Economic Modelling 95 (2021) 192–202

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

New nonlinear estimators of the gravity equation☆

Ayman Mnasri, Salem Nechi *

Department of Finance and Economics, College of Business and Economics, Qatar University, Doha, P.O. Box. 2713, Qatar

ARTICLE INFO

JEL classification:
F14
C01
C13
C15
C63

Keywords:
Gravity model
Heteroscedasticity
Structural zeros
Generalized Heckman two-step
Generalized nonlinear least squares
PPML

ABSTRACT

The gravity model of international trade is often applied by economists to explain bilateral trade between countries. Nevertheless, some estimation practices have been subject to criticism, namely how zero trade values and the heteroskedasticity are handled. This paper proposes new nonlinear estimation techniques to address these issues. In particular, we propose standard and generalized versions of the nonlinear Heckman two-step approach that do not require the log-linearization of the gravity equation and corrects for non-random selection bias, and a generalized nonlinear least squares estimator that can be viewed as an iterative version of the normal family Quasi-Generalized Pseudo-Maximum-Likelihood estimator. Monte Carlo simulations show that our proposed estimators outperform existent linear and nonlinear estimators and are very efficient in correcting the selection bias and reducing the standard deviation of the estimates. Empirical results show that previous studies have overestimated the contribution of variables such as importer's income, distance, remoteness, trade agreements, and openness.

1. Introduction

The gravity model of international trade predicts that the trade flows between two countries depend mainly on their economic sizes (measured by the gross domestic product, GDP) and the distance between them.[1] Broadly speaking, the gravity equation is built around a nonlinear relationship between trade flows and a set of explanatory variables. The strong theoretical foundations of the gravity model (e.g., Anderson, 1979; Anderson and van Wincoop, 2003) mean that it is adequate for explaining between-country trade flows. Nevertheless, some estimation practices have been subject to criticism. For instance, for estimation purposes, the gravity equation is usually log-linearized and then estimated by ordinary least squares (OLS). This technique suffers from two major issues: the way the log-linearized models treat zero bilateral trade values and the potential Jensen's inequality bias. For instance, Lin (2013) shows that the failure of OLS estimation of log-linearized gravity models to correct for the Jensen's inequality issue and the omission bias have led to a counter intuitive result of increasing negative impact of distance on trade over time.

Two methods have been widely used to deal with the first issue: (i) using censored data, in which only observations with non-zero trade flows are used for the estimation; or (ii) augmenting the dependent variable (i.e., trade) by 1 to avoid undefined cases of the log of zero. Excluding zero values could represent a substantial loss of information, as it would imply a significant reduction in the sample size, which can reach more than 50% in some cases. Obviously, augmenting the dependent variables by 1 could yield biased and inefficient estimators caused by model misspecification.

The second issue is related to the Jensen's inequality bias that could potentially arise when estimating the log-linearized version of the gravity model. Jensen's inequality suggests that the expected value of the logarithm of any random variable will not equal the logarithm of its expected value, but it will depend on the random variable's mean and variance. Hence, if the variance of the error term of the multiplicative gravity equation is correlated with the regressors, the estimation will be biased. In other words, OLS would yield unbiased results only if the trade conditional variance is proportional to the square of the mean.[2] However, if the trade variance is proportional to the mean, for example, or to any

☆ We are grateful to the editor in charge of this paper, Professor Sushanta Mallick, and two anonymous reviewers for their insightful comments. We also would like to thank the participants at the Canadian Economic Association 2018 Annual Meetings for their valuable comments. We are responsible for any remaining errors.

* Corresponding author.
E-mail address: snechi@qu.edu.qa (S. Nechi).

[1] Other variables such as price level and exchange rate have been shown to account for a significant amount of variance in the basic gravity equation.

[2] In this case, only the intercept estimate is biased.

https://doi.org/10.1016/j.econmod.2020.12.011
Received 4 February 2020; Received in revised form 12 December 2020; Accepted 13 December 2020
Available online 24 December 2020

other combination of the mean and its square, estimating the gravity model with the standard techniques that use log-transformed data could yield significantly biased results.

Santos Silva and Tenreyro (2006) (hereafter called SST (2006)) criticized the conventional estimation practices of log-linearized gravity trade models and proposed solutions for dealing with heteroscedasticity and zero trade values. Broadly, SST (2006) stated that the gravity equation should be estimated in its multiplicative form via a Poisson pseudo-maximum-likelihood (PPML) technique. SST (2006) showed that this method was robust to different patterns of heteroscedasticity and dealt with Jensen's inequality. The work of SST (2006) raised important questions about the findings of many seminal studies (e.g., Anderson and van Wincoop, 2003) that the predicted coefficients of GDP to be close to one. The improvements of gravity models by the PPML method made it more tractable and therefore, the PPML has been used extensively for estimating gravity equations (Bosquet and Boulhol, 2015; Egger and Tarlea, 2015; Dai et al., 2014; Lin, 2013; Yotov, 2012; de Sousa, 2012; Egger and Larch, 2011; Head et al., 2010; Shepherd, 2010; Fitzgerald, 2008; Tenreyro, 2007; among others).[3]

Despite the robustness of the PPML, some issues may persist. Firstly, as mentioned by SST (2006), although PPML yields consistent estimates regardless of the heteroscedasticity pattern, it is only efficient when the trade variance is proportional to the trade mean. Thus, inefficiency issues would partially persist if the true trade variance is not perfectly in line with the previous assumption. Secondly, although PPML allows estimating data with zero trade values, it does not explicitly model them. This makes PPML susceptible to mis-specification and sample selection bias particularly if zero trade values are correlated with gravity equation regressors. In this paper, we consider a set of nonlinear estimators that attempt to deal with the prevalence of zero values and heteroscedasticity, and we examine the performance of these estimators for situations with and without censoring. In addition to the PPML, our estimators include nonlinear Heckman Two-Step Least Squares (Heckman_2stp), and Nonlinear Least Squares (NLS). We also propose new generalized forms of NLS and Heckman_2stp estimators (GNLS and GHeckman_2stp, respectively). These generalized estimators allow for different trade variance processes: constant, proportional to the trade mean, proportional to the square of the trade mean or a combination of those scenarios. When correctly implemented, the new GNLS should, in principle, outperform the PPML, as the former allows for different variance process scenarios, whereas the latter is only optimal with the case where variance is proportional to the trade mean. The new GNLS can be viewed as an iterative version of the normal family Quasi-Generalized Pseudo-Maximum Likelihood introduced by Gourieroux et al. (1984).

To the best of our knowledge, previous studies using the Heckman two-step least squares technique to estimate the gravity model had to log-linearize the gravity model and estimate it linearly. In this paper, in addition to dealing with heteroscedasticity, we estimate the Heckman two-step least squares in a nonlinear manner (multiplicative form) and hence avoid the Jensen's inequality bias that would arise from the log-linearization of the gravity model. Moreover, using Heckman two-step technique should resolve, or at least minimize, any censoring bias that could arise when truncation is correlated with gravity model's explanatory variables. One could argue that, in such case, we can simply estimate the whole sample (including zero trade values) with the standard PPML. Yet, results can still be biased if zero trade observations correspond to pairs for which estimated trade values are small but not very close to zero.

In light of the performance of all the nonlinear estimators, we discuss their efficiency and their success in dealing with the issues highlighted earlier. For instance, by estimating the multiplicative form of the gravity equation (e.g., PPML, NLS, etc.), we can avoid Jensen's inequality bias but the estimators will not be fully efficient, given that heteroscedasticity is not fully addressed.

The simulation results suggest that the PPML yields unbiased and efficient estimates only when the simulated data do not include structural zeros. The results also show that the new GNLS slightly outperforms the PPML when data generator does not produce structural zeros. Finally, and most importantly, according to our Monte Carlo simulations, GHeckman_2stp estimator turns to be the best among all nonlinear estimators considered in this study. In fact, GHeckman_2stp was remarkably efficient in correcting for the sample selection bias as well as in reducing the standard deviations of the nonlinear Heckman estimates.

After evaluating the performance and efficiency of the nonlinear estimators, the next step in our methodology consists in estimating the gravity model with the real data and comparing the estimates of our new nonlinear estimators with those of PPML. The empirical results suggest that previous studies might have overestimated the contribution of many variables in the bilateral trade between countries.

This paper is structured as follows. Section 2 provides a brief review of the literature and Section 3 introduces the new nonlinear estimators and describes our approach for reporting robust standard errors. In section 4, we describe our data generating processes and report the Monte Carlo simulation results. Section 5 provides new estimates of the gravity equation and Section 6 concludes the paper.

2. Brief literature review

Estimating the coefficients of a gravity model has led to alternative methods based on linear and nonlinear least squares (OLS and NLS), a set of pseudo-maximum likelihood models and the Heckman two-step procedure. Here, we review the previous literature on gravity models and discuss the proposed estimation methods, the main findings and the drawbacks.

Arvis and Shepherd (2013) argued that, in addition to handling heteroscedasticity and zero trade flows, the PPML is the only estimator that produces estimates such that the total actual trade flows equals the sum of the fitted values. This finding confirms the superiority of the PPML over other gravity equation estimators. Although the PPML estimator outperformed those obtained with traditional methods (OLS, NLS etc.), it has been subjected to some criticism for problems of overdispersion and excess zero flows (Martin and Pham, 2019; Burger et al., 2009; Martínez-Zarzoso, 2013; Krisztin and Fischer; 2015). Many alternative estimation methods to the PPML have been proposed (the gamma pseudo-maximum likelihood (GPML), feasible generalized least squares, Eaton–Tamura Tobit, the negative binomial pseudo-maximum likelihood (NBPML) model, the zero-inflated Poisson pseudo-maximum likelihood model, the zero-inflated negative binomial pseudo-maximum likelihood model, etc.).

Krisztin and Fischer (2015) claimed that using PPML to estimate the gravity equation produces consistent but biased parameter estimates if spatial dependence between origin–destination pairs is ignored. On the basis of Monte Carlo simulations, Martínez-Zarzoso (2013) compared the PPML estimator with GPML, a NLS estimator and a feasible generalized least squares estimator by using three different datasets. Martínez-Zarzoso (2013) found that PPML was less affected by heteroscedasticity but did not outperform the other estimators in terms of bias and standard error. She concludes that for any application, the selection of the most appropriate estimator requires a number of tests and depends on the characteristics of each dataset.

Gómez-Herrera and Baleix (2010) reviewed the literature on the methods of specifying and estimating the gravity equation. They used a gravity equation based on Anderson and van Wincoop (2003) and discussed the fit of different estimation procedures (panel regressions with fixed and random effects, OLS, and simple and panel Poisson methodologies) applied to a large dataset of bilateral exports for 47 countries over 1980–2002. They found that none of the estimators outperformed the others in all aspects.

[3] According to Google Scholar, SST (2006) paper has been cited 4856 times, as of January 2020.

Xiong and Chen (2014) estimated the gravity equation in the presence of sample selection and heteroscedasticity by using a two-step method of moments (TS-MM) estimator. Their Monte Carlo experiment showed that the TS-MM estimates were robust to several combinations of sample selection and heteroscedasticity. Additionally, the TS-MM estimator performed reasonably well even when the data generation process deviated from the standard TS-MM assumptions. However, it is important to note that the simulations conducted by Xiong and Chen (2014) to confirm the outperformance of TS-MM over other estimators were based on a DGP where zeros are generated in a perfectly consistent way with the selection equation of the TS-MM. Hence, it is not surprising that their proposed estimator would outperform other candidates including PPML. However, as we will discuss later, structural zeros in our simulations are generated via a microfounded DGP that is different from our Heckman selection equation. This would increase the credibility of our comparative efficiency evaluation of the estimators being investigated.

Sukanuntathum (2012) performed two-step gravity model estimations under heteroscedasticity and data censoring with several estimators. He showed that the NBPML was robust to different forms of heteroscedasticity and is well able to manage zero flows. Burger et al. (2009) estimated a gravity model by using a set of modified PPML models (NBPML, zero-inflated, zero-inflated Poisson pseudo-maximum likelihood, and the zero-inflated negative binomial). Their zero-inflated estimation technique provided a good alternative to log-normal and standard Poisson specifications of the gravity model under assumptions of heteroscedasticity. Gómez-Herrera (2013) compared methods of estimating Anderson and van Wincoop's (2003) model (truncated OLS, OLS (1+X), Tobit, Probit, Heckman, panel fixed, panel random and PPML) and found that the *ad hoc* methods were not suitable for estimating the gravity equation because they provide inefficient and biased estimates.

Assane and Chiang (2014) found that the OLS and Heckman two-step procedure models produce coefficient estimates for distance and other trade cost parameters that were higher than those of the PPML estimator, consistent with SST (2006). They concluded that although the Heckman two-step and PPML models are appropriate for zero trade flows, the PPML estimator is the only estimator that deals with heteroscedasticity problems.

The gravity model has been used in other applications where the estimation issues related to zero data values and heteroscedasticity coexist. For instance, Santana-Gallego et al. (2016) extend Helpman et al. (2008) framework to investigate the impact of tourism on trade. Using a data set with around 70% of the observations take the value of zero, Nguyen et al. (2020) test several empirical specifications and estimation methods to explain bilateral FDI flows between 31 Asian countries. They find that PPML is among the appropriate estimators in this context. Dorakh (2020) argued that one drawback of the current practice of estimating an FDI model is the presence of zero and negative values in the FDI data, concluding that PPML is a more appropriate method. Mallick and Marques (2017) use two-stage Heckman model to analyze the export pricing behavior of Indian and Chinese firms when zero values are due to selection issues into exporting. Their results indicate systematic differences in pricing strategies of Chinese and Indian exporters whilst discovering a selection bias in exports to high-income countries.

3. Nonlinear estimators of the gravity equation

Consider the econometric formulation of the traditional gravity equation:

$$y_{ij} = \beta_0 \left(\frac{x_i^{\beta_1} x_j^{\beta_2}}{D_{ij}^{\beta_3}} \right) \eta_{ij}, \tag{1}$$

where y_{ij} represents the trade flows from country i to country$j$; x_i *and* x_j represent the GDP for countries i and j, respectively; D_{ij} is the distance from country i to country j; β_1, β_2 and β_3 are unknown parameters and η_{ij} is an error term with an expectation, $E(\eta_{ij}|x_i,x_j,D_{ij})$, of one that is assumed to be statistically independent of the regressors.

To estimate this model, traditional approaches in the trade literature start by log-linearizing Equation (1), then estimating the parameters of interest through least squares with Equation (2):

$$\ln(y_{ij}) = ln\beta_0 + \beta_1 ln\, x_i + \beta_2 ln\, x_j - \beta_3 ln\, D_{ij} + ln\, \eta_{ij}, \tag{2}$$

Two major issues are associated with this approach. First, complications associated with observations for the dependent variable with zero trade values can arise. Second, estimating Equation (2) with OLS can produce biased estimators if $var(\eta_{ij})$ is correlated with the regressors. In other words, OLS would produce unbiased results only in the special case where the variance of bilateral trade is proportional to the square of the trade mean. However, if the trade variance is an increasing function of the trade level but not necessarily proportional to the square of the mean, the variance of the error terms ($\ln \eta_{ij}$) and regressors ($x_i,\ x_j,\ or\ D_{ij}$) in Equation (2) would be correlated. In this case, OLS results will be certainly biased. Unfortunately, many studies neglect this fact, leading to a bias related to Jensen's inequality.

Many estimators have been proposed to estimate the traditional gravity equation. In the remainder of this section, we present a set of selected classical nonlinear estimators: PPML and NLS. We also introduce the nonlinear Heckman two-step (Heckman_2stp) and our new generalized nonlinear versions of NLS and Heckman_2stp, namely, GNLS and GHeckman_2stp.

3.1. NLS and PPML

Following SST (2006), the multiplicative gravity equation can be written as the following exponential function:

$$y_{ij} = \exp\left[ln\beta_0 + \beta_1 ln\, x_i + \beta_2 ln\, x_j - \beta_3 ln\, D_{ij}\right]\eta_{ij}, \tag{3}$$

where η_{ij} is a log-normal random variable with a mean of 1 and a variance of σ_i^2.

The objective function of the NLS estimator can be formulated as follows:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} [y_i - \exp(x_i\beta)]^2, \tag{4}$$

The FOC derived from this objective function can be written as:

$$\sum_{i=1}^{n} [y_i - \exp(x_i\widehat{\beta})]\exp(x_i\widehat{\beta})x_i = 0, \tag{5}$$

Clearly, NLS is optimal only when the trade variance is constant. However, trade variance is very likely to be an increasing function of the trade level. In this case, NLS will be overweighting nosier observations. Thus, we can conclude that although NLS estimates of the gravity equation are asymptotically consistent, they may be very inefficient.

SST (2006) argued that we can get an estimator that is more efficient if we follow McCullagh and Nelder (1989) and by estimating the parameters of interest with a Poisson pseudo-maximum likelihood estimator (PPML). In addition to be optimal when the conditional variance is in proportion to $\exp(x_i\beta)$, SST (2006) argue that PPML is robust to many other variance scenarios. The proposed estimator of SST (2006) consists of dividing the NLS FOC by the conditional variance, which yields the following FOC:

$$\sum_{i=1}^{n} [y_i - \exp(x_i\widetilde{\beta})]x_i = 0, \tag{6}$$

The advantages of the PPML can be summarized as follows. First, PPML avoids the Jensen's inequality bias, as it estimates the nonlinear form and allows the inclusion of zero observations. Second, the simulation exercises of SST (2006) and subsequent empirical investigations (including the present study) show that PPML is relatively robust to different variance process scenarios.[4] Third, the fact that PPML is already available in many econometric software packages (STATA, for instance) makes it a very attractive workhorse estimator for estimating gravity models.

3.2. The nonlinear Heckman two-step

Tobin (1958) had shown that zero values of the dependent variable could create potentially large biases in parameter estimates, even in linear models, if the estimator used does not allow for this feature of the data generating process. Hurd (1979) shows that estimations of gravity models with truncated data (exclusion of zero values) would result in large biases. Many studies attempted to deal with the prevalence of zero values. However, the most influential and significant work was conducted by Heckman (1979). Broadly speaking, Heckman (1979) generalized Tobin's (1958) approach to estimation in the presence of this problem based on non-random sample selection. A nonlinear version of Heckman's formulation in the context of our gravity model is presented as:

$$\begin{cases} y_{1i}^{*} = exp(x_i\beta) + u_{1i}, \\ y_{2i}^{*} = x_i\alpha + u_{2i}, \end{cases} \tag{7 and 8}$$

$$y_{1i} = y_{1i}^{*} \text{ if } y_{2i}^{*} > 0, \tag{9}$$

$$y_{1i} = 0 \text{ if } y_{2i}^{*} \leq 0, \tag{10}$$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim BN\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}\right]$$

Where x_{ji} is a vector of exogenous regressors, β and $\alpha$are $K\times 1$ vector of parameters. According to Heckman, Equation (7) is a behavioral equation of interest and Equation (8) is a sample selection rule (could be Probit type that describes the propensity to trade) that determines whether observations have a non-zero value. In principle, y_1^* and y_2^* are unobserved while y_1 is.

The error terms u_1 and u_2 can be expected to be positively correlated and they are commonly assumed to have bivariate normal distribution.

We estimate our sample selection model using Heckman two-step approach. In the first step, we estimate a Probit selection equation via maximum likelihood and we compute the Inverse Mills Ratio (IMR) for each country pair, denoted λ_{ij} (also known as Heckman's lambda) based on the Probit estimates. Formally, for the probability of trade, the Probit regression can be written as follows:

$$Prob(T = 1|x) = \Phi(x\alpha), \tag{11}$$

where T is a dummy for the occurrence of trade ($T = 1$ if trade occurs and $T = 0$ otherwise), x is a vector of explanatory variables, α is a vector of unknown parameters and Φ is the cumulative distribution function of the standard normal distribution. Estimation of this model yields the so-called Inverse Mill Ratio, λ:

$$\lambda(x\alpha) = \frac{\varphi(x\alpha)}{\Phi(x\alpha)}, \tag{12}$$

Where φ is the probability density function of the standard normal distribution.

In the second step, to correct for the non-random selection, we include Heckman's lambda as additional regressor in the estimation of the gravity equation:

$$y_{1i} = \exp(x_i\beta) + \lambda(x_i\widehat{\alpha})\theta + \upsilon_1, \tag{13}$$

θ is a vector of unknown parameters. Presented this way, sample selection can be viewed as a form of omitted-variables bias, conditional on both x and λ as if the sample is randomly selected. As long as u_1 has a normal distribution and υ_1 is independent from λ, Heckman's two-step estimator is consistent (Puhani, 2000). Hence, Heckman_2stp allows to estimate the multiplicative form of the gravity equation, which avoids the Jensen inequality bias that could arise from the log-linearization. Equation (13) is estimated by minimizing the following sum of squared residuals:

$$\underset{\beta,\theta}{\text{argmin}} \sum_{i=1}^{n} [y_{1i} - \exp(x_i\beta) - \lambda(x_i\widehat{\alpha})\theta]^2.$$

To the best of our knowledge, most of the previous studies that adopted Heckman two-step technique to estimate the gravity model had to log-linearize it. Although Helpman et al. (2008) used nonlinear least squares to estimate a two-step Heckman, the estimated equation was a log-linearized version of the gravity equation with an additional nonlinear term (equivalent to IMR) capturing the fraction of firms that participate in the bilateral trade.[5] Even though Xiong and Chen (2014) estimate a nonlinear Heckman model, their proposed estimator, TS-MM, does not allow for different variance processes as our Generalized Heckman_2sls that we will introduce in the next section. Moreover, the TS-MM is a standard method of moments estimator where the exogenous matrix (Ω) of the moment condition (equation (9) in Xiong and Chen (2014)) consists simply of the gravity equation regressors (x) and the IMR (λ).[6] Obviously, there is nothing to guarantee that the choice of Ω would lead to an efficient estimator. In fact, the FOC's of TS-MM can be viewed as an augmented version of the PPML FOC's, where IMR is added linearly as an extra explanatory variable.[7]

3.3. Generalized nonlinear estimators

In addition to the PPML and NLS (complete and censored data, i.e. $y > 0$) and the Heckman_2stp, we also estimate new generalized versions of NLS and Heckman_2stp.[8]

Consider the gravity model where errors terms are entered additively:

$$y_i = f(x_i, \beta, \theta) + \varepsilon_i, \quad \text{for} \quad i = 1, \ldots.., n \tag{14}$$

where ε_i has a mean zero and a conditional variance: h_i^2 .

For NLS, $f(x_i, \beta, \theta) = exp(x_i\beta)$, while for Heckman_2stp, $f(x_i, \beta, \theta) = \exp(x_i\beta) + \theta\widehat{\lambda}(x_i\widehat{\alpha})$, where $\widehat{\lambda}(x_i\widehat{\alpha})$ is the IMR already calculated from estimates of the Probit model in the first step.

We first estimate the constant trade variance model and we store the

[4] When data generators do generate structural zeros.

[5] Santos Silva and Tenreyro (2015) argued that the Helpman et al. (2008) model is only valid under the assumption of constant variance. These criticisms do not apply to our nonlinear Heckman two-step.

[6] These moment conditions are obviously different from those of our proposed Heckman estimator described in equation (13) where the FOC's for β are $\sum_{i=1}^{n}[y_{1i} - \exp(x_i\beta) - \lambda(x_i\widehat{\alpha})\theta]x_i \exp(x_i\beta) = 0$ and those of θ are $\sum_{i=1}^{n}[y_{1i} - \exp(x_i\beta) - \lambda(x_i\widehat{\alpha})\theta]\lambda(x_i\widehat{\alpha}) = 0$.

[7] Even in the case where the trade variance is proportional to the trade mean ($\propto \exp(x\beta)$), we doubt that the TS-MM would be optimal. The reason is that the variance of the adjusted error term ($y - \exp(x\beta) - \omega\lambda$) for the lowest positive trade values (close to zero such that λ is different from 0) would be asymptotically different from the variance of the original gravity equation error.($y - \exp(x\beta)$).

[8] Estimation was performed in Matlab, given the efficiency and the flexibility of the software in dealing with nonlinear optimization. Codes are available upon request.

estimated coefficients $(\widehat{\beta},\widehat{\theta})$ and the residuals $\widehat{\varepsilon}$, where:

$$(\widehat{\beta},\widehat{\theta}) = \underset{\beta,\theta}{\operatorname{argmin}} \sum_{i=1}^{n} [y_i - f(x_i,\beta,\theta)]^2 \quad \text{and} \quad \widehat{\varepsilon}_i = y_i - f(x_i,\widehat{\beta},\widehat{\theta}), \tag{15}$$

The generalized versions are then obtained by running the following loop:[9]

i. Run a regression of squared residuals ($\widehat{\varepsilon}^2$) against a constant term, $\exp(x\widehat{\beta})$ and the square of $\exp(x\widehat{\beta})$:

$$\widehat{\varepsilon}^2 = |\omega_0| + |\omega_1|\exp(x\widehat{\beta}) + |\omega_2|[\exp(x\widehat{\beta})]^2 + \psi, \tag{16}$$

Where ω_0, ω_1,ω_2 are the variance process coefficients expressed in absolute value to avoid negative predicted variance levels. ψ is an error term.

ii. Use Equation (16) to estimate the individual conditional variance $\widehat{h}_i$, such that:

$$\widehat{h}_i = |\widehat{\omega_0}| + |\widehat{\omega_1}|\exp(x_i\widehat{\beta}) + |\widehat{\omega_2}|[\exp(x_i\widehat{\beta})]^2, \tag{17}$$

iii. Estimate the new coefficients$\tilde{\beta}$ and $\tilde{\theta}$ by minimizing the following sum of squared variance-corrected residuals:

$$(\tilde{\beta},\tilde{\theta}) = \underset{\beta,\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \left[\frac{y_i - f(x_i,\beta,\theta)}{\widehat{h}_i}\right]^2, \tag{18}$$

iv. Update estimates: $\widehat{\beta} = \tilde{\beta}$; $\widehat{\theta} = \tilde{\theta}$ and $\widehat{\varepsilon}_i = y_i - f(x_{i,} \tilde{\beta},\tilde{\theta})$.

v. Repeat steps (i-iv) until $(\tilde{\beta},\tilde{\theta})$ converge to $(\widehat{\beta},\widehat{\theta})$.

These generalized estimators allow for different trade variance processes: constant, proportional to the trade mean, proportional to the square of the trade mean or a combination of these scenarios. Note that since our new GNLS is comparable to the feasible generalized least squares, the variance coefficients estimators ($\widehat{\omega_0}$, $\widehat{\omega_1}$and $\widehat{\omega_2}$)are asymptotically consistent when $\widehat{\beta}$ and $\widehat{\theta}$ are consistent estimators of the true β and θ (Davidson and Mackinnon, 2003; Amemiya, 1973). It is worth noting that our GNLS is similar to the GLS introduced by Delgado (1992) and Delgado and Kniesner (1997). However, the latter is based on nonparametric Nearest Neighbor (NN) estimates of the conditional variance process. One major drawback of this nonparametric generalized least squares estimator is the problem of choosing the number of nearest neighbors in an optimal way. Moreover, as highlighted by SST (2006), the implementation of this estimator to estimate the gravity equation is very challenging given the high number of regressors.[10]

This estimation of the generalized versions constitutes one of the main contributions of this paper. In fact, the new GNLS should, in principle, outperform the PPML as the former allows for more variance process scenarios than the latter, which is only optimal when the variance is proportional to$\exp(x\beta)$. It is interesting to note that the new GNLS is an iterative version of the normal family Quasi-Generalized Pseudo-Maximum Likelihood method introduced by Gourieroux et al. (1984). Table 1 describes all nonlinear estimators considered in this paper.

[9] To estimate these variance equations, we need to use a nonlinear least squares procedure, as the coefficients are in terms of absolute value. A significant challenge was to find good starting values in order to avoid convergence issues. The Matlab function "fminsearch" was very efficient at getting very good initial values.

[10] Since the nonparametric weights would depend on the Euclidean distance between regressors.

Table 1
Nonlinear estimators.

Estimator	Notation
Nonlinear Heckman two-step least squares	Heckman_2stp
Generalized Nonlinear Heckman two-step least squares	GHeckman_2stp
Nonlinear least squares	NLS
Nonlinear least squares – censored	NLS_C
Generalized nonlinear least squares	GNLS
Generalized nonlinear least squares – censored	GNLS_C
Poisson pseudo-maximum likelihood	PPML
Poisson pseudo-maximum likelihood – censored	PPML_C

3.4. Robust standard errors

In this section, we explain how we compute heteroskedasticity-robust standard errors for our new nonlinear estimators. Recall that the estimation of NLS and Heckman_2stp consists in minimizing the following sum-of-squared residuals (SSR) function:$Q(\beta) = \sum_{i=1}^{n} [y_i - f(x_i,\beta,\theta)]^2$. For GNLS and GHeckman_2stp, the criterion function to be minimized is simply the weighted SSR function: $Q(\beta) = \sum_{i=1}^{n} \left[\frac{y_i - f(x_i,\beta,\theta)}{\widehat{h}_i}\right]^2$, where $\widehat{h}_i$ is the estimated standard deviation of the i^{th}observation. Given that all these nonlinear estimators are obtained by minimizing a sum-of-squares function, a consistent estimate of the covariance matrix of (β,θ) can be computed by the means of Gauss-Newton regression (GNR). Following Davidson and Mackinnon (2003), we can show that a reasonable way of estimating a heteroscedasticity-consistent covariance matrix (HCCME) estimator of (β,θ) is to use the sandwich covariance matrix (HC1):

$$\left(\widehat{X}'\widehat{X}\right)^{-1}\widehat{X}'\widehat{\Omega}\widehat{X}(\widehat{X'X})^{-1}, \tag{19}$$

where $\widehat{\Omega}$ is a $n \times n$ diagonal matrix with the squared weighted residual $\frac{n}{n-k}\left(\frac{\widehat{\varepsilon}_i}{\widehat{h}_i}\right)^2$ as thet^{th} diagonal element, where $\frac{n}{n-k}$ is included to correct for the typically small size least-squares residuals. $\widehat{X}$is a $n \times K$ matrix with a typical $(i,j)^{th}$ element corresponding to an estimate of the derivative of $f(x_i,\beta,\theta)$with respect to β_j(or also with respect to θ for the Heckman_2stp and GHeckman_2stp) and divided by the estimated standard deviation $\widehat{h}_i$ for the case of generalized versions.[11]

4. Simulations

4.1. Data generation processes

We conducted different experiments to allow for different data generation processes (DGPs). We considered three dimensions in each DGP: the distribution of the errors (normal vs. log-normal), the frequency of censoring (different values of k) and five different patterns of heteroscedasticity. For all experiments, we adopt the multiplicative model used by SST (2006):

$$E[y_i|x] = exp(x_i\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}); i = 1, \ldots.., 10000, \tag{20}$$

where x_{1i} is the standard normal variable that captures the behavior of continuous explanatory variables (e.g., income levels and distance), whereas x_{2i} is a binary dummy that captures variables such as border and free trade agreements, and equals 1 with a probability of 0.4. A set of observations for each variable is generated in each replication as $\beta_0 = 0$, $\beta_1 = \beta_2 = 1$. Data on bilateral trade, y, are generated from the following

[11] For GNLS, $\widehat{X}_{i,j}$ is equal to $\frac{x_{ij}\exp(x_i\widehat{\beta})}{\widehat{h}_i}$, while for the GHeckman_2stp, $\widehat{X}_{i,j}$ is equal to $\frac{x_{ij}\exp(x_i\widehat{\beta})}{\widehat{h}_i}$ or to $\frac{\widehat{\lambda}(x_i\widehat{\alpha})}{\widehat{h}_i}$.

Table 2
Simulation results of the Eaton Tamura Data Generating Process under Different Forms of Heteroscedasticity and Censoring Levels.

Distribution		Normal Errors								Log-Normal Errors							
Variance of μ	Estimator	k = 0				k = 1				k = 0				k = 1			
		β1		β2		β1		β2		β1		β2		β1		β2	
		bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
Case 1: $1/(\exp(.))^2$	NLS	−0.009475	0.010192	−0.012855	0.017917	0.149097	0.011213	0.153677	0.024106	0.000092	0.009777	−0.001151	0.020299	0.156634	0.011474	0.162301	0.024918
	GNLS	−0.013532	0.010768	−0.016049	0.018323	0.193703	0.021461	0.198924	0.031888	0.000346	0.009862	−0.000904	0.020582	0.178746	0.026331	0.183997	0.035827
	NLS_C	−0.023135	0.009798	−0.028622	0.017497	0.117217	0.011373	0.116009	0.023531	0.000092	0.009777	−0.001151	0.020299	0.125074	0.011964	0.125549	0.02464
	GNLS_C	−0.033355	0.011166	−0.038335	0.018225	0.128466	0.013116	0.126289	0.024757	0.000346	0.009862	−0.000904	0.020582	0.129395	0.015062	0.129379	0.026382
	Heckman_2stp	0.000261	0.010139	−0.000137	0.020185	0.12708	0.011912	0.127514	0.023972	0.000092	0.009777	−0.001151	0.020299	0.139065	0.013628	0.14201	0.025236
	Gheckman_2stp	−0.000121	0.010249	−0.000645	0.020465	0.147301	0.014717	0.147528	0.025587	0.000346	0.009862	−0.000904	0.020582	0.146573	0.016835	0.148969	0.027213
	PPML	−0.063867	0.01418	−0.059943	0.022566	0.231669	0.016629	0.24572	0.03269	−0.00114	0.019069	−0.000966	0.030545	0.273547	0.028834	0.290752	0.0465
	PPML_C	−0.124253	0.011482	−0.123557	0.019616	0.086472	0.015449	0.085241	0.02941	−0.00114	0.019069	−0.000966	0.030545	0.098231	0.033798	0.099699	0.048577
Case 2: $1/\exp(.)$	NLS	−0.006163	0.031655	−0.016842	0.053782	0.142418	0.037905	0.145455	0.073001	0.002652	0.031416	−0.000507	0.058176	0.158032	0.038225	0.161758	0.075359
	NLS_C	−0.031302	0.031374	−0.046018	0.051724	0.081452	0.039031	0.078523	0.069632	0.002652	0.031416	−0.000507	0.058176	0.113988	0.040893	0.110003	0.074869
	GNLS	−0.029088	0.02051	−0.033244	0.036461	0.229963	0.034358	0.235574	0.063682	0.000707	0.01985	−0.002491	0.044471	0.230651	0.034965	0.236076	0.067296
	GNLS_C	−0.091823	0.023176	−0.099798	0.033625	0.075017	0.022999	0.068921	0.049679	0.000707	0.01985	−0.002491	0.044471	0.118574	0.031737	0.110773	0.061754
	Heckman_2stp	−0.013369	0.036735	−0.016955	0.060524	0.084405	0.044126	0.081808	0.073303	0.002652	0.031416	−0.000507	0.058176	0.125617	0.043994	0.123536	0.077602
	Gheckman_2stp	−0.056072	0.022896	−0.061612	0.041653	0.087218	0.027448	0.081922	0.052258	0.000712	0.019848	−0.002459	0.044425	0.160714	0.037427	0.157816	0.066022
	PPML	−0.033014	0.019035	−0.035683	0.035582	0.254988	0.023664	0.263316	0.056814	0.000708	0.019592	−0.001869	0.043993	0.286378	0.028359	0.298563	0.06583
	PPML_C	−0.091769	0.016682	−0.099448	0.03369	0.073418	0.022434	0.067722	0.048877	0.000708	0.019592	−0.001869	0.043993	0.08716	0.030707	0.07874	0.062515
Case 3: 1	NLS	0.027737	0.121464	0.000588	0.17327	0.144665	0.141628	0.152544	0.216254	0.01983	0.145841	0.007972	0.21133	0.175338	0.183032	0.174121	0.269021
	NLS_C	0.016051	0.104809	0.002177	0.147055	0.073983	0.115998	0.07955	0.173924	0.01983	0.145841	0.007972	0.21133	0.08462	0.195316	0.076484	0.269281
	GNLS	0.004126	0.035453	−0.002013	0.064972	0.307932	0.114812	0.313139	0.134037	−0.004681	0.043149	−0.010916	0.08117	0.30151	0.101093	0.301984	0.147965
	GNLS_C	0.001052	0.033243	0.00044	0.052456	0.150864	0.052426	0.151866	0.074261	−0.004681	0.043149	−0.010916	0.08117	−0.000349	0.081262	−0.017527	0.129727
	Heckman_2stp	0.010259	0.12742	0.009494	0.171189	0.05202	0.152695	0.056064	0.202497	0.047786	0.323882	0.024478	0.276366	0.086508	0.207032	0.085054	0.283903
	Gheckman_2stp	0.000273	0.040058	0.000611	0.0524	0.080191	0.056622	0.079576	0.077325	−0.004541	0.044072	−0.01145	0.082331	0.055103	0.100984	0.037045	0.141603
	PPML	0.006084	0.045189	−0.004222	0.076038	0.26401	0.055522	0.27362	0.109295	0.002928	0.056386	−0.006466	0.098592	0.292494	0.072739	0.294846	0.143516
	PPML_C	0.002121	0.040207	−0.00141	0.062162	0.111266	0.051526	0.11393	0.086712	0.002928	0.056386	−0.006466	0.098592	0.037693	0.085354	0.021038	0.139457
Case 4: $1/\exp(.) + \exp(x2)$	NLS	0.024339	0.166932	0.153192	0.203628	0.106065	0.187123	0.308853	0.239775	0.013927	0.194697	−0.002161	0.261645	0.177659	0.346839	0.160177	0.349581
	NLS_C	0.005691	0.13363	0.275843	0.162113	0.034585	0.150482	0.350621	0.18896	0.013927	0.194697	−0.002161	0.261645	−0.018859	0.308919	0.081082	0.352854
	GNLS	−0.037065	0.04408	0.107976	0.071066	0.194854	0.07865	0.421702	0.114398	−0.013274	0.065926	−0.020706	0.121002	0.210839	0.113896	0.245534	0.183971
	GNLS_C	−0.065012	0.042287	0.200504	0.064193	0.042947	0.042948	0.353521	0.070287	−0.013274	0.065926	−0.020706	0.121002	−0.234909	0.230711	−0.05584	0.210848
	Heckman_2stp	0.002217	0.167117	0.284995	0.196706	0.030968	0.190879	0.343026	0.233789	0.01251	0.19237	−0.00742	0.260508	0.005874	0.33923	0.101992	0.371598
	Gheckman_2stp	−0.029991	0.053761	0.22024	0.06917	0.03079	0.058947	0.339595	0.083248	−0.013542	0.06808	−0.020179	0.122767	−0.079348	0.188153	0.084279	0.211043
	PPML	−0.026391	0.055338	0.113823	0.085288	0.166316	0.065501	0.392535	0.114948	0.000236	0.083652	−0.013409	0.135289	0.227392	0.10589	0.257798	0.179698
	PPML_C	−0.05097	0.046071	0.211297	0.06603	0.033451	0.057857	0.346193	0.089176	0.000236	0.083652	−0.013409	0.135289	−0.1323	0.125993	0.01562	0.177909
Case 5: $1/(\exp(.))^2 + 1$	NLS	0.013136	0.121606	−0.014804	0.171009	0.127802	0.14166	0.133166	0.213375	0.020262	0.144409	0.002267	0.217364	0.16592	0.177973	0.164562	0.268095
	NLS_C	−0.011834	0.106787	−0.027664	0.146532	0.051168	0.116053	0.053809	0.172384	0.020262	0.144409	0.002267	0.217364	0.064407	0.196657	0.057107	0.271371
	GNLS	−0.082596	0.074447	−0.083107	0.087122	0.173574	0.110114	0.182916	0.122883	−0.00351	0.059653	−0.015613	0.095815	0.239703	0.08454	0.234614	0.146029
	GNLS_C	−0.136441	0.079877	−0.131686	0.077654	−0.083536	0.201532	−0.062999	0.174836	−0.00351	0.059653	−0.015613	0.095815	−0.178339	0.308172	−0.19144	0.309387
	Heckman_2stp	0.014609	0.121959	0.013174	0.167363	0.061534	0.15176	0.06579	0.20526	0.0198	0.144135	0.003068	0.217906	0.102449	0.216071	0.09841	0.294665
	Gheckman_2stp	−0.009806	0.116451	−0.00176	0.106332	0.099058	0.226382	0.083459	0.489654	−0.008053	0.100406	−0.022835	0.125567	0.065668	0.175797	0.056253	0.204103
	PPML	−0.058604	0.046058	−0.066132	0.074044	0.172506	0.054858	0.177775	0.104043	0.002144	0.05843	−0.01199	0.103886	0.239796	0.074936	0.239873	0.14516
	PPML_C	−0.102954	0.042171	−0.10567	0.061305	0.02475	0.049408	0.022625	0.08272	0.002144	0.05843	−0.01199	0.103886	−0.049592	0.08949	−0.070367	0.142749

Table 3
Simulation results of the Head and Mayer Data Generating Process under Different Forms of Heteroscedasticity and Censoring Levels.

Distribution		Normal errors								Log-normal errors							
Variance of μ	Estimator	k = 0.5				k = 1.25				k = 0.5				k = 1.25			
		β1		β2		β1		β2		β1		β2		β1		β2	
		bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
Case 1: $1/(\exp(.))^2$	NLS	0.030872	0.010055	0.033348	0.020026	0.071138	0.010807	0.078251	0.021493	0.036494	0.01031	0.039191	0.021521	0.073587	0.011055	0.081308	0.02206
	NLS_C	−0.027348	0.010023	−0.031616	0.018485	−0.028538	0.010643	−0.033613	0.019195	−0.015237	0.010448	−0.018389	0.020811	−0.024494	0.011175	−0.028781	0.0205
	GNLS	0.033241	0.010743	0.035576	0.020654	0.085454	0.018611	0.092499	0.02742	0.037744	0.011455	0.040291	0.022114	0.077921	0.016253	0.08551	0.025426
	GNLS_C	−0.031092	0.010951	−0.035549	0.019206	−0.03147	0.011635	−0.036773	0.01982	−0.015393	0.010424	−0.01859	0.020866	−0.024576	0.011137	−0.028875	0.020515
	Heckman_2stp	0.002772	0.010608	0.00124	0.020374	0.00365	0.011066	0.002995	0.021289	0.00029	0.011386	−0.002056	0.020976	0.002703	0.013013	0.001817	0.022714
	Gheckman_2stp	0.00273	0.010779	0.001056	0.020605	0.003732	0.011108	0.002972	0.021514	0.000244	0.011426	−0.002132	0.021139	0.00273	0.013041	0.001818	0.022797
	PPML	0.085155	0.017084	0.087718	0.030952	0.203966	0.017909	0.213294	0.035608	0.113716	0.022776	0.115949	0.037657	0.211753	0.026047	0.22378	0.042452
	PPML_C	−0.121526	0.013696	−0.123897	0.023785	−0.110729	0.014617	−0.118718	0.025064	−0.064338	0.024769	−0.065535	0.035166	−0.098656	0.030377	−0.104149	0.040662
Case 2: $1/\exp(.)$	NLS	0.02997	0.032122	0.030569	0.056424	0.067211	0.032935	0.072146	0.059206	0.037829	0.031624	0.038455	0.060755	0.071812	0.032793	0.07678	0.060324
	NLS_C	−0.04159	0.034412	−0.048531	0.054339	−0.044916	0.036944	−0.052623	0.056839	−0.01366	0.034092	−0.0189	0.060016	−0.027233	0.037103	−0.034234	0.059593
	GNLS	0.079662	0.022852	0.080365	0.047996	0.150488	0.033763	0.15817	0.057888	0.081719	0.026509	0.081174	0.051939	0.134811	0.034382	0.143076	0.058483
	GNLS_C	−0.130313	0.03823	−0.134604	0.047686	−0.138288	0.044837	−0.147483	0.055383	−0.048279	0.026306	−0.054107	0.047738	−0.075525	0.033552	−0.083941	0.054863
	Heckman_2stp	−0.014937	0.037414	−0.018387	0.060832	−0.014304	0.039257	−0.017551	0.06304	0.001255	0.035536	−0.001179	0.062385	0.001764	0.038285	−0.000618	0.06556
	Gheckman_2stp	−0.061925	0.023557	−0.068884	0.0451	−0.061945	0.027049	−0.069213	0.049959	−0.009183	0.023649	−0.013558	0.049297	−0.011494	0.028564	−0.016053	0.054378
	PPML	0.096657	0.021289	0.096662	0.046829	0.205958	0.023374	0.212657	0.052432	0.117201	0.022692	0.115939	0.051133	0.212527	0.026112	0.221895	0.055246
	PPML_C	−0.112361	0.020225	−0.119062	0.038861	−0.116658	0.023549	−0.127706	0.042572	−0.059724	0.024228	−0.06438	0.048209	−0.096925	0.028928	−0.105011	0.051788
Case 3: 1	NLS	0.044822	0.156767	0.04002	0.217665	0.077982	0.158468	0.077695	0.225973	0.048663	0.149161	0.038251	0.212786	0.091337	0.198495	0.081692	0.239368
	NLS_C	−0.007755	0.105299	−0.014652	0.147805	−0.016302	0.112728	−0.024424	0.155316	−0.003547	0.159314	−0.018679	0.214639	−0.018906	0.170412	−0.037168	0.225042
	GNLS	0.135869	0.067746	0.129963	0.100637	0.238594	0.095877	0.238475	0.125065	0.14396	0.078773	0.1347	0.10893	0.238694	0.104029	0.237772	0.131575
	GNLS_C	−0.06105	0.031977	−0.058642	0.052342	−0.091052	0.039365	−0.094136	0.060652	−0.099345	0.055289	−0.098126	0.091178	−0.170685	0.074673	−0.174198	0.10917
	Heckman_2stp	0.013748	0.134028	0.013752	0.176147	0.013223	0.139739	0.013294	0.183565	0.047896	0.349383	0.030236	0.298963	0.045924	0.351654	0.02851	0.306683
	Gheckman_2stp	−0.024973	0.038981	−0.025182	0.055888	−0.048362	0.04768	−0.050697	0.062914	−0.040969	0.056242	−0.050995	0.092073	−0.070919	0.070382	−0.084294	0.105459
	PPML	0.098594	0.063267	0.095479	0.107228	0.198622	0.066905	0.20185	0.116742	0.118087	0.06154	0.112859	0.106497	0.212765	0.065189	0.215622	0.116961
	PPML_C	−0.035508	0.0469	−0.039719	0.071364	−0.054047	0.054877	−0.062498	0.080243	−0.057709	0.071303	−0.065064	0.10739	−0.096144	0.083156	−0.110063	0.121044
Case 4: $1/\exp(.) + \exp(x2)$	NLS	0.054049	0.272773	0.052558	0.305227	0.08127	0.26667	0.090288	0.30817	0.041063	0.207323	0.031265	0.267677	0.076727	0.223887	0.073655	0.281307
	NLS_C	−0.021391	0.139217	**0.256561**	0.166969	−0.027309	0.147513	0.250556	0.176249	−0.009338	0.229345	−0.022135	0.286642	−0.029598	0.247347	−0.045379	0.297235
	GNLS	0.083692	0.072925	0.123127	0.128709	0.184153	0.095521	0.231189	0.147509	0.1081	0.099063	0.107178	0.139522	0.185153	0.121691	0.197202	0.162899
	GNLS_C	−0.145847	0.047504	0.11565	0.067014	−0.164749	0.053141	0.096438	0.073251	−0.139782	0.107057	−0.126574	0.146295	−0.242296	0.159999	−0.221778	0.182288
	Heckman_2stp	0.009758	0.176397	0.295281	0.203195	0.010183	0.181614	0.297824	0.211219	0.093097	0.544207	0.056481	0.427599	0.089549	0.545221	0.054255	0.434528
	Gheckman_2stp	−0.057416	0.058018	0.199657	0.07421	−0.077329	0.064525	0.184099	0.080755	−0.052001	0.088895	−0.058287	0.13694	−0.082514	0.112665	−0.085961	0.15248
	PPML	0.052432	0.097872	0.077404	0.148852	0.146683	0.102357	0.188374	0.157744	0.100822	0.092026	0.099473	0.143466	0.187077	0.097313	0.198648	0.154309
	PPML_C	−0.077133	0.055177	0.185574	0.076686	−0.088686	0.064138	0.173362	0.086501	−0.078905	0.108589	−0.082686	0.148948	−0.134554	0.12781	−0.139274	0.166984
Case 5: $1/(\exp(.))^2 + 1$	NLS	0.037883	0.157358	0.032268	0.217146	0.07195	0.158742	0.070582	0.225352	0.053556	0.195005	0.040447	0.233175	0.079593	0.153475	0.072149	0.220221
	NLS_C	−0.029495	0.106378	−0.03835	0.146537	−0.033916	0.113335	−0.043698	0.154299	0.000996	0.204957	−0.016955	0.23639	−0.023926	0.171275	−0.04151	0.225496
	GNLS	0.069584	0.061659	0.070627	0.100871	0.190227	0.079252	0.191826	0.115273	0.120459	0.0794	0.113463	0.114221	0.207143	0.108367	0.144585	1.386992
	GNLS_C	−0.302273	0.187291	−0.271865	0.1618	−0.275647	0.150376	−0.260418	0.136717	−0.126998	0.092059	−0.124622	0.127504	−0.250682	0.242606	−0.235877	0.255517
	Heckman_2stp	0.018273	0.128118	0.018539	0.1724	0.017263	0.134953	0.017858	0.180072	0.050171	0.358768	0.032276	0.303604	0.050166	0.358977	0.033185	0.310259
	Gheckman_2stp	−0.021186	0.267953	−0.010039	0.165875	−0.027282	0.090899	−0.027415	0.102749	−0.034175	0.073007	−0.044087	0.115777	−0.064013	0.105415	−0.072968	0.137927
	PPML	0.059634	0.063798	0.05735	0.107328	0.167221	0.067024	0.16876	0.116745	0.104198	0.063369	0.099324	0.110006	0.193281	0.067887	0.196142	0.120722
	PPML_C	−0.111641	0.048026	−0.11665	0.070722	−0.111262	0.055487	−0.12214	0.079701	−0.075535	0.074073	−0.081826	0.111602	−0.12659	0.087714	−0.139934	0.126485

main equation:

$$y_i = exp(x_i\beta)\eta_i, \tag{21}$$

where η_i is a log-normal random error with a mean of 1 and a variance of σ_i^2. Equivalently, data can be also normally simulated and generated as: $y_i = exp(x_i\beta) + \varepsilon_i$, where ε_i is normally distributed with a mean of 0 and a variance of $(exp(x_i\beta))^2\sigma_i^2$.[12] Negative values of y that can be potentially generated in this case are replaced with 0.

In addition to the four patterns of heteroscedasticity used by SST (2006), we introduce an additional pattern in our sensitivity analysis. The five cases are:

i. Case 1: $\sigma_i^2 = [exp(x_i\beta)]^{-2}$; $V[y_i|x] = 1$.
ii. Case 2: $\sigma_i^2 = [exp(x_i\beta)]^{-1}$; $V[y_i|x] = \exp(x_i\beta)$.
iii. Case 3: $\sigma_i^2 = 1$; $V[y_i|x] = [\exp(x_i\beta)]^2$.
iv. Case 4: $\sigma_i^2 = [\exp(x_i\beta)]^{-1} + \exp(x_{2i})$; $V[y_i|x] = \exp(x_i\beta) + \exp(x_{2i}) \times [\exp(x_i\beta)]^2$.
v. Case 5: $\sigma_i^2 = 1 + [exp(x_i\beta)]^{-2}$; $V[y_i|x] = 1 + \frac{1}{\exp(x_i\beta)^2}$.

We design two data generating processes. The first one is consistent with the threshold Tobit of Eaton and Tamura (1994); while the second data generating process is in line with Head and Mayer (2014).

4.1.1. The Eaton and Tamura data generating process

The Eaton-Tamura DGP (ET, hereafter) consists in adding a negative constant term, $-k$ to equation (21) to ensure the prevalence of a significant number of zero trade values:

$$y_i = \exp(x_i\beta)\eta_i - k, \tag{22}$$

Obviously, trade occurs only if a threshold level of potential trade is exceeded. In our simulations, we consider all possible combinations between the values of $k = (0, 1)$ and the distribution of the error terms (normal vs. log-normal). For normally simulated data, Equation (22) is equivalent to: $y_i = \exp(x_i\beta) - k + \varepsilon_i$. Negative values of y that can be generated even when $k = 0$ are replaced with 0. Obviously, when $k > 0$, the ET becomes inconsistent with the gravity equation specification. However, there are two reasons for considering this data generating process in our simulations. First, unlike the second DGP (described below), the ET allows us to generate data without structural zeros, when k is set to 0 and y is log-normally distributed. Second, reporting simulations results for ET with $k = 1$, would confirm that using this type of data generating process, already used in literature (see, for example, Martin and Pham, 2019), where positive trade values are generated from an equation that is not consistent with the standard gravity model, is not very useful in assessing the performance of estimators since all of these estimators would be mis-specified, by construction.

4.1.2. The Head and Mayer data generating process

Our second DGP is consistent with Head and Mayer (2014) (HM, hereafter), where data are generated via the following specifications:

$$y_{i,1} = \exp(x_i\beta)\eta_i, \tag{23}$$

$$y_{i,2} = \exp(x_i\beta) - k + \mu_i, \tag{24}$$

$$and \begin{cases} y_i = y_{i1},\ y_{i,2} > 0 \\ y_i = 0,\ y_{i,2} \le 0 \end{cases}, \tag{25}$$

Where μ_i is normally distributed with mean 0 and variance δ. We allow the error terms of the measure and selection equations to be correlated with each other (i.e., $corr(\eta,\mu) = \rho$). Different values for k are considered (0.5 and 1.25).

Clearly, the HM is perfectly consistent with the multiplicative form of the gravity equation. Moreover, the selection equation of the HM allows to generate structural zeros when the simulated potential trade mean is below a certain average threshold$k$. This is in line with Head and Mayer (2014), where exporting firms face market-specific entry costs so that zeros are generated if the predicted aggregate trade falls below a market-specific threshold. Hence, k in our HM can be interpreted as an average exporting fixed costs, and μ_i as the destination-specific random component of these costs.

4.2. Simulation results

Tables 2 and 3 provide the results of the ET and HM when the error terms follow normal and log-normal distributions for the five cases of the variance processes and considering different values of the parameter k. In each table, we report the bias of the two parameters of interest (β_1 and β_2) and the standard deviation of the estimators.

4.2.1. The ET results

Our first task was to match the results of Cases 1 to 4 of the experiments conducted by SST (2006) and to confirm their findings in relation to the superiority of PPML. Table 2 summarizes the simulation results of the ET.

As expected, when $k = 1$, all estimators yield significantly biased results. As highlighted earlier, this version of ET is inconsistent with the specification of the traditional gravity equation that does not include the parameter k.[13] It follows that this type of data generator might not be appropriate for assessing the performance of gravity equation estimators.

In the cases where simulated data do not include any zero values (lognormal errors with $k = 0$), all estimators yield unbiased results. GNLS performance is very similar to PPML, but with slightly lower standard deviations for variance cases 1, 2 and 3. One remarkable result of the ET simulations is the efficiency of the generalized estimators (GHeckman_2stp and GNLS) in significantly reducing the standard deviations of the estimates compared to Heckman_2stp and NLS, in particular for severe heteroskedasticity cases (3, 4 and 5). Moreover, for cases 3 and 5, the magnitude of the NLS and Heckman_2stp bias, although small, was further reduced with generalized versions.

It is interesting to note that although NLS yields relatively small bias, the standard deviations become relatively larger for severe heteroscedasticity of y (cases 3, 4 and 5). It is worth noting that in SST (2006) and others, although they report similar results for less severe heteroscedasticity cases, their NLS yields extremely large bias and standard deviations for more severe heteroscedasticity. We believe that the main reason behind this discrepancy is the impact of unrealistically large coefficients estimates on the reported average bias.[14] These extreme outliers are associated with convergence problems only encountered with NLS estimations in case of severe heteroscedasticity of y. This is not surprising since NLS assumes constant conditional variance of y. It seems that in many other papers, authors did not drop these extreme outliers in their NLS simulations exercise as we do here. In fact, when we do not

[12] Let's assume that $\eta_i = 1 + v_i$, where v_i has a mean 0 and variance σ_i^2, then $y_i = exp(x_i\beta)(1 + v_i) = exp(x_i\beta) + exp(x_i\beta)v_i$. Hence, y_i can be written as: $y_i = exp(x_i\beta) + \varepsilon_i$, where $\varepsilon_i = exp(x_i\beta)v_i$, with a zero mean and a variance equal to $(exp(x_i\beta))^2\sigma_i^2$. We choose to adopt the second presentation of the DG for the normally distributed simulations to avoid generating too many negatives simulated values of y. Obviously, this was not an issue for the lognormal case given the nature of the lognormal distribution where all values are by construction positive.

[13] According to the simulations results of OLS and log-linearized Heckman Two-Step Least Square estimations (not reported here, but available up on request), this issue becomes more pronounced when the estimation requires a logarithmic transformation, as the bias reaches high levels.

[14] Unrealistically large estimates refer to cases where $\widehat{\beta}_1$ is, for example, equal to 1000 for a true β_1 equal to 1.

Table 4
Estimation of the gravity equation.

Estimator	NLS	GNLS	NLS_C	GNLS_C	PPML	PPML_C	Heckman_2stp	GHeckman_2stp	Probit
Variable									
Constant	−45.0989***	−30.1638***	−45.085***	−28.7579***	−32.3261***	−31.5296***	−45.081***	−29.0386***	−16.583***
	(3.3792)	(1.4602)	(3.3832)	(1.4803)	(2.0595)	(2.161)	(3.3851)	(1.8458)	(0.7616)
Log of exporter's GDP	0.7378***	0.734***	0.7376***	0.7188***	0.7325***	0.7213***	0.7376***	0.7056***	0.4572***
	(0.0384)	(0.0253)	(0.0384)	(0.0242)	(0.0268)	(0.0268)	(0.0384)	(0.0304)	(0.0081)
Log of importer's GDP	0.8619***	0.7397***	0.8617***	0.7277***	0.7411***	0.7319***	0.8617***	0.7177***	0.3289***
	(0.041)	(0.0244)	(0.041)	(0.0237)	(0.0274)	(0.0279)	(0.041)	(0.0302)	(0.0074)
Log of exporter's per capita GDP	0.3957***	0.1775***	0.3953***	0.1805***	0.1567***	0.1544***	0.395***	0.1651***	0.1005***
	(0.1157)	(0.049)	(0.1158)	(0.0465)	(0.0533)	(0.0527)	(0.1158)	(0.0588)	(0.01)
Log of importer's per capita GDP	−0.0325	0.1638***	−0.0325	0.165***	0.135***	0.1327***	−0.0325	0.1605***	0.109***
	(0.0619)	(0.0419)	(0.0619)	(0.0397)	(0.0449)	(0.0445)	(0.0619)	(0.0506)	(0.0102)
Log of distance	−0.9237***	−0.8188***	−0.9234***	−0.8166***	−0.7838***	−0.7763***	−0.9233***	−0.8136***	−0.4495***
	(0.0725)	(0.0443)	(0.0725)	(0.0438)	(0.0546)	(0.0553)	(0.0725)	(0.0499)	(0.0224)
Contiguity dummy	−0.0813	0.1783	−0.0811	0.1768	0.1929*	0.2024*	−0.081	0.1705	−0.467***
	(0.1004)	(0.1345)	(0.1005)	(0.1358)	(0.1043)	(0.1052)	(0.1005)	(0.1379)	(0.1)
Common-language dummy	0.6894***	0.8582***	0.6894***	0.8872***	0.746***	0.7513***	0.6894***	0.9325***	0.3295***
	(0.085)	(0.1498)	(0.0851)	(0.148)	(0.1347)	(0.1342)	(0.0851)	(0.17)	(0.0372)
Colonial-tie dummy	0.0358	−0.097	0.0355	−0.1224	0.025	0.02	0.0355	−0.1944	0.1587***
	(0.1254)	(0.1559)	(0.1255)	(0.1534)	(0.1498)	(0.15)	(0.1255)	(0.1802)	(0.0395)
Landlocked exporter dummy	−1.3671***	−0.7355***	−1.367***	−0.7328***	−0.8635***	−0.8724***	−1.3667***	−0.8932***	0.0529
	(0.2022)	(0.1176)	(0.2023)	(0.1127)	(0.1572)	(0.1573)	(0.2023)	(0.1462)	(0.0361)
Landlocked importer dummy	−0.4715**	−0.7287***	−0.4716**	−0.7412***	−0.6964***	−0.7035***	−0.4716**	−0.7964***	−0.0661*
	(0.1838)	(0.1082)	(0.1839)	(0.1044)	(0.1408)	(0.1409)	(0.1839)	(0.134)	(0.0362)
Exporter's remoteness	1.1878***	0.5461***	1.1876***	0.5089***	0.6598***	0.6472***	1.1875***	0.5673***	0.1249**
	(0.1821)	(0.1158)	(0.1822)	(0.114)	(0.1338)	(0.1352)	(0.1823)	(0.1359)	(0.0537)
Importer's remoteness	1.0097***	0.4337***	1.0094***	0.3961***	0.5615***	0.5493***	1.0093***	0.4541***	−0.0574
	(0.1541)	(0.1036)	(0.1542)	(0.1022)	(0.1185)	(0.1197)	(0.1542)	(0.1186)	(0.0531)
Trade agreement dummy	0.4425***	0.0304	0.4426***	0.0071	0.1811**	0.1794**	0.4426***	0.0374	1.2488***
	(0.109)	(0.0724)	(0.1091)	(0.073)	(0.0886)	(0.0903)	(0.1091)	(0.0771)	(0.142)
Openness dummy	0.928***	−0.3121***	0.927***	−0.3766***	−0.1068	−0.1394	0.9268***	−0.3161**	0.2917***
	(0.1912)	(0.1085)	(0.1915)	(0.1062)	(0.1312)	(0.1329)	(0.1915)	(0.1329)	(0.0276)
Inverse Mill Ratio (λ)							5915.8445	−1596.1116*	
							(5921.4483)	(857.0481)	

All reported standard errors (in parentheses) are corrected for heteroskedasticity. *p > 0.1, **p > 0.05, ***p > 0.01.

drop the extreme values (representing less than 3% of the NLS simulation results), we get very similar results (extremely large bias and standard deviations) to what is usually reported in the literature.

As for the set of experiments with normally distributed errors ($k = 0$), results are relatively different. For this version of the ET, although$k$ is set to zero, a certain degree of censoring is still implied given the possible negative simulated values that are converted to zeros. This explains why censored and non-censored versions of the same estimator yield slightly different results. Despite the low censoring rate, except for case 3, PPML and PPML_C estimates are biased, even in the case where PPML is supposed to be optimal (case 2). As for the other estimators, they perform reasonably well with significantly lower bias than PPML. In particular, GHeckman_2stp turns to be the best estimator in terms of bias and standard deviation.

4.2.2. The HM results

Table 3 summarizes the simulation results of the HM.[15] Although most of estimators yield biased results in one or more of the scenarios of the cases of heteroscedasticity and/or the distribution of the error terms, some estimators are more efficient than others. In fact, GHeckman_2stp seems to yield the best results in terms of bias and standard deviation. Although NLS_C yields relatively low bias for most of the cases, the corresponding standards deviations are very large. It is important to notice that for all estimators, the truncated versions yield significantly better results than the versions estimating the whole samples, especially for high degree of censoring. In the same line, we can clearly notice that when k increases from 0.5 to 1.25, the bias of PPML and NLS increases to substantially higher magnitudes, while for PPML_C and NLS_C, the bias increases at significantly lower rates. This could mean that the mis-specification bias is more severe than the selection bias.

Moreover, HM results show that both PPML and PPML_C produce significantly biased estimates for all simulations scenarios. A closer look at those results can reveal that PPML tends to overestimate the true coefficients (positive mis-specification bias), while PPML_C tends to underestimate them (negative selection bias). This confirms that PPML technique can yield inconsistent results when the true data generator produces structural zeros.

As for the performance of the nonlinear Heckman estimators, the results confirm that Heckman_2stp and GHeckman_2stp were very successful in correcting for the sample selection bias, especially when the variance of simulated data is constant. Indeed, for heteroskedasticity pattern 1, the bias of both Heckman_2stp and GHeckman_2stp is very close to zero with very low standard deviations. Consistently with the ET findings, GHeckman_2stp was very efficient in reducing the standard deviations of the nonlinear Heckman estimates for severe heteroscedasticity cases (3, 4 and 5). However, GNLS and GNLS_C performed really badly, and had higher bias and standards deviations than those of NLS and NLS_C, respectively. It turns out that when the frequency of structural zeros is relatively high, GNLS, like PPML, becomes inconsistent and inefficient.

Finally, it is interesting to notice that the estimation bias seems to be quite affected by the data distribution function (normal vs lognormal), which means that the performance of the estimators is sensitive to the

[15] Since that the selection equation of the HM ($y_{i,2} = \exp(x_i\beta) - k + \mu_i$) is not consistent with the Heckman first stage equation ($y^*_{2i} = x_i\alpha + u_{2i}$), we decide not to report the Probit estimates. However, we should note that most of the Probit estimates ($\widehat{\alpha}$) are consistent with their counterparts of the gravity equation ($\widehat{\beta}$)in terms of sign and significance. As for the coefficient on the IMR, unreported simulations results show that for the first heteroskedasticity scenario, $\widehat{\theta}$is significantly positive with an average magnitude close to the correlation coefficient ρ (0.5). However, as the heteroskedasticity becomes more severe, the average estimate of $\widehat{\theta}$ significantly decreases (especially for case 3).

error distribution. This might indicate that a more accurate estimation of the gravity equation would probably require a maximum likelihood technique.

By way of review, the results of the ET and MH indicate that nonlinear Heckman estimators are very successful in correcting for the sample selection bias. Moreover, GHeckman_2stp, data even with high frequency of structural zeros. Although, none of the estimators yields perfectly unbiased efficient results when the data are censored, GHeckman_2stp seems to perform reasonably well and outperform all other estimators. PPML yields very good results only when there is no censoring and the error terms are log-normally distributed. Even in this case, the new GNLS seems to yield more efficient results than PPML for several variance scenarios.

5. Empirical results

In this section, we report the results of the nonlinear estimators that we propose and compare them with the main empirical findings in the literature, namely those of SST (2006). We start with a brief description of the data, and then we discuss the results.

5.1. Data

As our paper proposes refinements of the existing gravity equation estimators, mainly PPML, using the database of SST (2006) would be essential for comparison purposes. The data consist of a cross-section of bilateral export flows for 136 countries in 1990. A detailed description of the data, including the list of countries, variables, sources, etc., can be found in SST (2006). Briefly, the dependent variable is 'bilateral exports', and the explanatory variables are 'exporter's GDP′, 'importer's GDP′, 'real GDP per capita', 'distance', and 'remoteness'. Moreover, the explanatory variables include a set of dummy variables for contiguity, common-language, colonial-tie, landlocked, 'preferential trade agreements', and 'openness.'[16]

5.2. Estimates of the coefficients

Table 4 presents the results of the traditional nonlinear estimators widely used in the literature, namely NLS and PPML, and the newly introduced nonlinear estimators, GNLS, Heckman_2stp and GHeckman_2sp. We also present the Probit estimates for the first stage of the Heckman models (Heckman_2stp & GHeckman_2stp). NLS, GNLS, Heckman_2stp, GHeckman_2stp and PPML are estimated with the whole sample (including country pairs with zero bilateral trade flows), whereas NLS_C, GNLS_C and PPML_C represent the censored versions of the corresponding models.[17]

As expected, our estimates of NLS and PPML match those of SST (2006). We can also notice that GNLS and GNLS_C yield estimates that are relatively close to those of PPML and PPML_C, respectively. It is interesting to note that the Heckman estimate of the coefficient of the IMR, θ,[18] is not statistically significant, which explains why Heckman_2stp yields almost the same estimated coefficients as those of NLS_C, especially for the key variables. In fact, if the sample selection is random (or not significant), θ should be asymptotically equal to 0. In this case, the nonlinear Heckman two-step would simply consist in estimating the same nonlinear equation that NLS_C estimates. Hence, the estimates of NLS_C and Heckman_2stp should be asymptotically equivalent. Having said that, we should expect some numerical discrepancy with finite samples estimations since IMR coefficient is not going to be exactly equal to 0. Moreover, results show that there is no significant difference between the estimates of the full sample PPML and NLS, and their corresponding truncated subsample versions. Hence, we can conclude that for the actual bilateral trade database, including or excluding zero trade values does not seem to dramatically affect the estimation results, at least when heteroskedasticity is not estimated.[19]

Another interesting point to notice is that NLS, NLS_C and Heckman_2stp's estimated coefficients of log of importer's GDPs ($\approx$0.86) are significantly higher than those reported by PPML and the generalized estimators (between 0.71 and 0.74). This suggests that estimating the gravity equation with estimators that assume constant-trade variance (i.e. NLS, NLS_C and Heckman_2stp) can potentially lead to an overestimation of the contributions of importer's GDP in explaining the bilateral trade flows, at least for the data used in this paper.[20] Remarkably, GHeckman_2stp yields the lowest estimated coefficients of the log of importer's and exporter's GDPs. This indicates that, even when sample selection seems to be random, estimating the gravity equation with nonlinear estimators that explicitly model zero trade values and, most importantly, allow for different conditional trade variance processes (i.e. GHeckman_2stp), may result in relatively low coefficient estimates of the main key variables of the traditional gravity equation.

Coefficients of other variables such as the distance, exporter's and importer's per capita GDP, common language, landlocked exporter and importer, openness, and remoteness continue to be significant, especially for the generalized versions, but with some differences in the magnitudes, as compared to the PPML. For instance, the GNLS and GHeckma_2stp estimates of the exporters' and importers' remoteness coefficients are lower than their counterparts in the PPML with difference clustered around 0.1 and 0.11, respectively. The same holds true for the estimates of the openness but with larger differences in magnitude of around 0.2. In the same vein, PPML estimates for distance, exporter's and importer's per capita GDP, landlocked exporter and importer coefficients are higher than our generalized estimates but the divergence is small. This suggests that the PPML, which does not fully address the issue of heteroscedasticity, might tend to overestimate most of the regressors' coefficients of the gravity equation.

Finally, we report the first-stage Probit estimation, shared by both Heckman estimators. As expected, most of key variables seem to affect the extensive margin of trade (probability of doing a trade) in a similar way they affect the intensive margin of trade (volume of trade). Interestingly, colonial ties and preferential trade agreements, which do not seem to affect the trade volume according to our main estimators,[21] have significantly positive coefficients in the Probit regression. In other words, while sharing colonial past and/or being a part of preferential trade agreement do not affect the level of bilateral trade between two countries, they do play an important role in determining whether these countries would trade with each other.[22]

[16] See SST (2006) for a detailed description of the data sources and the variables' summary statistics.

[17] Results of OLS and log-linearized Heckman Two-Step Least Square estimations (not reported here, but available up on request) are consistent with previous empirical studies.

[18] For GHeckman_2stp, the estimate of θ is only significant at 0.1 level. The negative sign on the IMR would mean that some unobservable factors that make two countries more likely to trade with each other seem also to negatively affect their expected bilateral trade. However, it is important to note that the IMR coefficient was not significant for the standard Heckman and was only weakly significant for the GHeckman_2sls. Moreover, the estimates of the GHeckman_2sls were very close to those of the GNLS_C, which technically means that the selection bias is not a significant issue for the actual dataset.

[19] Possibly explained by non-random selection and very small potential trade for zero trade observations. However, zero trade values issues can potentially arise with different bilateral trade databases that might not necessarily share the same features as those of SST (2006)'s database.

[20] Same applies for trade friction variables (trade agreement and openness).

[21] Including the PPML for the case of colonial ties.

[22] Unlike the PPML that predicts a significant positive impact of trade agreements on the intensive margin of trade.

6. Conclusion

The prevalence of zero trade values and the presence of heteroscedasticity have raised many questions about the estimation methods of the gravity model. In this paper, we revisit the theoretical principles and estimation practices of many specifications of the gravity equation and show that improving the consistency and efficiency of the estimates is possible. Moreover, we propose new nonlinear estimators that deal with heteroskedasticity and correct for potential selection bias. In particular, we introduce a new nonlinear Heckman two-step (Heckman_2stp) estimator that corrects for the selection bias and avoid Jensen inequality. We also introduce generalized versions of the NLS and the Heckman two-step (GNLS and GHeckman_2stp) estimators that allow for different trade variance processes. We use Monte Carlo simulations to assess the performance of our proposed estimators and to compare them with the other popular nonlinear estimators, namely the NLS and PPML. Simulation results indicate that GHeckman_2stp was very successful in correcting for the sample selection and in reducing the standard errors of the estimates, even with high frequency of structural zeros. Overall, the generalized nonlinear Heckman two-step estimator seems to perform reasonably well and outperform all other estimators. According to the simulation results, PPML yields very good results only when there is no censoring and when error terms are log-normally distributed. Moreover, the new generalized NLS outperforms PPML for several variance scenarios.

Our empirical results suggest that estimating the gravity equation with estimators that assume constant-trade variance can lead to an overestimation of the contributions of importer's GDP in explaining the bilateral trade flows, at least for the database used in this paper. Moreover, when heteroskedasticity is not estimated, including or excluding zero trade values, does not seem to dramatically affect the estimation results. Finally, the GHeckman_2stp, our best estimator candidate, yields the lowest coefficient estimates for almost all gravity equation variables, especially for GDP elasticities of both importer and exporter.

In future work, we plan to introduce a new maximum likelihood nonlinear Heckman that endogenously models the occurrence of zero observations and estimates the heteroscedasticity.

Declaration of competing interest

None.

References

Amemiya, T., 1973. Generalized least squares with an estimated autocovariance matrix. Econometrica 41, 723–732.
Anderson, J., 1979. A theoretical foundation for the gravity equation. Am. Econ. Rev. 69, 106–116.
Anderson, J., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. Am. Econ. Rev. 93, 170–192.
Arvis, J.-F., Shepherd, B., 2013. The Poisson quasi-maximum likelihood estimator: a solution to the "adding up" problem in gravity models. Appl. Econ. Lett. 20 (6), 515–519.
Assane, D., Chiang, E.P., 2014. Trade, structural reform, and institutions in Sub-Saharan Africa. Contemp. Econ. Pol. 32 (1), 20–29.
Bosquet, C., Boulhol, H., 2015. What is really puzzling about the "distance puzzle. Rev. World Econ. 151 (1), 1–21.
Burger, M., van Oort, F., Linders, G.-J., 2009. On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. Spatial Econ. Anal. 4 (2), 167–190.
Dai, M., Yotov, Y.V., Zylkin, T., 2014. On the trade-diversion effects of free trade agreements. Econ. Lett. 122 (2), 321–325.
Davidson, R., MacKinnon, J.G., 2003. Econometric Theory and Methods. Oxford University Press, New York.
Delgado, M.A., 1992. Semiparametric generalized least squares in the multivariate nonlinear regression model. Econom. Theor. 8 (2), 203–222.
Delgado, M.A., Kniesner, T.J., 1997. Count data models with variance of unknown form: an application to a hedonic model of worker absenteeism. Rev. Econ. Stat. 79 (1), 41–49.
de Sousa, J., 2012. The currency union effect on trade is decreasing over time. Econ. Lett. 117, 917–920.
Eaton, J., A. Tamura, A., 1994. Bilateralism and regionalism in Japanese and US trade and direct foreign investment patterns. Journal of the Japanese and International Economics 8, 478–510.
Egger, P., Larch, M., 2011. An assessment of the Europe agreements' effects on bilateral trade, GDP, and welfare. Eur. Econ. Rev. 55, 263–279.
Egger, P.H., Tarlea, F., 2015. Multi-way clustering estimation of standard errors in gravity models. Econ. Lett. 134, 144–147.
Fitzgerald, D., 2008. Can trade costs explain why exchange rate volatility does not feed into consumer prices? J. Monetary Econ. 55 (3), 606–628.
Gómez-Herrera, E., 2013. Comparing alternative methods to estimate gravity models of bilateral trade. Empir. Econ. 44 (3), 1087–1111.
Gómez-Herrera, E., Baleix, J.M., 2010. Are Estimation Techniques Neutral to Estimate Gravity Equations? an Application to the Impact of EMU on Third Countries' Exports. Department of Theory and Economic History, Faculty of Economics and Business, University of Granada. http://www.etsg.org/ETSG2009/papers/gomez.pdf. (Accessed 11 October 2019).
Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: theory. Econometrica 52 (3), 681–700.
Head, K., Mayer, T., 2014. Gravity equations: workhorse, toolkit and cookbook. Handb. Int. Econ. 4, 131–201.
Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. J. Int. Econ. 81 (1), 1–14.
Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.
Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. Q. J. Econ. 123 (2), 441–487.
Hurd, M., 1979. Estimation in truncated samples when there is heteroscedasticity. J. Econom. 11 (2–3), 247–258.
Krisztin, T., Fischer, M.M., 2015. The gravity model for international trade: specification and estimation issues. Spatial Econ. Anal. 10 (4), 451–470.
Lin, F., 2013. Are distance effects really a puzzle? Econ. Modell. 31, 684–689.
Mallick, S., marques, H., 2017. Export prices, selection into exporting and market size: evidence from China and India. Int. Bus. Rev. 26 (6), 1034–1050.
Martin, W., Pham, C.S., 2019. Estimating the gravity model when zero trade flows are frequent and economically determined. Appl. Econ. https://doi.org/10.1080/00036846.2019.1687838.
Martínez-Zarzoso, I., 2013. The log of gravity revisited. Appl. Econ. 45 (3), 311–327.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall, London.
Nguyen, T.N.A., Haug, A.A., Owen, P.D., Gene, M., 2020. What drives bilateral foreign direct investment among Asian economies? Econ. Modell. 93, 125–141.
Puhani, P.A., 2000. The heckman correction for sample selection and its critique. J. Econ. Surv. 14 (1), 53–68.
Santana-Gallego, M., Ledesma-Rodríguez, F.J., Pérez-Rodríguez, J.V., 2016. International trade and tourism flows: an extension of the gravity model. Econ. Modell. 52, 1026–1033.
Santos Silva, J.M.C., Tenreyro, S., 2015. Trading partners and trading volumes: implementing the helpman–melitz–rubinstein model empirically. Oxf. Bull. Econ. Stat. 77 (1), 93–105.
Santos Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statitsics 88 (4), 641–658.
Shepherd, B., 2010. Geographic diversification of developing country exports. World Dev. 38 (9), 1217–1228.
Sukanuntathum, A., 2012. Robust estimation of gravity models under heteroskedasticity and data censoring. Procedia – Social and Behavioral Sciences 40, 731–735.
Tenreyro, S., 2007. On the trade impact of nominal exchange rate volatility. J. Dev. Econ. 82, 485–508.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26, 24–36.
Xiong, B., Chen, S., 2014. Estimating gravity equation models in the presence of sample selection and heteroscedasticity. Appl. Econ. 46 (24), 2993–3003.
Yotov, Y.V., 2012. A simple solution to the distance puzzle in international trade. Econ. Lett. 117, 794–798.