta_{tariff} = -1, \theta_{it} \sim \mathcal{N}(3, 2), \gamma_{jt} \sim \mathcal{N}(-6, 2), \eta_{ij} \sim \mathcal{N}(7.5, 2)$. Data on Y are generated as

$$Y_{ijt} = \mathbb{I}\left[\hat{Y}_{ijt}\varepsilon_{ijt} \geq Q_p(\hat{Y}_{ijt}\varepsilon_{ijt})\right]\hat{Y}_{ijt}\varepsilon_{ijt}, \tag{12}$$

where ε_{ijt} is a log-normal disturbance with mean 1 and variance σ^2_{ijt}. $\mathbb{I}$ is and indicator function, and $Q_p(\cdot)$ is a quantile function evaluated at the $p - th$ percentile.

To assess the performance of the estimators under different patterns of heteroscedasticity, we follow Silva and Tenreyro (2006) in considering the four following specifications of σ^2_{ijt}:

DGP 1: $\sigma^2_{ijt} = \hat{Y}^{-2}_{ijt}$; $V\left[Y_{ijt} \mid \cdot\right] = 1$;

DGP 2: $\sigma^2_{ijt} = \hat{Y}^{-1}_{ijt}$; $V\left[Y_{ijt} \mid \cdot\right] = \hat{Y}_{ijt}$;

DGP 3: $\sigma^2_{ijt} = 1$; $V\left[Y_{ijt} \mid \cdot\right] = \hat{Y}^2_{ijt}$;

DGP 4: $\sigma^2_{ijt} = \hat{Y}^{-1}_{ijt} + \exp\left(2ln(tariff_{ijt})\right)$; $V\left[Y_{ijt} \mid \cdot\right] = \hat{Y}_{ijt} + tariff_{ijt} \times \hat{Y}^2_{ijt}$,

where we allow for serial correlation within pairs by imposing

$$\text{Cov}\left[\varepsilon_{ijs}, \varepsilon_{ijt}\right] = \exp\left[0.3^{|s-t|} \times \sqrt{\ln\left(1 + \sigma^2_{ijt}\right)}\sqrt{\ln\left(1 + \sigma^2_{ijt}\right)}\right] - 1,$$

and normalize ε_{ijt} such that the average standard deviation equals one for each DGP to ensure comparability between the DGPs.

All simulation parameters have been chosen to ensure the simulated datasets align with key empirical moments of the observed data (see Online Appendix A.1. The average summary statistics for each DGP across all Monte Carlo simulations are reported in Online Appendix Table A.2. It can be observed that, across all DGPs, the MC dataset aligns with the observed data in terms of the signal-to-noise ratio, the average and standard deviations of the dependent variable, the RTA variable, the fixed effects, and the error term.

We perform 100 Monte Carlo simulations and estimate for each simulation (i) a naive gravity model assuming the fixed effect values $\theta_{it}, \gamma_{jt}, \eta_{ij}$ are observed, (i) a two-way fixed effects model assuming the pair fixed effects η_{ij} are observed, and (iii) a three-way fixed effect model. Our Naive estimations rely on estimators that differ in the underlying distributional assumption (Poisson (P), Gamma (G), and Log-normal (LN)), in the estimator type (PML, QGPML, and QEPML), and the variance function specification (Power variance (PV), and quadratic variance (QV)). The two-way and three-way fixed effects estimations consider, due to technical limitations (cf. infra), only estimators that differ in the underlying distributional assumption (Poisson (P), Gamma (G), and Log-normal (LN)) and in the estimator type (PML and QGPML).

The MC results for the naive gravity model are displayed in Table 4. The results confirm the results from similar MC performed by, amongst others, Silva and Tenreyro (2006) and Weidner and Zylkin (2021) for DGP 1 and 2, where the P-PML estimator demonstrates low bias and high efficiency. In contrast, as the DGP of the variance specification diverges from the theoretical variance specification of the P-PML for DGP 3 and 4, performance degrades in our MC. This performance degradation is starker than what is observed in the MCs performed by Silva and Tenreyro (2006) and Weidner and Zylkin (2021). This divergence can be ascribed to the fact that the average standard deviation of the error term and the Signal-to-Noise ratio are kept relatively constant across DGPs, whereas this is not the case for the MC results provided by Silva and Tenreyro (2006) and Weidner and Zylkin (2021).

The evaluation of novel alternative estimators demonstrates that the performance of the P-PML estimator can be matched or even improved upon. The G-PML estimator is relatively less efficient than the P-PML estimator for DGP 1 and DGP 2 but, in contrast with the P-PML estimator, provides similar performance across all DGP specifications. The G-QGPML estimator matches the performance of the P-PML estimator in DGP 1 and 2 and is less biased and more efficient than the P-PML estimator in DGP 3 and 4, where the variance diverges from the theoretical variance specification of the P-PML estimator. The G-QGPML estimator more efficient than the P-PML estimator in all instances. The heteroscedasticity robust Log-OLS estimator delivers relatively similar performance across all DGPs but is less efficient than the G-QGPML and G-PML estimators. Concerning the QEPML estimators, power variance specification delivers the best results for both the Normal and Lognormal assumption, with better performance for the Lognormal distribution. The LN-QEPML-PV estimator is best among its peers in recovering the correct θ_1 and θ_2 parameters. Compared to G-QGPML and G-PML, the LN-QEPML-PV estimator performs similarly in terms of bias and variability of the coefficient estimates in DGP 1, 2, and 3. However, the QEPML estimators are not as robust to a misspecification of the variance function, as shown by the results in DGP 4 of the Monte

Carlo exercise.

We extend the MC exercise to two-way and three-way gravity estimators in Table 5 and 6, respectively. Relative to the naive gravity setting, the bias seems to increase across all DGPs for the G-PML and OLS-het. estimators, while this is not the case for the P-PML estimators. The bias of the G-QGPML estimator only increases when moving towards a three-way gravity setting, delivering a performance almost equivalent to the P-PML estimator. The asymptotic bias of the G-PML estimator also becomes apparent when moving from the two-way gravity to the three-way gravity estimator. This bias is in line with, though smaller than, the demonstrated asymptotic bias of the G-PML estimator by Weidner and Zylkin (2021).

Overall, our MC results demonstrate that there is no uniformly estimator uniformly better than other estimators in terms of bias and/or efficiency. The performance of all estimators varies across specifications of the variance function and the specification of the fixed effects. The performance of the P-PML estimator seems least affected by the introduction of fixed effects but most affected by variation in the specification of the variance function. The G-PML estimator, on the other hand, is relatively less affected by different specifications of the variance function, but its performance degrades as fixed effects are added to the specification. Also OLS-het. estimator, then, proves to be a valid option when estimating gravity models.

Table 4: Monte Carlo Results for a Naive Gravity Model

	β_{rta}		β_{tariff}		θ_1		θ_2	
Method	Bias	S.E.	Bias	S.E.	Est.	S.E.	Est.	S.E.
DGP 1: $V[Y_{odt}\|\cdot] = 1$								
P-PML	0.00004	0.00014	0.00000	0.00003				
G-QGPML	0.00002	0.00015	0.00000	0.00003			0.00000	0.00000
G-PML	0.00477	0.06851	0.00064	0.02624				
OLS	0.01288	0.03135	-0.17027	0.01437				
OLS-het.	0.02563	0.05095	0.07705	0.02217				
N-QEPML-PV	0.00000	0.00002	-0.00000	0.00000	0.00000	0.00000	0.04013	0.03806
N-QEPML-QV	0.00139	0.00673	0.00023	0.00103	0.00061	0.00051	-0.00000	0.00000
LN-QEPML-PV	-0.00011	0.00078	0.00001	0.00008	0.00000	0.00000	0.17012	0.14961
LN-QEPML-QV	-0.00098	0.00260	-0.00008	0.00031	0.00029	0.00008	-0.00000	0.00000
DGP 2: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}$								
P-PML	0.00013	0.00191	-0.00001	0.00051				
G-QGPML	0.00011	0.00191	-0.00001	0.00052			0.00000	0.00000
G-PML	-0.00201	0.06344	-0.00261	0.02193				
OLS	0.02465	0.02251	-0.16683	0.01434				
OLS-het.	0.01629	0.04049	0.04151	0.01628				
N-QEPML-PV	0.00021	0.00222	-0.00001	0.00050	0.00169	0.00049	0.96608	0.03156
N-QEPML-QV	0.00039	0.00486	0.00007	0.00093	0.00194	0.00049	-0.00000	0.00000
LN-QEPML-PV	0.00009	0.00177	-0.00002	0.00050	0.00162	0.00044	0.95280	0.00564
LN-QEPML-QV	-0.00032	0.00227	-0.00023	0.00097	0.00190	0.00049	-0.00000	0.00000
DGP 3: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}^2$								
P-PML	-0.06362	0.28651	0.02038	0.15165				
G-QGPML	-0.00929	0.10790	-0.01268	0.04716			0.00143	0.00237
G-PML	0.00157	0.02599	-0.00043	0.01644				
OLS	-0.00120	0.01968	-0.00125	0.01314				
OLS-het.	-0.00107	0.02346	-0.00118	0.01481				
N-QEPML-PV	0.01039	0.07999	0.00340	0.04072	1.70273	0.12407	1.99908	0.01636
N-QEPML-QV	0.03701	0.12024	-0.01930	0.13832	1.09856	7.76799	1.72674	0.27966
LN-QEPML-PV	-0.00102	0.01960	-0.00170	0.01299	1.72744	0.03734	2.00137	0.00536
LN-QEPML-QV	-0.00094	0.01935	0.00022	0.01550	0.00000	0.00000	1.71418	0.02629
DGP 4: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$								
P-PML	-0.04224	0.37180	0.04720	0.22378				
G-QGPML	-0.01427	0.08906	0.00259	0.05412			0.00094	0.00100
G-PML	0.00007	0.02386	-0.00144	0.01430				
OLS	-0.03417	0.01714	-0.23435	0.01086				
OLS-het.	0.00303	0.02162	-0.00231	0.01339				
N-QEPML-PV	0.02663	0.07351	0.30609	0.04417	2.10596	0.42695	2.09033	0.02358
N-QEPML-QV	0.03220	0.05935	0.27397	0.04385	0.00003	0.00002	1.39553	0.06572
LN-QEPML-PV	-0.02471	0.01840	-0.24545	0.01137	2.01894	0.06781	2.06003	0.00708
LN-QEPML-QV	-0.03370	0.01715	-0.23083	0.01074	0.00001	0.00000	1.57905	0.00852

Notes: Monte Carlo results obtained from 100 simulations.

Table 5: Monte Carlo Results for a twoway Gravity Model

	β_{rta}		β_{tariff}		θ_1	
Method	Bias	S.E.	Bias	S.E.	Est.	S.E.
DGP 1: $V[Y_{odt}\|\cdot] = 1$						
P-PML	0.00006	0.00025	0.00003	0.00016		
G-QGPML	0.00005	0.00024	0.00003	0.00016	0.00000	0.00000
G-PML	0.01430	0.02196	-0.02571	0.03591		
OLS	0.07274	0.02838	-0.16108	0.03553		
OLS-het.	-0.01967	0.04186	0.04253	0.04982		
DGP 2: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}$						
P-PML	0.00015	0.00320	-0.00006	0.00312		
G-QGPML	0.00017	0.00327	-0.00008	0.00312	0.00000	0.00000
G-PML	0.01378	0.02541	-0.02986	0.03989		
OLS	0.07294	0.02096	-0.17122	0.04072		
OLS-het.	-0.00754	0.03717	0.02199	0.04213		
DGP 3: $V[Y_{odt}\|\cdot] = \hat{Y}^2_{odt}$						
P-PML	-0.02730	0.17213	-0.01769	0.24265		
G-QGPML	-0.00874	0.12511	0.00191	0.16737	0.00006	0.00007
G-PML	-0.00096	0.02410	-0.00763	0.05295		
OLS	-0.00120	0.01897	-0.01144	0.04493		
OLS-het.	-0.00164	0.02298	-0.01086	0.05268		
DGP 4: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$						
P-PML	-0.03294	0.22013	-0.04484	0.35984		
G-QGPML	-0.00996	0.14290	-0.04772	0.24568	0.00006	0.00006
G-PML	-0.00362	0.02183	-0.02963	0.04957		
OLS	-0.01555	0.01819	-0.22572	0.04699		
OLS-het.	0.00019	0.02155	-0.01283	0.05075		

Notes: Monte Carlo results obtained from 100 simulations.

Table 6: Monte Carlo Results for a threeway Gravity Model

	β_{rta}		β_{tariff}		θ_1	
Method	Bias	S.E.	Bias	S.E.	Est.	S.E.
DGP 1: $V[Y_{odt}\|\cdot] = 1$						
P-PML	0.00003	0.00036	0.00004	0.00032		
G-QGPML	0.00004	0.00034	0.00004	0.00032	0.00000	0.00000
G-PML	0.02790	0.01549	-0.05836	0.02641		
OLS	0.07730	0.02552	-0.16103	0.03061		
OLS-het.	0.00377	0.02906	-0.01890	0.04598		
DGP 2: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}$						
P-PML	-0.00016	0.00367	-0.00111	0.00433		
G-QGPML	-0.00016	0.00364	-0.00110	0.00434	0.00000	0.00000
G-PML	0.02880	0.02023	-0.06471	0.03616		
OLS	0.07608	0.02264	-0.16624	0.04298		
OLS-het.	0.01328	0.03143	-0.03158	0.04732		
DGP 3: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}^2$						
P-PML	0.02574	0.14441	-0.01700	0.19078		
G-QGPML	0.02539	0.13782	-0.01733	0.17785	0.00001	0.00001
G-PML	0.00068	0.02296	-0.01085	0.05578		
OLS	0.00071	0.01966	-0.01113	0.05137		
OLS-het.	0.00140	0.02310	-0.01061	0.05968		
DGP 4: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$						
P-PML	-0.02961	0.14723	-0.07746	0.19134		
G-QGPML	-0.02396	0.13786	-0.07618	0.18043	0.00001	0.00001
G-PML	-0.00703	0.02145	-0.09924	0.05038		
OLS	-0.01304	0.01969	-0.22382	0.04752		
OLS-het.	-0.00280	0.02180	-0.07490	0.05312		

Notes: Monte Carlo results obtained from 100 simulations.

5 Real-life Application

Provided the MC results, we expect that all four consistent estimators (P-PML, G-QGPML, G-PML, and OLS-het.) will deliver similar coefficient estimates when applied to real-life data. We rely on the trade database provided by Yotov et al. (2016). The sample consists of a balanced panel that covers 69 countries their aggregate manufacturing over the period 1986-2006. It includes consistently constructed international and intra-national trade flows data, and contains data on regional trade agreements. All standard gravity variables including distance, contiguous borders, common language, and colonial ties are from the CEPII Distances database.

We start by estimating a naive gravity model of the form

$$Y_{ijt} = e^{\beta_1 RTA_{ijt}+\beta_2 ln(dist_{ij})+\beta_3 colony_{ij}+\beta_4 contiguity_{ij}+\beta_5 commonlanguage_{ij}+\beta_6 D_{intra}+\beta_7 Y_{it}++\beta_8 E_{jt}}\varepsilon_{ijt}, \tag{13}$$

where D_{intra} represents a dummy for intra-national trade. Y_{ijt} and E_{jt} represent aggregate export revenue and import expenditure, and ε_{ijt} captures the multiplicative error term. The Two-way gravity model is estimated based on the naive specification enhanced with the respective fixed effects (θ_{it} and γ_{jt}) with the appropriate variables removed to avoid collinearity issues (Y_{it}, and E_{jt}). The three-way gravity model continues building on the two-way gravity model, additionally including a pair fixed effect (η_{ij}) with the appropriate variables removed to avoid collinearity issues (all variables with ij subscript).

Results for all three gravity specifications are displayed in Table 7. P-PML estimates and their evolution as an increasing number of fixed effects are included in the model and are in line with the literature (Piermartini and Yotov, 2016). Surprisingly, there are large differences in coefficient estimates between different estimators for each model specification. RTA coefficients range from 0.31 for the P-PML estimator to -0.08 for the LN-QEPML-PV estimator for a naive gravity specification. Moreover, this pattern remains even after increasingly controlling for unobserved heterogeneity by including fixed effects.

Even though Kwon et al. (2022) observe a similar difference in coefficient estimates between estimators, these differences are unexpected, as P-PML, G-QGPML, G-PML and log-OLS-het. are expected to be consistent estimators of the RTA coefficient. For if the conditional mean is correctly specified, the only difference expected between consistent estimators would be in terms of efficiency.

Further evidence of model misspecification can be obtained from the average of the error term for the P-PML model ($\bar{\varepsilon}_{ijt}$), which is equal to 0.388 for the naive gravity model, 0.693 for the two-way gravity model, and 1.031 for the three-way gravity model. A correctly specified model requires this mean to be equal to one. Additionally, we estimate the same regression specification on a sample with one of the 200 largest observations (out of 91.506 total observations) removed. We evaluate the variability of the Leave-One-Out (LOO) coefficient estimates for the RTA dummy and log distance in Figure 1 panels 1a and 1b respectively. It can be

Table 7: Gravity estimation results.

	Naive Gravity		**Two-way Gravity**		**Three-way Gravity**
Method	RTA	ln(dist)	RTA	ln(dist)	RTA
P-PML	0.305	-0.131	0.181	-0.701	0.554
G-QGPML	0.248	-0.398	0.101	-0.723	0.422
G-PML	-0.101	-0.788	-0.013	-1.182	0.195
OLS	-0.196	-0.998	-0.022	-1.178	0.209
OLS-het.	-0.004	-0.699	-0.042	-1.209	0.252
N-QEPML-PV	-0.086	-0.622			
N-QEPML-QV	-0.064	-0.620			
LN-QEPML-PV	-0.079	-0.706			
LN-QEPML-QV	-0.115	-0.704			
Origin-Year Fixed Effects	No	No	Yes	Yes	Yes
Destination-Year Fixed Effects	No	No	Yes	Yes	Yes
Pair Fixed Effects	No	No	No	No	Yes

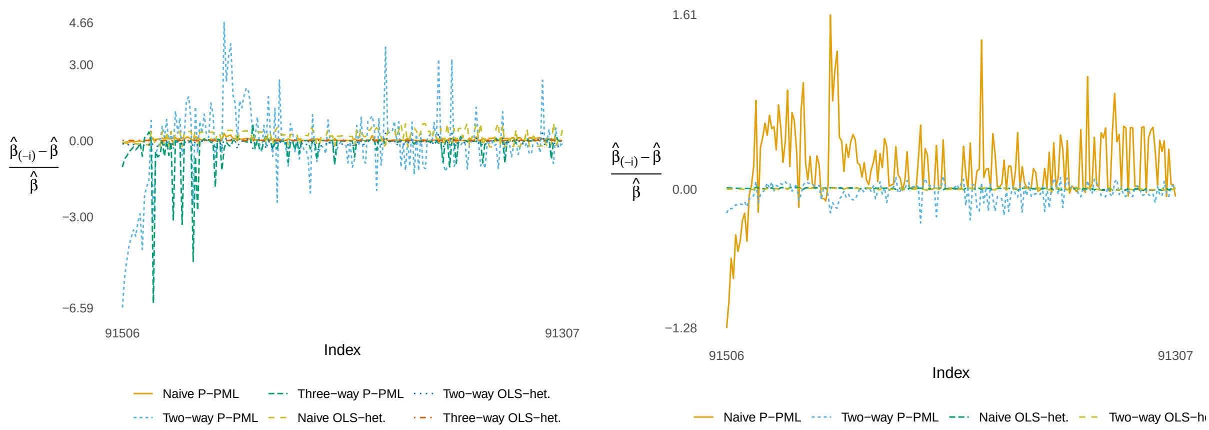
Notes: The sample consists of a balanced panel without zero trade flows that covers 69 countries their aggregate manufacturing over the period 1986-2006. Naive gravity is estimated based on a levels specification of the form $Y_{odt} = e^{\beta_1 RTA_{odt}+\beta_2 ln(dist_{od})+\beta_3 colony_{od}+\beta_4 contiguity_{od}+\beta_5 commonlanguage_{od}+\beta_6 D_{intra}+Y_{it}+E_{jt}}\varepsilon_{odt}$, where D_{intra} represents a dummy for intra-national trade. Y_{ijt} and E_{jt} represent aggregate export revenue and import expenditure, and ε_{ijt} captures the multiplicative error term. Two-way and three-way gravity are estimated based on the same specification enhanced with the respective fixed effects and the appropriate variables removed due to collinearity.

observed that omitting one large observation from the dataset is sufficient to decrease the RTA coefficient estimate by up to 6.59% for the two-way P-PML estimator and the coefficient on log-distance by up to 1.28% for the naive P-PML estimator. Provided the known efficiency of the P-PML estimator when models are correctly specified (see also the Monte Carlo exercise in the previous section), this variability likely points to model specification. In comparison, the heteroskedasticity-robust log-linear coefficient estimates of the complete sample are almost equivalent to the LOO coefficient estimates.

6 Causes and Effects of Model Misspecification

If the gravity model is misspecified, one wonders what the misspecification is and how it would affect coefficient estimates. Indications of answers to this question are scarce in the literature. ?Breinlich et al. (2022b) investigate whether a gravity model can be correctly specified as an aggregation of gravity models at a more disaggregated level. Breinlich et al. (2022b) demonstrates that it is only possible to recover unbiased estimates for a P-PML estimator in very restrictive cases and that the misspecification in other cases takes the form of an omitted variable. Gail et al. (1984), Neuhaus and Jewell (1993), Petersen and Deddens (2000) investigate the impact of omitted variables on PML estimators, but mainly focus on the case of omitted variables uncorrelated to the remaining explanatory variables.

What is clear, is that the choice of the variance function for a PML estimator is not solely a question of efficiency under model misspecification. When a model is misspecified, it does not fit the data equally well across the entire range of predicted values. An optimal fit in one



(a) RTA (b) ln(distance

Figure 1: Percentage difference between Leave-One-Out coefficient estimate and total sample estimate for 200 largest observations.

Note: The sample consists of a balanced panel without zero trade flows that covers 69 countries their aggregate manufacturing over the period 1986-2006. Naive gravity is estimated based on a levels specification of the form $Y_{odt} = e^{\beta_1 RTA_{odt}+\beta_2 ln(dist_{od})+\beta_3 colony_{od}+\beta_4 contiguity_{od}+\beta_5 commonlanguage_{od}+\beta_6 D_{intra}+Y_{it}+E_{jt}}\varepsilon_{odt}$, where D_{intra} represents a dummy for intra-national trade. Y_{ijt} and E_{jt} represent aggregate export revenue and import expenditure, and ε_{ijt} captures the multiplicative error term. Two-way and three-way gravity are estimated based on the same specification enhanced with the respective fixed effects and the appropriate variables removed due to collinearity. The database is sorted from small to large so that the largest trade observation in the dataset is indexed by the value 91,506.

part of the range implies at least a 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).

In the case of the four consistent estimators considered in this paper, these criteria of fit can be best exemplified by evaluating the FOC for one of the fixed effects, in this case, the importer-time fixed effect γ_{jt}:

$$\begin{aligned}
&\text{P-PML: } \frac{\sum_{i=1}^{I}\hat{Y}_{ijt}\varepsilon_{ijt}}{\sum_{i=1}^{I}\hat{Y}_{ijt}} = 1;\\
&\text{G-QGPML: } \frac{\sum_{i=1}^{I}V^{-1}(\hat{Y}_{ijt}|\cdot;\hat{\theta})\varepsilon_{ijt}}{\sum_{i=1}^{I}V^{-1}(\hat{Y}_{ijt}|\cdot;\hat{\theta})} = 1;\\
&\text{G-PML: } \frac{\sum_{i=1}^{I}\varepsilon_{ijt}}{I} = 1;\\
&\text{log-OLS: } \frac{\sum_{i=1}^{I}(ln\varepsilon_{ijt}+1)}{I} = 1.
\end{aligned} \tag{14}$$

These FOC reveal that the P-PML estimator focuses on a trade-weighted average error term, whereas the G-QGPML relies on an inverse-variance weighted average error term. G-PML focuses on the simple average of the error term, while log-OLS focuses on the average of the log of the error term + 1. In Table 7, we observe that the RTA and log-distance coefficient estimates are generally largest for P-PML, followed by G-QGPML, and smallest for G-PML, but closely matched by the heteroskedasticity-robust log-OLS estimates. This seems to indicate that countries with larger trade flows and/or a smaller variance exhibit a stronger correlation between an RTA and trade and a smaller correlation between log-distance and trade.

Building on earlier MC results, simulations results presented in Online Appendix Table A.3, A.4, and A.5 evaluate the performance of the estimators using the same MC setup, but with the continuous tariff variable omitted from the estimation specification. As expected, we find that omitting a relevant variable not only results in biased coefficient estimates but also affects the efficiency of the estimators. This to a point where the G-PML estimator is more efficient than P-PML and G-QGPML for all DGPs except for the case where the variance is correctly specified.

Overall, it can be concluded that model misspecification affects the performance of gravity estimators, and that further research is required into the causes and effects of model misspecification within gravity models

7 Conclusion

This paper evaluates the performance of gravity estimation methodologies in non-negative trade data that is heteroscedastic, positively skewed, and heavy-tailed. To this aim, we introduce several novel estimation methodologies to the trade literature. Using a Monte Carlo analysis, we demonstrate that no gravity estimation methodology, including P-PML, performs better in terms of efficiency than all other estimation methodologies. Rather, the efficiency of a gravity estimator depends on the specification of the variance function of the underlying model and on the number of fixed effects relied on. When taking the models to the data, we find large discrepancies between the MC results and real-life results. We attribute these discrepancies to model misspecification and discuss the possible influence of model misspecification on gravity estimators.

References

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 (2022a, July). Trade, Gravity and Aggregation. *The Review of Economics and Statistics*, 1–29.

Breinlich, H., D. Novy, and J. Santos Silva (2022b). 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.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, D. Pierola, and A. Rodríguez-Clare (2018, October). The Intensive Margin in Trade.

Fernández-Val, I. and M. Weidner (2016, May). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics 192*(1), 291–312.

Gail, M. H., S. Wieand, and S. Piantadosi (1984). Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates. *Biometrika 71*(3), 431–444. Publisher: [Oxford University Press, Biometrika Trust].

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.

Kwon, O., J. Yoon, and Y. Yotov (2022). A Generalized Poisson-Pseudo Maximum Likelihood Estimator. *School of Economics Working Paper Series* (2022-13).

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.

Neuhaus, J. M. and N. P. Jewell (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika 80*(4), 807–815.

Petersen, M. R. and J. A. Deddens (2000). Effects of omitting a covariate in poisson models when the data are balanced. *Canadian Journal of Statistics 28*(2), 439–445. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/3315990.

Piermartini, R. and Y. Yotov (2016, July). Estimating Trade Policy Effects with Structural Gravity. WTO Working Papers 2016/10. Series: WTO Working Papers Volume: 2016/10.

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.

Silva, J. M. C. S. and S. Tenreyro (2008). Comments on "The Log of Gravity revisited". *mimeo*.

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

Weidner, M. and T. Zylkin (2021, September). Bias and consistency in three-way gravity models. *Journal of International Economics 132*, 103513.

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*.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016, November). *An Advanced Guid To Trade Policy Analysis: The Structural Gravity Model.* WTO iLibrary.

Online Appendix to "The Log of Heavy-Tailed Gravity"

Ruben Dewitte*

September, 2023

*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.

Appendix A Additional Figures and Tables

A.1 Figures

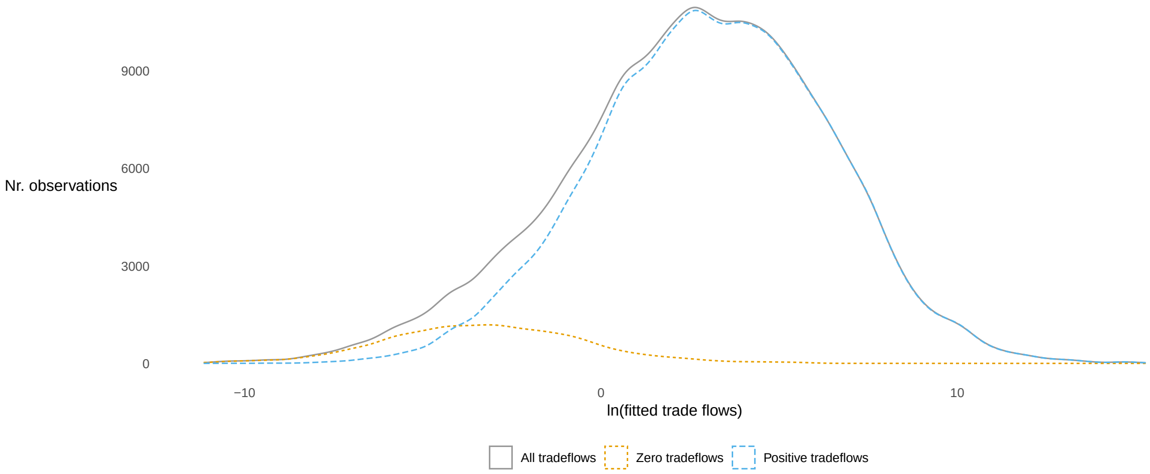


Figure A.1: Density of the log fitted values from a three-way gravity Poisson model by tradeflow values.
Note: Results obtained from a balanced panel that covers 69 countries their aggregate manufacturing over the period 1986-2006. The three-way Poisson Gravity model specification equals $trade_{ijt} = e^{\beta_{RTA} RTA_{ijt} + \theta_{it} + \gamma_{jt} + \eta_{ij}} \varepsilon_{ijt}$.

A.2 Tables

Table A.1: Summary Statistics of the input and output of a three-way Poisson Gravity model.

	Average	**Std. Dev.**
$ln(trade_{odt})$	2.801	3.631
RTA_{odt}	0.117	0.321
$ln(\theta_{ot})$	2.556	3.245
$ln(\gamma_{dt})$	-7.508	2.236
$ln(\eta_{od})$	7.612	2.833
$ln(\varepsilon_{odt})$	-0.351	1.136
Signal-to-Noise Ratio	10.208	
Prop. of zeroes	0.082	

Notes: The sample consists of a balanced panel that covers 69 countries their aggregate manufacturing over the period 1986-2006. θ_{ot}, γ_{dt}, and η_{od} represent, respectively, the origin-year, destination-year and origin-destination fixed effects recovered from a three-way Poisson Gravity model with specification $trade_{odt} = e^{\beta_{RTA} RTA_{odt} + \theta_{ot} + \gamma_{dt} + \eta_{od}} \varepsilon_{odt}$, where ε_{odt} captures the multiplicative error term. The signal-to-noise ratio is calculated as $SNR = \frac{V(\beta_{RTA} RTA_{odt} + \theta_{ot} + \gamma_{dt} + \eta_{od})}{V(ln(\varepsilon_{odt}))}$.

Table A.2: Monte Carlo Summary Statistics

	Average	**Std. Dev.**
	DGP 1: $V[Y_{odt}\|\cdot] = 1$	
$ln(trade_{odt})$	2.589	3.404
RTA_{odt}	0.100	0.300
$ln(\theta_{ot})$	2.999	1.987
$ln(\gamma_{dt})$	-6.007	1.993
$ln(\eta_{od})$	7.502	1.994
$ln(\varepsilon_{odt})$	-0.255	1.000
Signal-to-Noise Ratio	11.592	
Prop. of zeroes	0.000	
	DGP 2: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}$	
$ln(trade_{odt})$	2.533	3.419
RTA_{odt}	0.100	0.300
$ln(\theta_{ot})$	3.005	2.003
$ln(\gamma_{dt})$	-6.014	1.994
$ln(\eta_{od})$	7.498	1.999
$ln(\varepsilon_{odt})$	-0.306	1.000
Signal-to-Noise Ratio	11.688	
Prop. of zeroes	0.000	
	DGP 3: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}^2$	
$ln(trade_{odt})$	2.345	3.005
RTA_{odt}	0.100	0.300
$ln(\theta_{ot})$	2.989	1.986
$ln(\gamma_{dt})$	-5.995	1.979
$ln(\eta_{od})$	7.500	2.002
$ln(\varepsilon_{odt})$	-0.500	1.000
Signal-to-Noise Ratio	9.029	
Prop. of zeroes	0.000	
	DGP 4: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$	
$ln(trade_{odt})$	2.361	2.948
RTA_{odt}	0.100	0.300
$ln(\theta_{ot})$	3.003	2.004
$ln(\gamma_{dt})$	-6.007	1.983
$ln(\eta_{od})$	7.490	2.006
$ln(\varepsilon_{odt})$	-0.476	1.000
Signal-to-Noise Ratio	8.692	
Prop. of zeroes	0.000	

Notes: Monte Carlo summary statistics averaged across 100 simulations.

Table A.3: Monte Carlo Results for a Naive Gravity Model without tariffs

	β_{rta}		θ_1		θ_2	
Method	Bias	S.E.	Bias	S.E.	Est.	S.E.
			DGP 1: $V[Y_{odt}\vert\cdot] = 1$			
P-PML	0.51814	0.16991				
G-QGPML	0.46453	0.07636			0.00303	0.00571
G-PML	0.37930	0.05474				
OLS	0.44356	0.04990				
OLS-het.	0.39933	0.06207				
N-QEPML-PV	0.41689	0.29237	0.22887	0.02330	1.65292	0.06606
N-QEPML-QV	0.40458	0.08755	0.00173	0.00096	0.29057	0.01393
LN-QEPML-PV	0.33941	0.04190	0.24636	0.02216	1.60802	0.01030
LN-QEPML-QV	0.37337	0.03571	0.00298	0.00079	0.29564	0.01049
			DGP 2: $V[Y_{odt}\vert\cdot] = \hat{Y}_{odt}$			
P-PML	0.54482	0.17504				
G-QGPML	0.47288	0.06127			0.03125	0.20278
G-PML	0.38283	0.06016				
OLS	0.45440	0.04151				
OLS-het.	0.40669	0.05184				
N-QEPML-PV	0.41879	0.29319	0.25164	0.02604	1.65190	0.05094
N-QEPML-QV	0.41286	0.08364	0.00221	0.00065	0.33357	0.01507
LN-QEPML-PV	0.34439	0.03590	0.28665	0.02584	1.61211	0.00946
LN-QEPML-QV	0.37548	0.03036	0.00520	0.00133	0.33524	0.01327
			DGP 3: $V[Y_{odt}\vert\cdot] = \hat{Y}^2_{odt}$			
P-PML	0.45358	0.37580				
G-QGPML	0.47281	0.11688			0.00121	0.00170
G-PML	0.38632	0.03749				
OLS	0.36807	0.03322				
OLS-het.	0.39170	0.03553				
N-QEPML-PV	0.38716	0.08427	2.60026	0.30629	2.00227	0.02164
N-QEPML-QV	0.38479	0.08567	0.00001	0.00002	2.51125	0.24009
LN-QEPML-PV	0.36869	0.03305	2.67476	0.07321	2.00242	0.00620
LN-QEPML-QV	0.36814	0.03347	0.00000	0.00000	2.64574	0.04362
			DGP 4: $V[Y_{odt}\vert\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$			
P-PML	0.41819	0.44642				
G-QGPML	0.44239	0.13539			0.00098	0.00085
G-PML	0.37922	0.03031				
OLS	0.41881	0.03298				
OLS-het.	0.39568	0.03183				
N-QEPML-PV	0.28866	0.04715	2.40073	0.35362	2.09201	0.01982
N-QEPML-QV	0.31989	0.04117	0.00004	0.00003	1.61256	0.05458
LN-QEPML-PV	0.43910	0.03179	4.22905	0.15485	2.08780	0.00857
LN-QEPML-QV	0.41681	0.03303	0.00003	0.00001	3.03656	0.04805

Notes: Monte Carlo results obtained from 100 simulations.

Table A.4: Monte Carlo Results for a twoway Gravity Model without tariffs

	β_{rta}		θ_1	
Method	Bias	S.E.	Est.	S.E.
		DGP 1: $V\left[Y_{odt}\|\cdot\right] = 1$		
P-PML	0.05697	0.03283		
G-QGPML	0.05815	0.02284	0.00008	0.00008
G-PML	0.06557	0.02189		
OLS	0.13166	0.03077		
OLS-het.	0.02638	0.04037		
		DGP 2: $V\left[Y_{odt}\|\cdot\right] = \hat{Y}_{odt}$		
P-PML	0.06095	0.02446		
G-QGPML	0.05870	0.02006	0.00008	0.00007
G-PML	0.06501	0.02614		
OLS	0.13134	0.02345		
OLS-het.	0.04052	0.03810		
		DGP 3: $V\left[Y_{odt}\|\cdot\right] = \hat{Y}_{odt}^2$		
P-PML	0.02347	0.17553		
G-QGPML	0.04234	0.12935	0.00007	0.00007
G-PML	0.05131	0.02326		
OLS	0.05047	0.01902		
OLS-het.	0.05052	0.02219		
		DGP 4: $V\left[Y_{odt}\|\cdot\right] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$		
P-PML	0.02684	0.21463		
G-QGPML	0.05023	0.14126	0.00006	0.00006
G-PML	0.04788	0.02351		
OLS	0.04618	0.01934		
OLS-het.	0.05037	0.02265		

Notes: Monte Carlo results obtained from 100 simulations.

Table A.5: Monte Carlo Results for a threeway Gravity Model without tariffs

	β_{rta}		θ_1	
Method	Bias	S.E.	Est.	S.E.
		DGP 1: $V[Y_{odt}\|\cdot] = 1$		
P-PML	0.04499	0.02117		
G-QGPML	0.04513	0.01967	0.00002	0.00002
G-PML	0.07172	0.01706		
OLS	0.12611	0.02740		
OLS-het.	0.04468	0.02965		
		DGP 2: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt}$		
P-PML	0.05051	0.02196		
G-QGPML	0.05014	0.02131	0.00001	0.00001
G-PML	0.07251	0.02126		
OLS	0.12422	0.02359		
OLS-het.	0.05483	0.03267		
		DGP 3: $V[Y_{odt}\|\cdot] = \hat{Y}^2_{odt}$		
P-PML	0.06958	0.14927		
G-QGPML	0.06953	0.14291	0.00001	0.00001
G-PML	0.04229	0.02253		
OLS	0.04310	0.01927		
OLS-het.	0.04373	0.02257		
		DGP 4: $V[Y_{odt}\|\cdot] = \hat{Y}_{odt} + exp(2ln(tariff_{odt})$		
P-PML	0.01611	0.14835		
G-QGPML	0.02196	0.13993	0.00001	0.00001
G-PML	0.03810	0.02257		
OLS	0.03833	0.02066		
OLS-het.	0.04163	0.02288		

Notes: Monte Carlo results obtained from 100 simulations.

Appendix References