

Estimating the Gravity Model When Zero Trade Flows Are Frequent and Economically Determined

Will Martin
Cong S. Pham



WORLD BANK GROUP

Development Research Group

Agriculture and Rural Development Team

June 2015

Abstract

This paper evaluates the performance of alternative estimators of the gravity equation when zero trade flows result from economically-based data-generating processes with heteroscedastic residuals and potentially-omitted variables. In a standard Monte Carlo analysis, the paper finds that this combination can create seriously biased estimates in gravity models with frequencies of zero frequently observed in real-world data, and that Poisson Pseudo-Maximum-Likelihood models can be important in solving this problem. Standard threshold-Tobit estimators perform well in a Tobit-based data-generating process only if the analysis deals with the heteroscedasticity problem. When the data are generated by a Heckman sample selection model, the

Zero-Inflated Poisson model appears to have the lowest bias. When the data are generated by a Helpman, Melitz, and Rubinstein-type model with heterogeneous firms, a Zero-Inflated Poisson estimator including firm numbers appears to provide the best results. Testing on real-world data for total trade throws up additional puzzles with truncated Poisson Pseudo-Maximum-Likelihood and Poisson Pseudo-Maximum-Likelihood estimators being very similar, and Zero-Inflated Poisson and truncated Poisson Pseudo-Maximum-Likelihood identical. Repeating the Monte Carlo analysis taking into account the high frequency of very small predicted trade flows in real-world data reconciles these findings and leads to specific recommendations for estimators.

This paper is a product of the Agriculture and Rural Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at w.martin@cgiar.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Estimating the Gravity Model When Zero Trade Flows Are Frequent and Economically Determined

Will Martin and Cong S. Pham *

JEL: C21, F10, F11, F12, F15

* At the time of writing Will Martin was Research Manager, Development Research Group, the World Bank, and Cong S. Pham was Lecturer, School of Accounting, Economics and Finance, Deakin University, respectively. Martin is now at the International Food Policy Research Institute (IPRI). Their email addresses are w.martin@cgiar.org and cpham@deakin.edu.au. The authors would like to thank João Santos Silva and Silvana Tenreyro for their helpful comments and for access to their data. They also would like to thank Jonathan Eaton, Caroline Freund, Russell Hillberry, Hiau Looi Kee, Daniel Lederman, Douglas Nelson, Guido Porto, Shang-Jin Wei and Tiemen Woutersen for valuable comments on earlier versions of this paper. The usual disclaimer applies.

I. Introduction

The gravity model is now enormously popular for analysis of a wide range of trade questions. This popularity is due to its apparently good performance in representing trade flows and to the strong theoretical foundations provided in papers such as Anderson (1979) and Anderson and van Wincoop (2003). Until recently, the gravity equation was almost invariably estimated using data sets converted into logarithms and truncated to contain only positive trade flows. The problem of zero trade flows was rarely emphasized, partly because trade theories were silent on their causes, partly because of a lack of recognition of their frequency, and partly because of the convenience of the log-linear estimator.¹

Recently, there has been growing recognition among trade economists that zero trade flows do not occur randomly and that using truncated samples in logarithms may yield misleading results. Melitz (2003) and myriad subsequent studies have added firm-level foundations to earlier models of zero trade flows, such as those based on the Tobin (1958) and Heckman (1979) models. Santos Silva and Tenreyro (SST) (2006) raised an important additional concern by pointing to the possibility of serious bias from combinations of heteroscedasticity and nonlinearity in the gravity model. Helpman, Melitz and Rubinstein (HMR) (2008) pointed to another potential source of bias in coefficient estimates—failure to account for the increase in demand for exports as the number of exporting firms increases.

Like SST (2006), we believe that at least part of the process of identifying the right estimator must involve analysis of generated data whose statistical properties are known. In this paper we evaluate different estimators of the gravity equation in the presence of pervasive zero trade flows *and* under different patterns of heteroscedasticity. Because we are unsure of the process generating real-world data, we test alternative estimators on data sets generated using three different economic models in which the probability of a nonzero trade flow and the volume of trade when trade is nonzero are jointly determined: the threshold-Tobit model of Eaton and Tamura (1994); the Heckman model; and an extended Heckman model of the type used by HMR (2008).

¹ As shown in Annex A, zero trade flows account for more than 40% of the possible bilateral trade flows in country-level data and more than 60% in U.S. 10-digit product-level export data, and the probability of zero trade flows is strongly linked to the explanatory variables in gravity models. The frequency of zero trade flows is much higher at the more disaggregated levels of product and firm trade that are increasingly being used for analysis.

Each of the three data-generating processes used in our Monte Carlo simulations provides a plausible economic representation of observed trade flows, and can be parameterized to represent the frequency of zero trade flows observed in real trade data. The resulting data sets differ from those used by SST (2006) in containing true zero observations, and in being based on economic models that jointly determine whether to trade and, if so, the volume of trade. They are *not* constant-elasticity models because constant-elasticity models cannot generate finite probability mass at zero trade. Instead of moving from a constant-elasticity model to a model containing zeros by transforming the latent variables from constant-elasticity models with a negative-binomial distribution as in SST (2011), we use limited-dependent variable models whose properties have been thoroughly explored in the literature.

The paper is organized as follows. In the next section we provide a brief review of the relevant literature. Section III discusses the econometric problems associated with zero trade flows. Potential approaches to estimating the gravity equation in the presence of frequent zero trade flows are considered in Section IV. Section V covers the DGPs of our Monte Carlo simulations and the simulation results. Section VI provides an empirical implementation of the gravity estimation using country-level trade data and section VII discusses reconciling Monte Carlo and empirical findings. Finally, the paper concludes in Section VIII.

II. Review of Relevant Literature on Estimating the Gravity Model

Eaton and Tamura (1994) appear to have been the first to provide a systematic approach to estimation of the gravity equation as a limited-dependent model. Their model builds on the Tobin (1958) model, where non-limit observations arise when the latent dependent variable (including the error term) crosses a threshold value.² Another popular approach to estimation uses the Heckman (1979) model to deal with the presence of zero trade flows (see, for example, Francois and Manchin (2007) and Lederman and Ozden (2007)). The third approach that we consider builds on the Heckman model to take into account the extensive margin of firm involvement in trade (HMR 2008).

² Note that in the Tobit model the effect of the coefficient estimate on the explanatory variable x cannot be interpreted as the effect of x on the dependent variable y . Rather, as McDonald and Moffitt (1980) correctly point out, it is the effect of x on two components of y : “(1) the change in y for those observations above the limit, weighted by the probability of being above the limit; and (2) the change in the probability of being above the limit, weighted by the expected value of y if above” (pg. 318-319).

In an important and influential article, SST (2006) emphasized the potential econometric problems resulting from a combination of heteroscedastic residuals and zero trade flows. Using Monte Carlo simulations, they showed that traditional estimators may yield severely-biased parameter estimates and identified the Poisson Pseudo Maximum Likelihood (PPML) estimator as the preferred estimator for dealing with these problems. The data sets used for the main experiments in their 2006 paper included no zeros (SST 2011, p220), suggesting that their important findings about the bias and variance of standard estimators arise from the combination of nonlinearity and heteroscedasticity highlighted in the paper, rather than from success in dealing with the presence of zero observations. While their paper included some experiments with zeros generated by rounding errors, a process quite different from economic processes generating zeros such as those assumed when applying limited-dependent variable estimators.

The tractable and apparently robust PPML estimator proposed by SST is now widely used in estimation of gravity models (see, for example, Shepherd 2010; Anderson and Yotov 2012). SST show why the normal equations used to solve the PPML estimator should make it more robust than OLS-based estimators when the function is nonlinear and the errors are heteroscedastic, and show using Monte Carlo simulations that its coefficients are much less subject to bias in this situation. Based on their evidence, the case for the PPML estimator seems extremely strong for analysis of nonlinear relationships in models where zero values are infrequent.³ However, because the tests in their 2006 paper were based on samples containing no zero values, the results obtained cannot be expected to provide a reliable guide to the robustness of the PPML estimator in the presence of economically-determined zero values.

Perhaps a more fundamental problem for reliance on the PPML estimator applied to censored samples (including the zeros) is that the move from the traditional, discrete version of the Poisson distribution to the continuous needed for the gravity equation removes the ability to have positive probability mass at the limit observation (Iliencko 2013, p2). While a discrete version of the Poisson distribution could adequately capture the presence of large shares of limit observations of the type identified by Tobin (1958), the

³ Heteroscedasticity in nonlinear models with relatively few zero observations arises in many important applications such as consumption, investment and money demand systems; and representative models for consumer demand, firm costs and profits. The results of SST (2006) appear to deserve much more attention for estimation in this context.

continuous version of this distribution is unable to adequately represent these cases. This problem appears worse than the common concern that empirical distributions may be over-dispersed relative to the Poisson, in which the mean and the variance are equal. Especially in cases where large fractions of a sample are zero trade flows that are retained in the analysis, the continuous Poisson specification appears likely to be seriously mis-specified. The Zero-inflated Poisson estimator developed by Lambert (1992) and applied to the gravity model by Burger, van Oordt and Linders (2009) might be one way to retain the advantages of the Poisson estimator while allowing for substantial numbers of zeros.

On the theoretical front, there have also been major recent developments. HMR (2008) provide a heterogeneous-firm trade model that both allows for frequent and asymmetric patterns of zeros in bilateral trade data and introduces a number-of-firms variable, correlated with the probability of non-zero trade that affects trade volume through a love-of-variety effect. They extend the Heckman model so that the standard gravity explanatory variables determine not only the probability of positive bilateral trade (i.e. the probability that the most productive firm finds it profitable to export to a foreign destination) but also the fraction of firms that are productive enough to enter foreign markets when positive trade occurs. This overcomes an omitted-variable problem different from that created by the limited-dependent feature of the data.

III. Intuition on the Econometric Problems Associated with Zero Flows

Since Tobin's famous (1958) paper, it has been known that zero values of the dependent variable can create potentially large limited-dependent variable biases in parameter estimates, even in linear models, if the estimator used does not allow for this feature of the data generating process. Heckman (1979) generalized Tobin's approach to estimation in the presence of this problem, casting it in the context of estimation using samples with non-random sample selection. A nonlinear version of Heckman's formulation of the problem is presented in a two equation context:

$$(1) \quad y^*_{1i} = f(x_i, \beta) + u_{1i}$$

$$(2) \quad y^*_{2i} = f(z_i, \alpha) + u_{2i}$$

where x_i and z_i are vectors of exogenous regressors while β and α are vectors of parameters, and

$$E(u_{ji}) = 0,$$

$$\begin{aligned} E(u_{ji} u_{j'i'}) &= \sigma_{jj'} \text{ if } i = i' \\ &= 0 \text{ if } i \neq i' \end{aligned}$$

In Heckman's formulation, equation (1) is a behavioral equation and equation (2) is a sample selection rule that determines whether observations in equation (1) are observed or not. In our context, this model results in:

$$(3) \quad y_{li} = y_{li}^* \quad \text{if } y_{2i}^* > 0 \text{ or}$$

$$(4) \quad y_{li} = 0 \quad \text{if } y_{2i}^* \leq 0$$

A key problem for estimation of equation (1) in the presence of sample selection is that its residuals no longer have the properties assumed in standard regression theory, particularly a zero mean. In this situation, standard regression procedures result in biased estimates of the coefficient vector β because they omit relevant explanatory variables.

The Tobit model is a special case of the Heckman model in which the right hand sides of (1) and (2) are identical (Heckman 1979). In this case, a simple diagram (Figure 1) provides important insights into the key problems arising in applying standard approaches to estimation. Our version of the figure includes both the nonlinearity and heteroscedasticity of the gravity equation emphasized by SST (2006) and the limited-dependent variable problem.

Figure 1. Limited-dependent variable bias in the gravity model

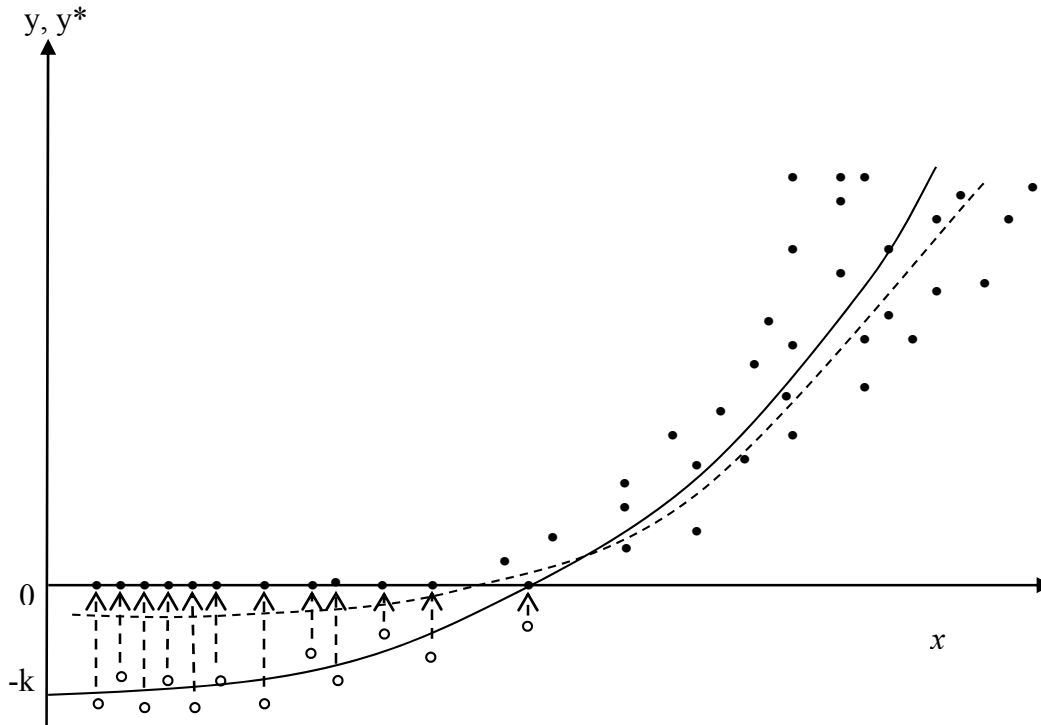


Figure 1 shows the relationship between a latent variable, y^* , and an explanatory variable, x , with the non-stochastic relationship represented by the solid curve, and individual data points by the bullets. The hollow bullets correspond to any value of y^* less than or equal to zero. In the censored regression case, the residuals associated with these low values of the latent variable will be replaced by positive residuals that lead to a zero value of the dependent variable, a change represented by the dashed arrows in the diagram. With this model, using a sample containing the zero values—for which $E(u_i) > 0$ —biases the coefficient on the slope and results in an estimate something like the dashed curve (Greene 1981).

Another insight from careful examination of Figure 1 is the difference between the case of censoring shown in the diagram and the case of truncation where all of the observations at the limit (zero in this case) are discarded from the sample. In the censoring case, the error terms on all limit observations but those where y^* was precisely zero have been increased. In the case of truncation, observations with negative errors, like the rightmost hollow bullet in the diagram, are more likely to be excluded from the truncated sample and this creates a situation where $E(u_i) > 0$. Intuitively, the transformation of the

error terms associated with the censored sample seems likely to be greater than that associated with the truncated sample. This suggests that moving from a truncated estimation procedure to one which retains the zero values without changing the estimation procedure may result in greater bias.

The diagram shows another feature of the residuals from application of a regression estimator designed for non-censored problems. Such estimators are likely to find residuals that are both large and serially dependent at both ends of the regression line—note the consistently positive apparent residuals relative to the dashed line in Figure 1 for observations near zero and near the largest observations in the sample. The estimated regression line is likely to be strongly influenced by such extreme observations, particularly when using ordinary least squares (Beggs 1988). However, the implications of this are likely to vary considerably between estimators, given the different weighting implied by the normal equations for the different estimators (SST 2006).

With a little more imagination, Figure 1 can also help visualize a different problem associated with heteroscedasticity even in the absence of the nonlinearity problem identified by SST—a problem associated with mis-measurement of the probability of observations at the limit. Following the notation of Terza (1985), the log-likelihood function for a sample generated by a Tobit-type process is given by

$$L = \sum_i d_i \log(f(y_i)) + (1-d_i) \log(F(0))$$

where d_i is an indicator value taking the value 1 for a non-limit observation; $f(y_i)$ is the density function for non-limit observation i , and $F(0)$ is the distribution function showing the probabilities of zero observations. If the variance relevant to assessing the value of the distribution function is incorrectly specified—as it would be with a model assuming homoscedasticity—this is clearly likely to change the realized values of the distribution function for limit observations, causing potentially-serious bias in parameter estimates, as pointed out by SST (2014) in their assessment of the HMR (2008) estimator.

IV. Potential Approaches to Estimating Gravity Models with Zeros

In estimating the gravity model when there are many zero observations, some key questions must be confronted:

- i. Which functional form to use for the explanatory variables?
- ii. Whether to truncate or censor the zero observations? and

iii. What estimator to use?

There seems to be widespread agreement that the gravity model involves relationships between variables that are nonlinear in levels, with a constant-elasticity functional form almost always used. This implies a relationship:

$$(5) \quad y_i = \exp(x_i\beta) + \varepsilon_i$$

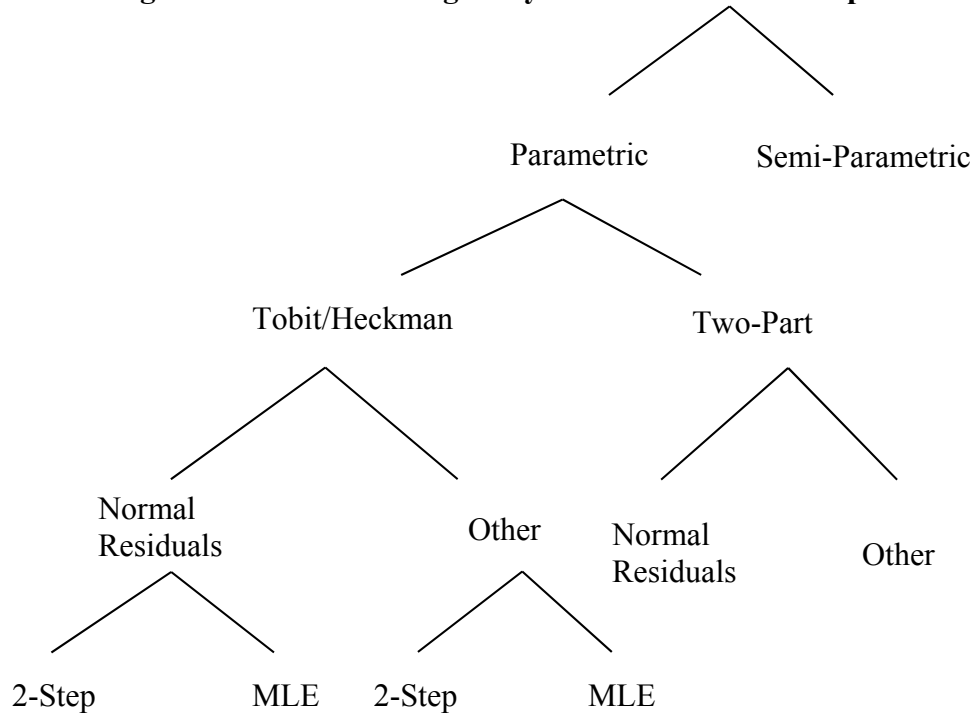
where y_i represents bilateral trade; x_i is a vector of explanatory variables (some of which may be linear, some in logarithms and some dummy variables) for observation i ; β is a vector of coefficients; and ε_i is an error term whose variance, unlike those of equations (1) and (2), need not be constant across observations. As noted by SST, the gravity model has traditionally been estimated after taking logarithms, which allows estimation using linear regression techniques. However, this specification will only involve homoscedastic residuals if $\varepsilon_i = \exp(x_i\beta) \cdot v_i$ where v_i is distributed independently of x_i with a zero mean and constant variance. Of course, if this is the true model, the constant-elasticity model estimated in levels will have heteroscedastic errors.

The use of the logarithmic transformation for the dependent variable creates another immediate difficulty when trade is zero, since the log of zero is undefined. The most common response to this problem has been to truncate the sample by deleting the observations with zero trade. This is inefficient, since it ignores the information in the limit observations, and may lead to bias as noted in the previous section. Many studies have replaced the value of imports by the value of imports plus one, allowing the log of the zero values to take a zero value. Others have estimated in levels using nonlinear estimation techniques which allow the zero values to be retained. However, as we have seen from Figure 1, retaining the zero observations without using an estimator that accounts for the limited-dependent nature of the model may lead to more serious bias than simply truncating the sample. What seems to be needed is an approach to estimation that systematically takes into account the information in the limit observations. Once the decision to include the limit observations is taken, a number of other decisions must be confronted. We find it useful to lay out the choices with the decision tree in Figure 2.

At the first stage of the decision tree, analysts must decide between parametric models and semi-parametric models. Semi-parametric models (Chay and Powell 2001) avoid specifying a distribution for the residuals, sometimes at the expense of computational efficiency, and estimate parameters using methods such as Powell's (1984) Censored Least

Absolute Deviations (CLAD) model. While such models have been extended to deal with nonlinear models (Berg 1998), it appears that such applications have been infrequent to date. Certainly, most of the focus in estimation of gravity models has been on the parametric branch of the decision tree.

Figure 2. Choosing an estimator for the gravity model with limited-dependent trade



If a parametric approach is taken, the first decision required is whether to adopt a Tobit/Heckman-type model (Amemiya 1984), or a Two-Part model (Jones 2000). The Two-Part model has the desirable feature of allowing the sample selection and the behavioral equations to be estimated independently (Duan et al. 1983). However, this simplification comes at the expense of assuming that these decisions are taken independently, something that seems implausible in a world where decisions on whether to export and on how much to export and how much to ship should they choose to export are taken by individual firms based on the profitability of exports of their particular products to particular markets. In most cases, it seems to us that the variable of interest is the latent variable for the desired level of trade, y^* , for which the Tobit/Heckman approach seems the most suited.

A key argument for the Two-Part model has been a belief that the performance of the sample-selection models is irretrievably compromised by statistical problems, and particularly multicollinearity. Leung and Yu (1996) show that these problems may be less of a problem for practical implementation than was thought based on earlier studies such as Manning, Duan and Rogers (1987), which included insufficient variation in the exogenous variables to mitigate the multicollinearity between these variables and the inverse Mills ratio added to the behavioral equation when estimating using the two-step estimator of the Heckman (1979) sample selection model.⁴ Based on a detailed review of the literature on the Heckman correction for sample selection bias, Puhani (2000) found that the full-information maximum likelihood estimator of Heckman's model generally gives better results than either the two-step Heckman estimator or the Two-Part model, although the Two-Part model is more robust to multicollinearity problems than the other standard estimators. Clearly, these results suggest that the consistency of the data with the assumptions of the Tobin/Heckman models should be examined carefully.

Whichever parametric estimator is chosen, an assumption about the distribution of the residuals must be made. The most common approach is to assume that the residuals are distributed normally. However, alternative assumptions are sometimes used, including the Poisson distribution highlighted by SST (2006) or the Gamma distribution that they also examined. The decision about which distribution to use need not be based solely on judgments about the actual distribution of the residuals. The essence of SST's recommendation of the PPML estimator was that it is more robust to heteroscedasticity than one based on the normal distribution when the residuals are actually normally distributed.

The first part of the Two-Part approach is the use of a qualitative-dependent model such as Probit to determine whether a particular bilateral trade flow will be zero or positive. The second part is to estimate the relationship between trade values and explanatory variables using only the truncated sample of observations with positive trade (Leung and Yu 1996, p198). Potential estimators for this stage include the standard approach of OLS in logarithms; the nonlinear least squares (NLS) model used by Frankel and Wei (1993);

⁴ Leung and Yu (1996, p201) show that problems with Heckman's two-step estimator are more likely when there are few exclusion restrictions; a high degree of censoring; limited variability among the exogenous regressors; or a large error variance in the choice equation.

and the PPML and Gamma Pseudo-Maximum Likelihood (GPML) estimators discussed by SST. Estimation of the bias of truncated single-equation estimators will therefore help assess the suitability of Two-Part models.

Under the Tobit/Heckman limited-dependent branches of the decision tree, a decision must be made about whether to estimate using two-step estimators of the type proposed by Heckman (1979) or a maximum likelihood approach (see Tobin (1958), Puhani (2000) and Jones (2000)). One concern with the Heckman two-step estimator is that the variable that adjusts for the probability of sample selection⁵ may be close to a linear function of the other explanatory variables, resulting in multicollinearity problems (Puhani 2000). A second concern is that this approach introduces heteroscedasticity into the residuals. An alternative, nonlinear, approach to estimating Heckman's second-step equation is provided by Wales and Woodland (1980, p461). We do not consider this estimator because it performed poorly in their simulations.

The performance of the Heckman-based models appears to depend heavily on whether both equations (1) and (2) are active, and whether there are at least some variables included in equation (1) but excluded from equation (2). Leung and Yu (1996) concluded that this estimator with some excluded variables in the behavioral equation outperformed other estimators for limited-dependent variables.

V. Monte Carlo Simulations

In this section of the paper, we provide a description of the approach we used to create zero observations with frequencies similar to those observed in real-world data and to test the ability of different estimators to retrieve the parameters of these data generating processes. We also followed the approach of SST and of Westerlund and Wilhelmsson (2011) to creating heteroscedastic errors—that is, we posited a range of different types of heteroscedasticity, and tested their implications for the performance of different estimators. Because we know the true parameters, we focus on the bias and imprecision of the coefficient estimates, rather than the goodness of fit measures emphasized by Gomez-Herrera (2013).

⁵ Which is the inverse Mills ratio for the particular observation (Heckman 1979).

We adopted the following specification of the gravity equation from SST (2006).⁶

$$(6) \quad y_i = \exp(x_i\beta) + \varepsilon_i = \exp(x_i\beta) \cdot \eta_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \cdot \eta_i$$

where $\eta_i \equiv 1 + [\varepsilon_i / \exp(x_i\beta)]$; x_1 is a binary dummy, designed to mimic variables such as border dummies, that equals 1 with probability 0.4; x_2 is a standard normal variable to mimic the behavior of continuous explanatory variables such as the log of distance or income; and the data are randomly generated using $\beta_0=0$, $\beta_1 = \beta_2 = 1$. Following SST, we assumed that η_i is log-normally distributed with mean 1 and variance σ_i^2 .

To assess the sensitivity of the different estimators to different patterns of heteroskedasticity, we used Santos Silva and Tenreiro's four cases:

$$\text{Case 1: } \sigma_i^2 = (\exp(x_i\beta))^{-2}; \quad V(y_i|x) = 1$$

$$\text{Case 2: } \sigma_i^2 = (\exp(x_i\beta))^{-1}; \quad V(y_i|x) = \exp(x_i\beta)$$

$$\text{Case 3: } \sigma_i^2 = 1; \quad V(y_i|x) = (\exp(x_i\beta))^2$$

$$\text{Case 4: } \sigma_i^2 = (\exp(x_i\beta))^{-1} + \exp(x_{2i}); \quad V(y_i|x) = \exp(x_i\beta) + \exp(x_{2i}) \cdot (\exp(x_i\beta))^2$$

where Case 1 involves an error term that is homoscedastic when estimated in levels; Case 3 is homoscedastic when estimated in logarithms; Case 2 is an intermediate case between 1 and 3; and Case 4 represents a situation in which the variance of the residual is related to the level of a subset of the explanatory variables, as well as to the expected value of the dependent variable. To help interpret such a large number of results, we also present, in Appendix Table 4, measures of the average absolute bias for each coefficient across the four cases and the average absolute bias across both coefficients, together with averages of their standard errors.

We consider two broad types of data-generating-process (DGP): one based on the Eaton-Tamura model and another based on the Heckman model. We also examine a special case of the Heckman DGP, including an additional variable for the number of products traded, proposed by HMR (2008), whose robustness to heteroscedasticity bias does not appear to have been examined on data sets with known parameters, although SST (2014) raise concerns about potentially serious bias. These DGPs have solid economic foundations and can generate data sets with the required fractions of zero trade values. We considered using the SST (2011) DGP but its economic underpinnings and statistical properties were less obvious to us than was the case for the three more conventional limited-dependent

⁶ These equations correspond exactly to the SST (2006) specification for equations (14) and (15).

variable DGPs that we used and, in any event, its properties have already been examined using Monte Carlo techniques. We investigated the Two-Step Method of Moments estimator of Xiong and Chen (forthcoming), but did not find encouraging results with our sample.

V.1. ET-Tobit data generating process (DGP)

For the threshold-Tobit model, we ensured that a significant number of observations would have zero values by adding a negative intercept term, $-k$, in the levels version of the data-generating equation, and then transforming all realizations of the latent variable with a value below zero into zero values. We refer to this DGP as the ET-Tobit DGP because it is the model underlying the Eaton and Tamura (1994) estimator. It has the interpretation of introducing a threshold that must be exceeded before trade actually occurs. It differs fundamentally from the rounding approach used by SST (2006) and the approach of setting observations to zero randomly or for particular groups used by Martinez-Zarzoso (2013).

The data generating process for these initial simulations was equation (6) augmented with the negative intercept term, $-k$, that defines the ET-Tobit model

$$(7) \quad y_i^* = \exp(x_i\beta) + \varepsilon_i - k \equiv \exp(x_i\beta) \cdot \eta_i - k = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \cdot \eta_i - k$$

where $y_i = y_i^*$ if $y_i^* \geq 0$; $y_i = 0$ if $y_i^* < 0$

Within these samples, we found that a value for k of 1 provides numbers of zero trade values consistent with the 40-50 percent of zero values frequently observed in analyses of total bilateral trade. As shown in Appendix Table 1, a k value of 1.5 generated higher shares of zeros. The analysis was performed in Stata 9.2, using double precision to minimize numerical errors, with samples of 1000 observations, replicated 10,000 times as in SST (2006).⁷

Semi-Parametric Estimators

Given the pervasive uncertainty about the distribution of the errors, and the relatively poor results obtained using standard limited-dependent variable estimators, it seemed important to investigate the performance of the semi-parametric, Least-Absolute-Deviations (LAD) approach proposed by Powell (1984). Although this model has not, to our knowledge, been

⁷ Our first estimation task was to replicate the simulations of SST (2006). While our results are not exactly the same as theirs because of the stochastic nature of the analysis, they are completely consistent. The replicated simulation results are available upon request.

applied to gravity models, Paarsch (1984) found the censored version of this model to give satisfactory results in large samples with censored data. Because the estimator for this model in Stata requires linearity, we examined the model in logarithms to make an initial assessment of its suitability. We considered first the truncated LAD estimator based only on the positive observations, and then turned to Powell's (1984) censored LAD estimator. The results are presented in Table 1.

From the results in Table 1, it appears that the Truncated LAD estimator has quite small bias in Case (3), when the heteroscedasticity is consistent with the functional form adopted for estimation. However, it appears to be very vulnerable to heteroscedasticity. In cases (1), (2) and (4), the bias of the Truncated LAD was typically 20 to 30 percent. The Censored LAD estimator performed worse than the truncated estimator in all cases using $k=1$, with bias of 0.5 or greater. With $k=1.5$, the bias averaged 70 percent in cases (1) to (3). Given its overall poor performance we did not pursue the LAD estimator further.

Standard Single-Equation Estimators

In this section, we first considered the performance of eight parametric estimators that do not account for the limited-dependent nature of the DGP. Despite their failure to account for the limited-dependent nature of the data-generating-process, they might plausibly be the best estimators of the key underlying parameters, just as truncated OLS was widely believed to be the best estimator for limited-dependent variable models for the reasons outlined in Manning, Duan and Rogers (1987). The specific estimators considered were: (i) the traditional truncated OLS-in-logs regression, (ii) its censored counterpart with 0.1 added to the log of output to allow taking logs of the dependent variable, (iii) Truncated nonlinear least squares (NLS) in levels, (iv) Censored NLS in levels, (v) a Poisson Pseudo-Maximum Likelihood (PPML) estimator, (vi) a truncated Pseudo-Maximum Likelihood estimator, (vii) a Gaussian Pseudo-Maximum Likelihood (GPML) estimator, and (viii) a Negative Binomial Maximum Likelihood estimator (NBML). We report the results for $k=1$ in Table 2. Summary results reporting the average absolute bias and standard errors across the four cases, and across the two coefficients, are included in Appendix Table 4.

An important feature of the results for the truncated OLS-in-logarithms model is its apparently strong sensitivity to heteroscedasticity. In Case 3, where the error distribution is homoscedastic in the log-linear model, the bias in the estimates is around 6 percent for both coefficients. However, when we move to cases involving errors that are

heteroscedastic in the log-linear equation, the bias of the estimated coefficient on the normally-distributed variable (β_2) rises to around 20 percent in cases 1, 2 and 4. For this estimator, unlike many others, the biases in the coefficients are generally similar for the dummy variable, x_1 and the normally distributed explanatory variable, x_2 . The censored OLS model in logarithms (with 0.1 added to allow inclusion of the zeros) produces results that are generally inferior to those obtained from the truncated OLS estimator. Except in Case 4, the biases were smaller for the coefficient estimate on x_2 than on x_1 .

Truncated NLS is the levels counterpart to the traditional OLS-in-logs estimator and the second stage of a two-part estimator. It has lower bias than its logarithmic counterpart in cases 1 and 2, but is inferior in cases 3 and 4. In Case 3, the bias of this estimator is 40 percent for β_2 , around seven times the bias of truncated OLS in logs. Perhaps the best thing that can be said for the truncated NLS estimator is that it is consistently less biased than the censored NLS regression model. The superiority of the truncated OLS and NLS models over their censored regression counterparts reinforces our intuition that just solving the “zero problem” and adding the zero-valued observations to the sample is likely quite an unhelpful strategy when the data are generated by a process that determines the levels of observed trade and the probability of zero trade together.

The PPML estimator in levels yielded estimates that were strongly biased in all cases. Because this equation was estimated with the dependent variable in levels, the underlying error structure is consistent with the estimator in case 1, but, In this case, the bias in the estimate of β_1 was 0.28 and of β_2 was 0.25. The bias was very similar for each of the other cases. This confirms SST’s hypothesis that the PPML estimator is robust to heteroscedasticity in the residuals while demonstrating its potential vulnerability to limited-dependent variable bias.

The truncated PPML estimator has much lower bias than the PPML estimator in all cases, again reinforcing our belief that truncated estimators are to be preferred over censored estimators. The truncated PPML, in fact, outperforms truncated OLS in logs in all cases considered, including Case 3 where heteroscedasticity is not a problem for the OLS estimator.

The GPML estimator has very large bias in almost all cases, reinforcing the SST conclusion that this estimator has little role to play in estimation of the gravity model. The NBML estimator is inferior to the two preferred estimators—truncated OLS in logs and

truncated PPML—in all except Case 1 with $k=1$. This good performance seems likely to be an outlier, since the NBML is strongly biased in Case 1 when $k=1.5$ (see Appendix Table 2).

When the $k=1.5$ sample is considered (see Appendix Table 2), the superior performance of truncated PPML relative to full-sample, censored PPML is again evident. It has by far the lowest mean absolute bias across the four cases considered and is superior to its nearest competitors (the two OLS in logs models) even in case 3, where the pattern of heteroscedasticity is consistent with the model in logarithms. This result suggests a case for use of the truncated PPML, at least in exploratory analyses. It retains the robustness to heteroscedasticity emphasized by SST (2006), but reduces the vulnerability to the limited-dependent variable bias emphasized in the earlier literature originated by Tobin (1958).

The summary results in Appendix Table 4 give the average absolute bias and standard errors across the four cases and again highlight the superiority of truncated PPML in terms of bias. For both coefficients, it has by far the lowest average absolute bias—less than half that of other estimators in almost every case, and on average across the two coefficients. The standard errors associated with this estimator are almost always among the lowest—although OLS in logs and the NBML have lower standard errors around their highly biased estimates.

Limited-Dependent Variable Estimators

In this section, we considered eight estimators designed specifically for estimation with limited-dependent variables. The estimators considered are: (i) the Eaton-Tamura model with the dependent variable in levels; (ii) the Eaton-Tamura model in logs; (iii) a Tobit-Poisson model⁸ like that of Terza (1985); (iv) the maximum likelihood version of the Heckman model estimated in logs (see Amemiya (1984)); (v) the Heckman (1979) 2SLS estimator in logs; (vi) the ML version of the Heckman estimator in levels; (vii) the 2SLS version of the Heckman model estimated in levels by nonlinear least squares; and (viii) the Zero-Inflated Poisson Maximum likelihood estimator (ZIP) (Lambert 1992).

Key results for cases with $k = 1$ are presented in Table 3, while results for $k=1.5$ are presented in Appendix Table 3. The Eaton-Tamura Tobit estimator in levels has quite

⁸ To program the likelihood function in Stata, we needed to replace the factorial function with $\exp(\ln(\text{gamma}(y+1)))$ to allow evaluation with non-integer values of the dependent variable.

low bias for both coefficients—around three percent—in Case 1. However, it exhibited much larger bias in all other cases. The same model in logarithms suffers from quite large bias (25 percent) in cases 1 and 2 but around 6 percent in Case 3, where the pattern of heteroscedasticity is consistent with the estimator. The average bias across cases was much lower for the log version than for the version in levels. The poor result for the ET estimator in levels is somewhat surprising given that the ET-Tobit model is the data-generating process, and that the ET-Tobit model in levels might be expected to avoid the bias associated with the combination of heteroscedasticity and log-linearization.

The performance of the Poisson-Tobit estimator is much worse than the ET-Tobit in all cases. It yields large bias—20 percent or higher in all but one case—and large standard errors in all cases. Because this model lacks the additive intercept that is central to the ET-Tobit data generating process, it suffers from similar problems to the censored PPML estimator.

The Heckman ML estimator in logs performed poorly in virtually every case. The bias of this estimator was smallest in case (3), where the form of heteroscedasticity is consistent with the estimator, but was still 14 percent for each coefficient. The Heckman 2SLS estimator in logs was worse in all cases, while the Heckman ML in levels was sharply inferior in cases 3 and 4. The Heckman 2SLS estimator in levels had the lowest average absolute bias of any of the limited-dependent variable estimators, although this bias was over 16 percent for β_1 in all four cases. The ZIP estimator was uniformly inferior to the Heckman 2SLS estimator.

We had expected the ET-Tobit models to outperform alternative models such as Heckman given that the DGP in this case is ET-Tobit. One potential approach to improving the performance of the ET Tobit model might be to adjust for heteroscedasticity. In Table 4, we report results including the adjustments proposed by Maddala (1985), in which the error variance is specified by the process $\sigma_i^2 = (\gamma + \delta(x_i\beta))^2$ for the log-linear model, with γ and δ are estimated using nonlinear least squares. For the linear-in-levels model, the specified error process is $\sigma_i^2 = (\gamma + \delta \exp(x_i\beta))^2$. These models should capture heteroscedasticity of the types considered in Cases 1 to 3. The resulting estimates were reasonably satisfactory for Cases 1 and 3 using estimators in logs and for Case 1 using estimators in levels. When the true restriction that $\beta_0=0$ was imposed, the bias in β_1 and β_2 fell to very low levels. This suggests that it may be important in estimation of the

heteroscedasticity-corrected model to use restrictions on the parameter values when prior information or prior testing suggests that these are valid.

A feature of the results for this DGP is that the limited-dependent variable models were, in almost all cases, outperformed in terms of both bias and standard error by the truncated PPML estimator. This suggests that a two-part approach to estimation might be best if the underlying DGP appears to be of the ET-Tobit type, with a first stage used to assess whether trade occurs or not, and the PPML estimator applied in the second stage. This is clearly an imperfect strategy given that the truncated PPML estimator has an average bias across the two coefficients of 7 percent, but the average bias of even the best limited-dependent estimator—the Heckman 2SLS in levels—is around 50 percent higher. The superiority of the truncated PPML estimator over all other estimators appears to be robust to the frequency of zeros in the sample. It also arises in the sample with $k=-1.5$. It appears worthwhile to pursue limited-dependent-variable estimation only if an appropriate correction for heteroscedasticity can be made—perhaps by following a general to specific approach to estimation in order to introduce as many justified restrictions as is possible.

V.2. Heckman sample selection DGPs

In this section, we examine the performance of the estimators when the data are generated using Heckman Sample Selection models. For this purpose, we first generate the data with frequent zeros based on a standard Heckman sample selection framework. We then construct HMR-type data sets with two sources of omitted variable bias— sample selection and omission of the number of firms engaged in exporting.

The Standard Heckman DGP

For the standard Heckman DGPs, we assume that the selection equation contains at least one variable that is excluded from the behavioral equation. To generate the data for this test, we use equation (6) for the behavioral equation, but add the following sample selection equation:

$$(8) \quad y_{2i} = \exp(z_i\alpha) \cdot \eta_i^* \equiv \exp(z_i\alpha) + u_{2i}^*$$

where $z_i\alpha = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_4 x_{4i}$; $\eta^*_i \equiv 1 + [u^*_{2i}/\exp(z_i\alpha)]$; x_4 is a variable included in the selection equation but excluded from the behavioral equation; and the correlation between $\ln(\eta^*_i)$ and $\ln(\eta_i)$ is set at 0.5.⁹ The dependent variable in (6) is observed when $y_{2i} > 1$.

The results for standard estimators with this DGP are presented in Table 5 while those based on limited-dependent estimators are presented in Table 6. Results for average bias and standard errors across the four cases are summarized in Appendix Table 4. Zero trade flows account approximately for 30% of the observations in this sample.

In Table 5, many of the results are similar to those obtained using Eaton-Tamura data sets. For example, comparison of the truncated and censored versions of each estimator reveals consistently better performance of the truncated versions. Also similarly, the NLS estimators generally have smaller biases and standard errors in cases 1 and 2 while the than in cases 3 and 4. With this DGP, the censored OLS in logs estimator has quite serious bias in case 3, even though the form of the dependent variable is consistent with homoscedasticity.

The Poisson estimator outperforms the standard log-linear estimators in most cases. Specifically, in cases 1 2 and 3 the bias on the estimate for β_1 is about 14%. The bias is smaller in case 4 but the standard error is substantially larger. The bias on the—frequently more important—estimate of β_2 is substantially lower than for β_1 . The truncated PPML performs better than the censored PPML in almost all cases, reinforcing our finding using the ET-Tobit DGP. As in the case of the Tobit DGP we find that as the number of zeros increases so does the bias and standard error of the parameter estimates generated by the PPML estimator.¹⁰

The GPML and NBML estimators suffer from very large bias in almost all cases. The standard errors associated with these estimators are also large. This finding reinforces our earlier findings—and those of SST—that these estimators do not appear to perform well for estimation of the gravity model.

When we turn to limited-dependent variable estimators, we find that the ET-Tobit estimators are frequently seriously biased. In some cases, such as the ET-Tobit estimator

⁹ Note that $\ln(\eta^*_i)$ and $\ln(\eta_i)$ correspond exactly to the two error terms U_{1i} and U_{2i} in Heckman (1979), respectively. The correlation between $\ln(\eta^*_i)$ and $\ln(\eta_i)$ is assumed to be 0.5 in order to rule out the case of independence between the two error terms in which case, the standard least squares estimator may be used on the selected subsample without introducing bias.

¹⁰ The simulation results with higher percentages of zero observations are available upon request by the authors.

in levels in Case 1, this seems related to the mismatch between the data generating process and the more restricted estimator. In Case 3 with this estimator in logs, the bias is again large with the error process homoscedastic. The Heckman estimators generally have lower bias and standard errors than the ET-Tobit estimators, which is not surprising given that the data were generated using a Heckman process. Perhaps somewhat surprisingly, however, the Heckman estimators in logs performed better than those in levels, despite the problems associated with the log transformation. The Poisson-Tobit estimator in levels performed reasonably well, although the Heckman-Maximum Likelihood estimator had lower average bias and smaller standard errors.

As shown in Appendix Table 4, the Zero-inflated Poisson (ZIP) estimator has, on average, substantially lower bias than any of the other estimators considered. The standard errors of this estimator are slightly larger than for the Heckman ML estimator in logs, the estimator with the second-lowest average absolute bias, suggesting that it is important to have a large sample when considering the use of the ZIP estimator¹¹. The Heckman ML estimator has lower bias than the ZIP for categorical explanatory variables, but greater bias for the normally-distributed variable. The Heckman 2SLS estimator in logs may also be worth considering if convergence problems with the Heckman ML prove too difficult although its average bias across the four cases is more than twice as high as that for the ZIP estimator. By contrast with the ET-Tobit DGP, the ZIP estimator has clearly lower bias than the truncated PPML model both on average and in all but one case.

The HMR Data Generating Process

To capture the essence of the DGP à la HMR we need a firm number variable that is non-negative and positively correlated with the probability of nonzero trade. It should also be correlated with the economic mass variable, and take the value zero when there is no trade. We generate synthetic data sets with our four different types of heteroscedasticity for Monte-Carlo analysis using the following two equations and the procedures laid out in more detail in Annex B:

$$(10.1) \quad y_{1i} = \exp(x_i\beta) \cdot \eta_i \equiv \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) + \varepsilon_i$$

$$(10.2) \quad y_{2i} = \exp(z_i\alpha) \cdot \eta^*_{i} \equiv \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_4 x_{4i}) + u^*_{2i}$$

¹¹ Another practical problem with ML estimators with the dependent variable in levels is frequent difficulty with convergence in Stata. In many cases we had to vary initial values to overcome this problem. In other cases, even varying initial values does not lead the ML estimators to convergence, especially when the dependent variable is in levels.

where x_3 is the firm number variable, constructed to be correlated with x_2 ; and all other variables are as defined in equations (6) and (9). Equation (10.2) is the selection equation which determines whether there is nonzero trade. Specifically, if $y_{2i} > 1$ then y_{1i} is positive. Again, as in the standard Heckman sample selection model, we set the correlation between $\ln(\eta_i)$ and $\ln(\eta^*_i)$ equal to 0.5. Variable x_3 in the behavioral equation mimics the behavior of the firm heterogeneity variable in the HMR theoretical framework. It is—for want of better information—assumed to be uniformly distributed¹². As in the standard Heckman case, x_4 is excluded from the behavioral equation.

Finally, we look at some simulation results when the data are generated à la HMR. The first source of potential omitted-variable bias comes from the standard sample selection problem: similar factors determine the volume of trade (y_{1i}) and the probability ($\text{Prob}(y_{2i} > 1)$) that trade occurs between two countries. The second source of omitted variable bias results from exclusion of a variable for the number of exporting firms (x_3). Like the firm number proxy used by HMR, x_3 is uncorrelated with the error term in the trade equation.

From Table 7 and Appendix Table 4, it is clear that the truncated PPML estimator has the lowest average absolute bias of the standard estimators. The truncated OLS estimator in logs has the next lowest average absolute bias, and its standard errors are lower than for the truncated PPML, but the average absolute bias is four times as large as for truncated PPML. The NLS estimators perform extremely poorly, as do the GPML and NBML standard estimators.

The coefficient on x_2 is of particular interest when the DGP is based on a Heckman sample selection with omitted variable bias. Since, by construction $\text{Cov}(x_2, x_3) = 0.5 > 0$, we expect upward bias in the estimated coefficient on x_2 when omitted-variable bias is not controlled for.

From Table 8 and Appendix Table 4, it is clear that the ZIP2 estimator with the firm number variable included has by far the lowest average absolute bias. The ZIP estimator without this variable is the next best-performing in terms of bias, with the truncated PPML coming next. The HMR estimator follows in terms of bias, although it has much higher bias on the important β_2 coefficient. ZIP2 type estimators assume that

¹² We also considered the case with a normally distributed variable and reached the same conclusions.

researchers have available estimates of firm numbers, and have the ability to deal with the potential endogeneity of such a variable. If the available variable for firm numbers suffers from endogeneity, the approaches suggested by Wooldridge (2010, section 18.5) are likely to be useful. One concern with all of the Poisson-based estimators is that they have somewhat higher standard errors than the HMR estimator.

VI. Empirical Estimation

While we are very conscious of the importance of the multilateral resistance terms emphasized by Head and Mayer (2013), and hence concerned about the empirical validity of traditional gravity models, we consider the implications of different estimators for parameter estimates from both traditional (including GDP as explanatory variables) and Anderson-van Wincoop (using country fixed effects) specifications. The results for these analyses are presented in Tables 9 and 10.¹⁶ Since the specification with country fixed effects controls for the multilateral resistance terms the results in Table 10 are our preferred ones.

The results in Table 9 are very informative about the coefficients on GDP and the log of distance. The truncated PPML has very similar but slightly smaller coefficients than the censored PPML. But a more striking feature of the table is that all of the Poisson-based estimators have quite similar coefficients on GDP and distance that are well below (in absolute value) those of traditional estimators such as Truncated OLS and the Heckman estimators. A striking feature of these results is that the ZIP coefficient estimates are identical to those from the truncated PPML estimator. Two important points are noteworthy with respect to this result. First, the fact that the coefficient estimates of the two estimators are the same hides the fact that the marginal effects from each model are different. Second, while the coefficient estimates of the truncated Poisson and the ZIP estimators are the same and equal to the elasticities (i.e. $\partial \ln(y_i)/\partial \ln(x_i)$) they are evaluated at different predicted values of the dependent variable.¹⁷ The ZIP2 estimator, which performed the best in our

¹⁶ Note that our truncated sample contains 8857 observations and our full sample 17028. Both are smaller than SST's truncated sample (9613) and full sample (18360) because we used common religion as our excluded restriction variable. Data for this variable are not available for all the samples used by SST (2006). For comparison purpose the sample sizes for our truncated OLS and Poisson estimator were kept the same as for our Heckman estimator.

¹⁷ Specifically, the elasticities are evaluated at $\hat{y}_i = 18644.519$ for the truncated Poisson and $\hat{y}_i = 1704.81$ for the ZIP.

Monte-Carlo tests with an HMR-based DGP, yields coefficient estimates above those of the other Poisson estimators, but still well below those for traditional estimators. The HMR estimator is an outlier in the other direction, with much lower coefficients on GDP, and somewhat lower coefficients on distance, than the PPML-based estimators.

The results for the Anderson-van Wincoop formulation are presented in Table 10 with model estimates when we control for firm heterogeneity à la Helpman, Melitz and Rubinstein. The results on the log of distance variable are very similar with this model to those for the traditional gravity model. The PPML model results for this coefficient are clustered around -0.8, while those for the traditional model are around -1.4. The HMR Model is again an outlier with a coefficient of -1.14.

VII. Reconciling Monte Carlo and Empirical Findings

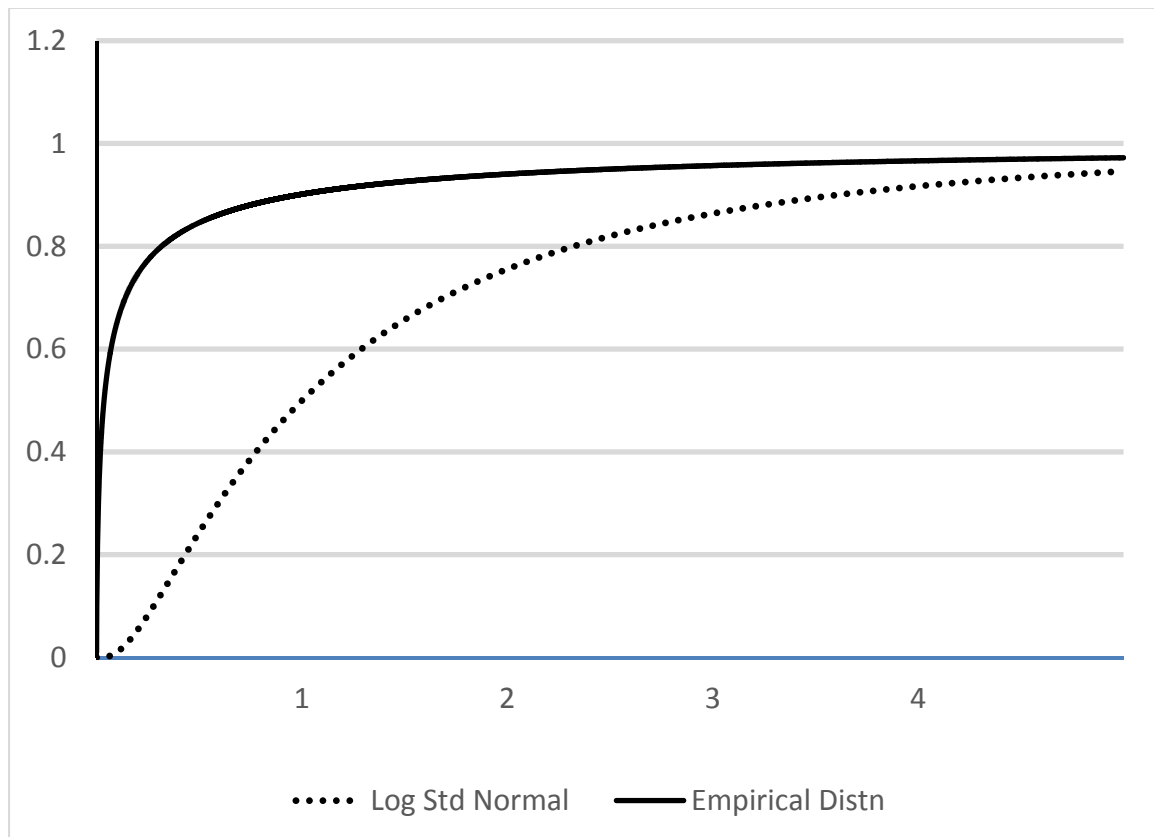
The contrast between the empirical results and those from the Monte Carlo analysis raises important questions. Why, in particular, is there so little difference between the results from the PPML model and the truncated PPML, and why are the truncated PPML and the ZIP model results identical? One possible explanation lies with the distribution of the continuous explanatory variable used in the Monte Carlo analysis. By using a standard normal variable, we have ensured that, with the unit coefficient used in the Monte Carlo analysis (and assuming a zero value for the 0-1 variable, a relatively small fraction of the $X_i\beta$ values have values near zero. For any given distribution of the error term this reduces the frequency of zero observations and requires larger adjustment terms such as the $-k$ in the ET-Tobit model to obtain any given frequency of these observations.

One simple approach—consistent with the Leung-Yu approach of ensuring that the distribution of the explanatory variables in a Monte Carlo analysis is consistent with the parameters of the empirical distribution being studied—is to examine the distribution of a composite, continuous explanatory variable consistent with the empirical data set. To do this for the SST empirical dataset of, we combined the three signature explanatory variables of the SST empirical data set into a composite variable¹⁸ $\text{Log}[(\text{GDP_Exporter} * \text{GDP_Importer}) / (\text{Bilateral distance})^{1.3}]$, where the coefficients on GDP and on distance are based on averages of the estimates in SST (2006, p650). To allow

¹⁸ We performed the same experiment with the parameters of the PPML model and found qualitatively similar results.

comparability of the density and distribution functions of this variable and the log standard normal, we rescaled this composite driving variable to have the same mean as the log standard normal. We found that the distribution of this variable was well-represented by a log-normal distribution with mean -5.728896 and standard deviation 3.523514, rather than 0 and 1 as assumed with the log standard normal. To provide a visual comparison of the two distributions, we compare their cumulative distribution functions¹⁹ (CDFs) in Figure 3. The figure reveals that the empirically-based distribution has a much higher fraction of the explanatory variable near zero than the standard normal variable.

Figure 3. CDFs for the Log Standard Normal and the Empirical Distribution



Note: Generated using MS EXCEL with 50000 data points per distribution at a spacing of 0.0001.

A focus on a couple of data points in Figure 3 highlights the stark differences between the two distributions. With the empirical distribution, 80 (90) percent of the observations are below 0.06 (0.27). With the log standard normal distribution, only 0.3 (11) percent of the

¹⁹ We had intended to compare their density functions but it proved to be impossible to plot these on the same scale.

observations lie within this range. The very large fraction of empirical observations near zero has important implications for the limited-dependent data-generating process of interest to us. For any given distribution of errors, the presence of a large share of the observations of the driving variable near zero makes it much more likely that the outcome will be transformed into a zero observation.

We next use a continuous explanatory variable with this distribution and the simplest model previously examined—the ET Tobit model—and repeat our Monte Carlo analysis. When we do this, we find that the absolute value of the k parameter needed to obtain any given share of zeros falls precipitously. Specifically, with a tiny value for k of 0.01, the percentages of zero observations in heteroscedasticity cases 1, 2, 3 and 4 are 71 percent, 64 percent, 53 percent and 64 percent, respectively. The 53 percent share with heteroscedasticity of type 3 is almost exactly the share of zeros observed in the SST data for total trade between 136 countries. It is much higher than the fraction of zeros reported in Table 1 using k values of 1.0 and 1.5 with the log standard normal distribution. Moving to slightly larger values of k would allow us to replicate the shares of zeros observed in studies using disaggregated trade data.

More importantly, we obtain results, shown in Table 11, that have features that are consistent with our empirical findings. The estimated values on the continuous variable are very similar for both the PPML and the truncated PPML models. And the ZIP and the truncated PPML models are identical. Another interesting feature of these results is that the Negative Binomial model—which performed so poorly in our earlier analysis appears to perform very well, although it remains subject to the concerns expressed by SST (2006) about its parameter estimates depending upon the units of measurement.

VIII. Conclusions

The purpose of this paper was to evaluate the performance of different estimators of the gravity equation in the presence of zero trade flows, which are a common feature of trade data both at the aggregate and disaggregate level. We adopt two different data-generating processes in our Monte Carlo simulations—the Tobit and the Heckman sample-selection models. For the Tobit DGP we used Monte Carlo simulations based on the design of Santos Silva and Tenreyro (SST 2006), modified to include a threshold level of trade that must be surmounted before positive trade levels are observed. For the Heckman sample selection

DGP, an additional equation determines which observations of the outcome equation are included. We also analyze a data set generated for consistency with the Helpman, Melitz and Rubinstein (2008) model.

In our initial simulations, we follow SST in using Monte-Carlo simulations with a log standard normal distribution of the continuous driving variable. These tests confirm the importance of heteroscedasticity and the prevalence of zeros as sources of bias with many estimators. However, estimators such as the censored PPML that are known to perform well in the presence of heteroscedasticity are found to be potentially susceptible to limited-dependent variable bias when a substantial fraction of the observations are censored.

When the data are generated using an Eaton-Tamura Tobit model, the smallest biases were sometimes found with Eaton-Tamura Tobit estimators consistent with the data generating process, but only when the functional form was consistent with the form of heteroscedasticity, or an appropriate correction had been made to deal with heteroscedasticity. On average across the four cases considered, the truncated PPML had the lowest mean absolute bias, including relative to the explicitly limited-dependent estimators consistent with the data-generating process.

When the data are generated by a Heckman sample-selection DGP, the truncated PPML estimator outperformed all other models that do not explicitly account for the limited-dependent nature of the data set. The Zero-inflated Poisson (ZIP) model appeared to have the lowest bias of the models considered, and particularly for the continuously-distributed variables that are frequently of greatest interest in a gravity-model context.

Applying a range of estimators to the SST empirical dataset on international trade resulted in a number of surprising findings. In particular, we found very little difference between the PPML estimator and its truncated version. Further, the results from the truncated PPML estimator were identical with those from the ZIP estimator. To try to better understand the relationship between our Monte Carlo results and our empirical findings, we repeated the Monte Carlo analysis with a continuous explanatory variable distributed like a composite of the key explanatory variables in the gravity model, with a much higher fraction of the explanatory variables clustered near zero. This analysis helped explain two key findings in the empirical analysis—that the PPML and truncated PPML estimates were similar, and the truncated PPML and the ZIP were almost identical.

Our results, in an important sense, restore the SST findings about the importance of heteroscedasticity in the gravity model and in favor of the PPML model over traditional estimators for the gravity model of international trade. However, they also highlight the potential risks of limited-dependent variable bias. They also highlight the importance for Monte-Carlo testing of taking into account the characteristics of the distribution of the driving variables as well as of the residuals. Using a driving variable that is distributed as a log standard normal variable resulted in a relatively qualified recommendation for the PPML estimator as a least-bad option. When we moved to Monte-Carlo testing with a driving variable based on the empirical distribution of the key explanatory variables for trade, we were able to easily explain the large shares of zeros observed in real-world trade data and to validate the recommendation of the PPML as a good estimator for the gravity model, at least for models of aggregate trade flows.

References

- Amemiya, T. "Tobit models: a survey." *Journal of Econometrics* 24(1984): 3-61.
- Anderson, J. and E. van Wincoop. "Gravity with gravitas: a solution to the border puzzle." *American Economic Review* 93(2003): 170-92.
- Anderson, J. E. "A theoretical foundation for the gravity equation." *American Economic Review* 69(1979): 106-16.
- Anderson, J.E and Yotov, Y. "Gold standard gravity" Working Paper 17835, National Bureau of Economic Research, Cambridge, MA (2012).
- Beggs, J. "Diagnostic testing in applied econometrics." *Economic Record* 64(1988): 88-101.
- Berg, G. D. "Extending Powell's semiparametric censored estimator to include non-linear functional forms and extending Buchinsky's estimation technique." *University of Colorado Discussion Paper in Economics* 98-27(1998).
- Burger, M., van Oort, F. and Linders, G.-J. "On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation," *Spatial Economic Analysis*, 4 (2009): 167-190.
- Chay, K. Y. and J. L. Powell. "Semiparametric censored regression models." *Journal of Economic Perspectives* 15(2001): 29-42.
- Clark, A. K. "Testing nonnested models of international relations: reevaluating realism". *American Journal of Political Science*, 45 (2001): 724-744.
- Duan, N., W. G. Manning Jr., C. N. Morris and J. P. Newhouse. "A comparison of alternative models for the demand for medical care." *Journal of Business and Economic Statistics* 1(1983): 115-27.
- Eaton, J. and A. Tamura "Bilateralism and regionalism in Japanese and US trade and direct foreign investment." *Journal of the Japanese and International Economies* 8(1994): 478-510.
- Feenstra, R., Romalis, J. and Schott P. "U.S. Imports, Exports, and Tariff Data, 1989-2001". *NBER Working Paper No. 9387* (2002).
- Frankel, J., and Wei, S. "Trade blocs and currency blocs," NBER Working Paper No. 4335 (1993).
- Francois, J. and M. Manchin. "Institutions, infrastructure, and trade." Centro Studi Luca D'Agliano's Development Studies Working Paper Series 224 (2007).
- Gómez-Herrera, E. "Comparing alternative methods to estimate gravity models of bilateral trade" *Empirical Economics* (2013) 44:1087–1111
- Greene, W. "On the asymptotic bias of the Ordinary Least Squares Estimator of the Tobit model." *Econometrica* 49(1981): 505-13.
- Head, K. and T. Mayer. Gravity equations: workhorse, toolkit and cookbook, *Handbook of International Economics*, Volume 4 (2014): 131-201.
- Heckman, J. "Sample selection bias as a specification error." *Econometrica* 47(1979): 153-61.
- Helpman, E., M. Melitz, and Y. Rubinstein "Estimating trade flows: trading partners and trading volumes." *Quarterly Journal of Economics* 2 (2008): 441-487.

- Hillberry, R. and D. Hummels. "Trade response to geographic frictions: a decomposition using micro-data." *European Economic Review* 52 (2008): 527-550.
- Iliencko, A. "Continuous counterparts of Poisson and Binomial distributions and their properties" arXiv:1303.5990v1, 24 March 2013, Downloaded from <http://arxiv.org/pdf/1303.5990v1.pdf> 12 May, 2013.
- Jones, A. "Health econometrics." *Handbook of Health Economics, Volume 1*. A.Culyer and J.Newhouse, eds., 267-344. Amsterdam: Elsevier Science, 2000.
- Lambert, D. "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing." *Technometrics* 34(1)(1992):1-14.
- Lederman, D. and C. Ozden. "Geopolitical interests and preferential access to US markets." *Economics and Politics* 19(2007): 235-258.
- Leung, S. F. and S. Yu. "On the choice between sample selection and two-part models." *Journal of Econometrics* 72(1996): 197-229.
- Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK.: Cambridge, 1985.
- Manning, W., N. Duan, and W. Rogers. "Monte Carlo evidence on the choice between sample selection and Two-Part models." *Journal of Econometrics* 35(1987): 59-82.
- Martin, W. and Pham, C. "Estimating the gravity model when zero trade flows are frequent". Deakin University, School of Accounting, Economics and Finance Working Paper Series (2008).
- Martínez-Zarzoso, I. "The log of gravity revisited." *Applied Economics* 45(3)(2013): 311-27.
- McDonald, J. and Moffitt, R. "The Use of Tobit Analysis". *Review of Economics and Statistics*, 62 (1980): 318-321.
- Melitz, M. "The impact of trade on intra-industry reallocations and aggregate industry productivity." *Econometrica* 71(6)(2003):1695-725.
- Paarsch, H. J. "A Monte-Carlo comparison of estimators for censored regression models." *Journal of Econometrics* 24(1984): 197-213.
- Powell, J. L. "Least absolute deviations estimation for the censored regression model." *Journal of Econometrics* 25(1984): 303-25.
- Puhani, P. "The Heckman correction for sample selection and its critique." *Journal of Economic Surveys* 14(2000): 53-68.
- Santos Silva, J. and Tenreyro, S. "The log of gravity." *The Review of Economics and Statistics* 88(2006): 641-58.
- Santos Silva, J. and Tenreyro, S. "On the existence of the maximum likelihood estimates in Poisson regression" *Economics Letters* 107 (2010) 310-12.
- Santos Silva, J. and Tenreyro, S. "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator." *Economics Letters* 112 (2011) 220-2.
- Santos Silva, J. and Tenreyro, S. "Trading partners and trading volumes: implementing the Helpman-Melitz-Rubinstein model empirically." *Oxford Bulletin of Economics and Statistics* 0305-9049 doi: 10.1111/obes.12055, (2014) Accepted article.

- Santos Silva, J., Tenreyro, S. and Windmeijer, F. "Testing competing models for non-negative data with many zeros" DOI 10.1515/jem-2013-0005 *Journal of Econometric Methods* (2014) 1-18.
- Shepherd, B. "Geographical diversification of Developing Country exports." *World Development* 38(9)(2010):1217–28.
- Terza, J. "A tobit-type estimator for the censored Poisson regression model." *Economics Letters* 18(1985): 361-5.
- Tobin, J. "Estimation of relationships for limited dependent variables." *Econometrica* 26(1958): 24-36.
- Westerlund, J. and F. Wilhelmsson "Estimating the gravity model without gravity using panel data." *Applied Economics*, 43(6) (2011): 641-649.
- Vuong, Q. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis." *Econometrica* 57(1989): 307-333.
- Wales, T. J. and Woodland, A. D. "Sample Selectivity and the Estimation of Labor Supply Functions" *International Economic Review* 21 (2) (1980): 437-68.
- Wooldridge, J. *Econometric Analysis of Cross-Section and Panel Data*, 2nd Edn MIT Press, Cambridge, MA.(2010).
- Xiong, Bo., and Chen, Sixia, 'Estimating gravity equations in the presence of sample selection and heteroscedasticity' *Applied Economics*, forthcoming.

Table 1. Monte Carlo results from Least-Absolute-Deviations Estimators, ET-Tobit DGP with $k=1$

Estimator	Dependent Variable Form	β_1		β_2		β_1		β_2	
		Bias		Bias		Bias		Bias	
		$(k=1.0)$				$(k=1.5)$			
		Case 1: $V[y_i x]=1$				Case 1: $V[y_i x]=1$			
Truncated LAD	Log	0.2289	0.048	0.219	0.034	0.291	0.056	0.283	0.039
Censored LAD	Log	0.507	0.067	0.491	0.050	0.690	0.028	0.670	0.020
		Case 2: $V[y_i x]=\exp(x_i\beta)$				Case 2: $V[y_i x]=\exp(x_i\beta)$			
Truncated LAD	Log	0.195	0.079	0.190	0.045	0.211	0.092	0.211	0.053
Censored LAD	Log	0.561	0.035	0.531	0.024	0.734	0.043	0.711	0.027
		Case 3: $V[y_i x]=(\exp(x_i\beta))^2$				Case 3: $V[y_i x]=(\exp(x_i\beta))^2$			
Truncated LAD	Log	0.028	0.127	0.029	0.075	-0.013	0.151	-0.017	0.090
Censored LAD	Log	0.531	0.063	0.515	0.042	0.715	0.086	0.703	0.053
		Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i})$ $(\exp(x_i\beta))^2$				Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i})$ $(\exp(x_i\beta))^2$			
Truncated LAD	Log	-0.184	0.155	-0.192	0.092	-0.186	0.177	-0.250	0.107
Censored LAD	Log	0.125	0.245	0.575	0.186	0.014	0.139	0.018	0.178

Table 2. Monte Carlo results with standard estimators, ET-Tobit DGP with k=1

Estimator	Dependent	β_1		β_2		β_1		β_2	
	Variable Form	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.
Truncated OLS OLS (ln(y+0.1)) Truncated NLS Censored NLS PPML Truncated PPML GPML NBML	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$			
	Log	0.220	0.085	0.205	0.059	0.197	0.093	0.190	0.058
	Log	0.354	0.058	0.268	0.029	0.315	0.064	0.233	0.030
	Level	0.110	0.027	0.098	0.022	0.099	0.074	0.089	0.043
	Level	0.169	0.036	0.138	0.034	0.143	0.076	0.120	0.046
	Level	0.280	0.045	0.251	0.036	0.288	0.062	0.258	0.039
	Level	0.098	0.041	0.096	0.028	0.081	0.057	0.082	0.030
	Level	0.416	0.236	0.424	0.206	0.571	0.215	0.593	0.186
	Level	0.083	0.033	0.069	0.016	0.124	0.049	0.116	0.026
	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$			
Truncated OLS	Log	0.057	0.115	0.059	0.067	-0.158	0.137	-0.189	0.079
OLS (ln(y+0.1))	Log	0.162	0.077	0.096	0.036	0.152	0.0900	0.012	0.041
Truncated NLS	Level	0.091	2.814	0.396	20.469	0.236	7.112	0.668	25.202
Censored NLS	Level	0.190	2.160	0.503	24.282	0.324	6.364	0.819	24.46
PPML	Level	0.290	0.144	0.255	0.090	0.258	0.189	0.198	0.122
Truncated PPML	Level	0.029	0.146	0.028	0.028	0.032	0.196	-0.127	0.151
GPML	Level	0.864	0.206	0.922	0.156	0.558	0.247	0.499	0.179
NBML	Level	0.254	0.090	0.247	0.049	0.210	0.136	0.203	0.073

Table 3. Monte Carlo results with limited dependent variable estimators, ET-Tobit DGP with k=1

Estimator	Dependent	β_1		β_2		β_1		β_2		
	Variable Form	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	
ET-Tobit ET-Tobit Poisson-Tobit Heckman-ML Heckman-2SLS Heckman-ML Heckman-2SLS ZIP	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$				
	Level	0.034	0.03	0.034	0.025	0.159	0.085	0.148	0.076	
	Log	-0.252	0.045	-0.251	0.038	-0.16	0.055	-0.151	0.043	
	Level	0.28	0.044	0.251	0.034	0.288	0.062	0.258	0.038	
	Log	0.302	0.077	0.284	0.048	0.288	0.087	0.278	0.048	
	Log	0.624	0.098	0.598	0.073	0.596	0.116	0.578	0.085	
	Level	0.11	0.209	0.078	0.101	0.105	0.117	0.069	0.056	
	Level	-0.163	0.013	0.04	0.005	-0.169	0.039	0.04	0.023	
	Level	0.222	0.045	0.195	0.022	0.215	0.063	0.195	0.027	
	ET-Tobit ET-Tobit Poisson-Tobit Heckman-ML Heckman-2SLS Heckman-ML Heckman-2SLS ZIP	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$			
Level		0.668	0.171	0.657	0.182	0.727	0.302	0.685	0.084	
Log		0.062	0.089	0.067	0.062	0.008	0.114	0.064	0.075	
Level		0.281	0.138	0.242	0.079	0.203	0.161	0.163	0.092	
Log		0.14	0.119	0.14	0.074	-0.142	0.147	-0.029	0.108	
Log		0.303	0.175	0.299	0.145	0.239	0.231	0.358	0.262	
Level		0.209	22.808	0.263	1.646	0.26	6.736	0.43	4.001	
Level		-0.179	0.161	0.021	0.153	-0.194	0.227	0.001	0.196	
Level		0.168	0.171	0.142	0.119	0.195	0.232	-0.031	0.189	

Table 4. Monte Carlo Results: ET Tobit Adjusted for Heteroscedasticity: ET Tobit DGP with k=1

	β_0		β_1		β_2	
	Bias	Std Error	Bias	Std Error	Bias	Std Error
Logs						
Case 1	0.022	0.337	-0.027	0.163	-0.057	0.209
Case 2	-0.063	0.404	-0.041	0.238	-0.122	0.35
Case 3	-0.333	0.497	-0.008	0.264	-0.066	0.359
Case 4	-0.257	0.419	-0.182	0.39	-0.275	0.4
Levels						
Case 1	0.083	0.38	-0.019	0.12	-0.021	0.12
Case 2	0.513	0.251	-0.166	0.101	-0.177	0.12
Case 3	0.782	0.688	-0.263	0.265	-0.271	0.311
Case 4	1.358	0.961	-0.472	0.317	-0.366	0.487
Levels, $\beta_0=0$						
Case 1	-	-	-0.005	0.159	-0.038	0.615
Case 2	-	-	0.007	0.015	0.037	0.032
Case 3	-	-	0.037	0.047	0.058	0.078
Case 4	-	-	-0.024	0.065	0.347	0.084

Table 5. Monte Carlo results with standard estimators, Heckman DGP

Estimator	Dependent	β_1		β_2		β_1		β_2	
	Variable Form	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.
		Case 1: $V[y_i x]=1$				Case 2: $V[y_i x]=\exp(x_i\beta)$			
Truncated OLS	Log	-0.099	0.032	-0.029	0.040	-0.112	0.033	-0.017	0.047
OLS (ln(y+0.1))	Log	0.438	0.013	-0.031	0.057	0.420	0.015	-0.067	0.059
Truncated NLS	Level	-0.007	0.006	-0.008	0.019	-0.010	0.036	-0.006	0.050
Censored NLS	Level	0.021	0.006	-0.002	0.019	0.022	0.032	-0.001	0.051
PPML	Level	0.134	0.014	0.023	0.029	0.145	0.020	0.032	0.047
Truncated PPML	Level	-0.058	0.018	-0.047	0.028	-0.062	0.025	-0.036	0.042
GPML	Level	0.863	0.279	0.223	0.227	1.053	0.221	0.350	0.197
NBML	Level	0.118	0.017	0.128	0.136	0.191	0.026	0.190	0.054
		Case 3: $V[y_i x]=(\exp(x_i\beta))^2$				Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i})(\exp(x_i\beta))^2$			
Truncated OLS	Log	-0.242	0.047	-0.089	0.066	-0.345	0.059	-0.258	0.086
OLS (ln(y+0.1))	Log	0.330	0.022	-0.171	0.068	0.253	0.026	-0.350	0.077
Truncated NLS	Level	0.096	1.196	-0.172	1.172	0.311	2.464	-0.468	2.5481
Censored NLS	Level	0.144	1.186	-0.168	1.173	0.376	2.450	-0.475	2.547
PPML	Level	0.141	0.087	0.038	0.107	0.090	0.114	0.022	0.150
Truncated PPML	Level	-0.083	0.119	-0.021	0.113	-0.157	0.165	-0.015	0.154
GPML	Level	1.186	0.167	0.474	0.178	0.762	0.198	0.316	0.222
NBML	Level	0.244	0.047	0.244	0.088	0.204	0.068	0.220	0.130

Table 6. Monte Carlo results with limited-dependent estimators, Heckman DGP

Estimator	Dependent Variable Form	β_1		β_2		β_1		β_2		
		Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	
ET-Tobit ET-Tobit Poisson-Tobit Heckman-ML Heckman-2SLS Heckman-ML Heckman-2SLS ZIP	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$				
	Level	-0.059	0.013	-0.095	0.025	-0.133	0.054	-0.202	0.057	
	Log	-0.398	0.018	-0.435	0.024	-0.353	0.026	-0.408	0.034	
	Level	0.135	0.015	0.024	0.03	0.145	0.02	0.032	0.047	
	Log	-0.001	0.029	0.029	0.039	0.034	0.032	0.056	0.047	
	Log	0.105	0.029	0.077	0.039	0.102	0.037	0.082	0.05	
	Level	0.013	0.127	0.082	1.478	0.028	0.312	-0.051	0.59	
	Level	0.076	0.006	0.051	0.025	0.083	0.03	0.053	0.06	
	Level	-0.023	0.017	-0.01	0.027	-0.027	0.025	0.001	0.043	
	ET-Tobit ET-Tobit Poisson-Tobit Heckman-ML Heckman-2SLS Heckman-ML Heckman-2SLS ZIP	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i})(\exp(x_i\beta))^2$			
Level		-0.229	1.738	-0.546	0.933	-0.244	2.642	-0.662	1.504	
Log		-0.171	0.051	-0.285	0.061	-0.097	0.108	-0.076	0.06	
Level		0.108	0.065	0.01	0.12	0.031	0.081	-0.036	0.157	
Log		0	0.045	-0.002	0.064	0.033	0.057	-0.168	0.085	
Log		-0.002	0.058	-0.002	0.07	0.042	0.078	-0.162	0.09	
Level		0.196	1.392	0.155	10.93	0.502	2.546	-0.408	2.996	
Level		0.223	1.17	-0.116	1.17	0.448	5.548	-0.434	2.567	
Level		-0.054	0.117	0.008	0.113	-0.12	0.165	0.017	0.154	

Table 7. Monte Carlo results with standard estimators, Heckman DGP à la HMR

Estimator	Dependent	β_1		β_2		β_1		β_2	
	Variable Form	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.
Truncated OLS OLS (ln(y+0.1)) Truncated NLS Censored NLS PPML Truncated PPML GPML NBML	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$			
	Log	0.02	0.021	0.067	0.019	0.041	0.035	0.086	0.023
	Log	0.036	0.027	0.204	0.014	0.036	0.033	0.202	0.015
	Level	-0.002	0.029	0.05	0.015	-0.002	0.04	0.05	0.023
	Level	0.01	0.079	0.36	0.042	0.01	0.085	0.359	0.047
	Level	-0.003	0.071	0.407	0.014	-0.003	0.073	0.407	0.017
	Level	-0.001	0.017	0.039	0.006	-0.001	0.027	0.039	0.011
	Level	0.002	0.113	0.62	0.054	0.002	0.119	0.619	0.054
	Level	-0.009	0.094	-	0.101	-0.009	0.101	0.586	0.031
	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$			
Truncated OLS	Log	0.001	0.092	0.037	0.045	0.026	0.114	-0.228	0.064
OLS (ln(y+0.1))	Log	-0.046	0.055	0.104	0.023	0.001	0.064	-0.065	0.033
Truncated NLS	Level	0.13	3.739	0.382	5.925	0.183	8.807	2.324	19.734
Censored NLS	Level	0.165	3.683	0.656	5.79	0.393	10.134	2.985	21.05
PPML	Level	-0.006	0.171	0.402	0.104	-0.011	0.35	0.369	0.243
Truncated PPML	Level	-0.004	0.156	0.033	0.108	-0.011	0.343	-0.003	0.263
GPML	Level	-0.003	0.158	0.619	0.069	-0.002	0.166	0.613	0.084
NBML	Level	-0.011	0.145	0.589	0.057	-0.001	0.157	0.582	0.076

Table 8. Monte Carlo results with limited-dependent variable estimators, Heckman DGP à la HMR

Estimator	Dependent Variable Form	β_1		β_2		β_1		β_2	
		Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.
		Case 1: $V[y_i x]=1$				Case 2: $V[y_i x]=\exp(x_i\beta)$			
ET-Tobit	Level	-0.082	0.036	-0.023	0.027	-0.124	0.048	-0.058	0.038
ET-Tobit	Log	-0.072	0.025	-0.044	0.021	-0.21	0.041	-0.178	0.032
Poisson-Tobit	Level	~	~	~	~	~	~	~	~
Heckman-ML	Log	0.019	0.021	0.067	0.019	0.041	0.035	0.086	0.023
Heckman-2SLS	Log	0.035	0.021	0.821	0.059	0.045	0.030	0.787	0.079
Heckman-ML	Level	0.020	0.021	0.067	0.019	0.041	0.035	0.086	0.023
Heckman-2SLS	Level	-0.016	0.072	0.313	0.040	-0.016	0.079	0.314	0.046
ZIP	Level	0.001	0.014	0.011	0.005	-0.001	0.029	0.011	0.015
ZIP2	Level	0.005	0.005	0.001	0.003	0.006	0.021	0.001	0.110
HMR	Level	0.018	0.021	0.042	0.022	0.041	0.035	0.084	0.026
		Case 3: $V[y_i x]=(\exp(x_i\beta))^2$				Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$			
ET-Tobit	Level	-0.500	0.248	-0.323	0.246	-0.834	1.391	-0.644	0.252
ET-Tobit	Log	-0.327	0.091	-0.044	0.07	-0.323	0.126	-0.043	0.07
Poisson-Tobit	Level	~	~	~	~	~	~	~	~
Heckman-ML	Log	0.002	0.091	0.037	0.045	0.026	0.114	-0.229	0.064
Heckman-2SLS	Log	0.002	0.091	0.037	0.045	0.026	0.115	-0.227	0.064
Heckman-ML	Level	0.123	0.734	0.216	0.767	0.935	14.708	3.048	25.905
Heckman-2SLS	Level	0.154	3.686	0.637	5.802	0.344	9.704	2.909	20.398
ZIP	Level	-0.01	0.233	-0.001	0.154	-0.025	0.508	-0.064	0.352
ZIP2	Level	0.002	0.157	-0.006	0.106	-0.005	0.343	-0.047	0.252
HMR	Level	0.001	0.093	0.224	0.052	0.032	0.119	0.113	0.072

Note: ~ denotes cases in which either the maximum likelihood estimator does not converge or it does but yields unreasonably huge bias and standard errors.

Table 9: Traditional Gravity Equation

Independent Variables	Truncated OLS	PPML	Truncated PPML	Heckman 2SLS	Heckman ML	ZIP	ZIP2	ET-Tobit	HMR
Log exporter's GDP	0.959 ^c (0.012)	0.715 ^c (0.030)	0.702 ^c (0.030)	1.079 ^c (0.019)	1.002 ^c (0.015)	0.702 ^c (0.030)	0.792 ^c (0.050)	1.091 ^c (0.012)	0.922 ^c (0.074)
Log importer's GDP	0.827 ^c (0.012)	0.740 ^c (0.032)	0.729 ^c (0.032)	0.908 ^c (0.015)	0.856 ^c (0.013)	0.729 ^c (0.032)	0.795 ^c (0.049)	0.917 ^c (0.011)	0.721 ^c (0.053)
Log exporter's GDPC	0.199 ^c (0.018)	0.185 ^c (0.060)	0.183 ^c (0.059)	0.213 ^c (0.018)	0.204 ^c (0.017)	0.183 ^c (0.059)	0.211 ^c (0.061)	0.216 ^c (0.017)	0.134 ^c (0.020)
Log importer's GDPC	0.082 ^c (0.018)	0.150 ^c (0.051)	0.148 ^c (0.050)	0.103 ^c (0.018)	0.089 ^c (0.018)	0.148 ^c (0.050)	0.173 ^c (0.049)	0.137 ^c (0.016)	-0.119 ^c (0.023)
Log distance	-1.195 ^c (0.035)	-0.764 ^c (0.060)	-0.755 ^c (0.060)	-1.289 ^c (0.037)	-1.229 ^c (0.036)	-0.755 ^c (0.060)	-0.859 ^c (0.069)	-1.243 ^c (0.032)	-0.922 ^c (0.078)
Contiguity	0.284 ^a (0.133)	0.341 ^b (0.124)	0.351 ^b (0.125)	0.153 (0.152)	0.239 ^c (0.150)	0.351 ^b (0.125)	0.207 (0.140)	-0.142 ^b (0.129)	-0.071 ^c (0.230)
Common language	0.736 ^c (0.067)	0.692 ^c (0.153)	0.698 ^c (0.152)	0.830 ^c (0.067)	0.780 ^c (0.066)	0.698 ^c (0.152)	0.737 ^c (0.152)	0.856 ^c (0.064)	0.646 ^c (0.083)
Colonial tie	0.374 ^c (0.073)	0.023 (0.160)	0.018 (0.160)	0.414 ^c (0.071)	0.389 ^c (0.071)	0.018 (0.160)	0.055 (0.158)	0.364 ^c (0.067)	0.334 ^c (0.076)
Landlocked_exporter	-0.050 (0.065)	-0.807 ^c (0.183)	-0.814 ^c (0.184)	-0.048 (0.066)	-0.049 (0.066)	-0.814 ^c (0.184)	-0.812 ^c (0.190)	-0.238 ^c (0.058)	-0.079 ^b (0.051)
Landlocked_importer	-0.682 ^c (0.063)	-0.789 ^c (0.152)	-0.795 ^c (0.152)	-0.710 ^c (0.065)	-0.692 ^c (0.064)	-0.795 ^c (0.152)	-0.828 ^c (0.158)	-0.744 ^c (0.054)	-0.384 ^c (0.053)
Exporter's remoteness	0.533 ^c (0.082)	0.745 ^c (0.141)	0.736 ^c (0.142)	0.563 ^c (0.081)	0.543 ^c (0.080)	0.736 ^c (0.142)	0.756 ^c (0.135)	0.442 ^c (0.073)	0.464 ^c (0.085)
Importer's remoteness	-0.196 ^b	0.548 ^c	0.539 ^c	-0.207 ^c	-0.201 ^b	0.539 ^c	0.511 ^c	0.120	-0.120

	(0.088)	(0.122)	(0.123)	(0.084)	(0.083)	(0.123)	(0.119)	(0.078)	(0.088)
FTA	0.488 ^c	0.218 ^b	0.214 ^b	0.470 ^c	0.483 ^c	0.214 ^b	0.341 ^c	-0.082 ^b	-0.765 ^c
	(0.101)	(0.097)	(0.099)	(0.112)	(0.110)	(0.099)	(0.104)	(0.082)	(0.215)
Openness	-0.217 ^c	-0.154	-0.182	-0.112	-0.179 ^b	-0.182	-0.133	0.060	-0.234 ^c
	(0.054)	(0.132)	(0.134)	(0.052)	(0.051)	(0.134)	(0.140)	(0.051)	(0.063)
Z							0.867		15.247 ^c
							(0.700)		(0.377)
Z ²							-0.282		-4.766 ^c
							(0.225)		(0.136)
Z ³							0.022		0.471 ^c
							(0.023)		(0.017)
Inverse Mills ratio									3.326 ^c
									(0.159)
Number of observations	8857	17028	8857	17028	17028	17028	17028	17028	17028
Excluded variable:	No	No	No	Yes	Yes	No	No	No	Yes
Common religion									
Hetero. correction	No	No	No	No	No	No	Yes	Yes	No
γ								2.418 ^c	
								(0.042)	
δ								-0.090 ^c	
								(0.004)	

Notes: (1) ^a, ^b and ^c denote 10%, 5% and 1% level of significance, respectively. (2) For the HMR estimator a polynomial with degree 3 in z (i.e. z^1 , z^2 and z^3) are included in order to control for the firm heterogeneity. Specifically, we first estimate the Probit equation predicting whether or not country j and country k have trade with each other. From the Probit equation estimates we obtain \hat{p}_{jk} the predicted probability of exports from j to k . For exporter j and importer k , the inverse Mills ratio is equal to $\hat{v}_{jk}^* = \phi(\hat{z}_{jk}^*) / \Phi(\hat{z}_{jk}^*)$ where ϕ and Φ denote the density and the cdf. of the unit-normal distribution, respectively and $\hat{z}_{jk}^* = \Phi^{-1}(\hat{p}_{jk})$. The variable controlling for firm heterogeneity, z , is defined as $z_{jk} = \hat{v}_{jk}^* + \hat{z}_{jk}^*$. The ZIP2 model includes the firm number variable from the HMR model.

Table 10: Anderson-vanWincoop Gravity Equation

Independent Variables	Truncated OLS	PPML	Truncated PPML	Heckman 2SLS	Heckman ML	ZIP	ZIP2	ET-Tobit	HMR
Log distance	-1.366 ^c (0.032)	-0.747 ^c (0.043)	-0.768 ^c (0.045)	-1.378 ^c (0.032)	-1.371 ^c (0.032)	-0.768 ^c (0.045)	-0.693 ^c (0.069)	-1.302 ^c (0.031)	-0.945 ^c (0.045)
Contiguity	0.179 (0.132)	0.632 ^c (0.107)	0.605 ^c (0.106)	0.164 (0.125)	0.172 (0.131)	0.605 ^c (0.106)	0.821 ^c (0.168)	-0.067 (0.049)	0.089 ^c (0.192)
Common language	0.432 ^c (0.071)	0.144 (0.099)	0.191 ^b (0.102)	0.443 ^c (0.067)	0.437 ^c (0.067)	0.191 ^b (0.102)	0.417 ^c (0.133)	0.532 ^c (0.061)	0.322 ^c (0.068)
Colonial tie	0.639 ^c (0.074)	0.164 (0.140)	0.117 (0.141)	0.916 ^c (0.055)	0.643 ^c (0.070)	0.117 ^c (0.141)	-0.027 (0.168)	0.635 ^c (0.064)	0.398 ^c (0.070)
FTA	0.370 ^b (0.103)	0.390 ^c (0.083)	0.383 ^c (0.082)	0.291 ^c (0.098)	0.361 ^b (0.103)	0.383 ^c (0.082)	0.500 ^c (0.133)	-0.197 (0.095)	0.623 ^c (0.140)
Importer F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Obs.	8857	17028	8857	17028	17028	17028	17028	17028	17028
Excluded Res. Variable									
Common religion	No	No	No	Yes	Yes	No	No	No	Yes
Hetero. Correction	No	No	No	No	No	No	No	Yes	No
γ								1.948 ^c (0.049)	
δ								-0.066 ^c (0.005)	
Inverse Mills ratio									2.077 ^c (0.073)
Control for firm hetero.	No	No	No	No	No	No	Yes	No	Yes

Table 10: Anderson-vanWincoop Gravity Equation

Independent Variables	Truncated OLS	PPML	Truncated PPML	Heckman 2SLS	Heckman ML	ZIP	ZIP2	ET-Tobit	HMR
z									12.564 ^c (0.298)
z^2									-3.720 ^c (0.115)
z^3									0.360 ^c (0.013)

Notes: (1) ^a, ^b and ^c denotes 10%, 5% and 1% level of significance, respectively. (2) For the HMR estimator a polynomial with degree 3 in z (i.e. z^1 , z^2 and z^3) are included in order to control for the firm heterogeneity. Specifically, we first estimate the Probit equation predicting whether or not country j and country k have trade with each other. From the Probit equation estimates we obtain \hat{p}_{jk} the predicted probability of exports from j to k . For exporter j and importer k , the inverse Mills ratio is equal to $\hat{v}_{jk}^* \equiv \phi(\hat{z}_{jk}^*) / \Phi(\hat{z}_{jk}^*)$ where Φ and ϕ denote the density and the c.d.f. of the unit-normal distribution, respectively and $\hat{z}_{jk}^* = \Phi^{-1}(\hat{p}_{jk})$. The variable controlling for firm heterogeneity, z , is defined as $z_{jk} \equiv \hat{v}_{jk}^* + \hat{z}_{jk}^*$. The ZIP2 estimator includes the firm number variable from the HMR model

Table 11. Monte Carlo Simulation Results with x Values Concentrated near Zero

Estimator		Bias	Std. Error	Bias	Std. Error	Bias	Std. Error	Bias	Std. Error
	Variable form	Case 1		Case 2		Case 3		Case 4	
Standard Estimators									
Truncated OLS	Level	0.0874	0.0096	0.0394	0.0074	0.0635	0.0051	-0.2838	0.0081
OLS (ln(y+0.01))	Log	-0.5264	0.0018	-0.5097	0.0012	-0.4966	0.0013	-0.5532	0.0021
Truncated NLS	Level	-0.0003	0.0008	-0.0027	0.034	-1.4765	5.7881	-3.9862	18.9271
Censored NLS	Log	-0.0003	0.0008	-0.0027	0.034	-1.4765	5.7881	-3.9857	18.9265
PPML	Log	0.0013	0.0016	0.0017	0.0016	0.0079	0.0649	-0.0274	0.3547
Truncated PPML	Level	-0.0039	0.0016	-0.0013	0.0017	0.0068	0.0653	-0.0341	0.3531
GPML	Level	0.1727	0.1242	0.1622	0.0794	0.9248	0.0399	0.1619	0.0794
NBML	Level	0.0001	0.0052	0.0018	0.0016	0.0188	0.0085	0.0106	0.0318
Lim. Dep. Estimators									
ET-Tobit	Log	0.0691	0.0827	0.0404	0.0786	0.0536	0.1229	-0.0623	0.0179
Heckman-ML	Log	0.1692	0.0071	0.1163	0.0053	0.0644	0.0048	-0.0505	0.0104
Heckman-2SLS	Log	0.3353	0.0105	0.0268	0.0086	0.0653	0.0056	0.0009	0.0153
ZIP	Level	0.0013	0.0016	0.0017	0.0016	0.0079	0.0649	-0.0274	0.3547

Above are the simulation results when $y = \exp(1 + x_1) * \eta$ where x_1 has mean -5.161901 and standard deviation 3.503101. Please note that x_1 has the same mean and the same standard deviation as $\text{Log}[(\text{GDP_Exporter} * \text{GDP_Importer}) / (1.3 \text{ Bilateral distance})]$. GDP_Exporter, GDP_Importer and Bilateral Distance are the real data GDPs of exporting countries and importing countries and their bilateral distance normalized by their respective mean. Each simulated sample has 18360 observations, which is equal to the empirical sample used by SS&T.

**Appendix Table 1: Indicators of the degree of censoring for the ET-Tobit
DGP**

<i>Percentage of Zero Trade Values</i>		
	<i>k = 1.0</i>	<i>k = 1.5</i>
	%	%
Case 1	41	51
Case 2	44	54
Case 3	49	60
Case 4	55	64
<i>Percentage change in Mean</i>		
Case 1	15	38
Case 2	15	40
Case 3	16	43
Case 4	19	50

Notes: the results are the averages we obtain after we simulate 10000 samples of 1000 observations for each pattern of heteroscedasticity.

Appendix Table 2. Monte Carlo results with standard estimators, ET-Tobit DGP with k=1.5

Estimator	Dependent Variable Form	β_1		β_2		β_1		β_2		
		Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	
Truncated OLS OLS (ln(y+0.1)) Truncated NLS Censored NLS PPML Truncated PPML GPML NBML	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$				
	Log	0.227	0.072	0.236	0.102	0.187	0.069	0.188	0.109	
	Log	0.230	0.031	0.329	0.062	0.186	0.033	0.278	0.069	
	Level	0.143	0.031	0.160	0.035	0.124	0.050	0.135	0.083	
	Level	0.210	0.052	0.262	0.053	0.173	0.055	0.205	0.087	
	Level	0.356	0.049	0.400	0.056	0.258	0.039	0.288	0.062	
	Level	0.151	0.035	0.153	0.049	0.111	0.030	0.107	0.065	
	Level	0.532	0.269	0.521	0.315	0.781	0.266	0.749	0.266	
	Level	0.345	0.034	0.369	0.048	0.487	0.043	0.490	0.078	
Truncated OLS OLS (ln(y+0.1)) Truncated NLS Censored NLS PPML Truncated PPML GPML NBML	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$				
	Log	0.015	0.079	0.008	0.135	0.248	0.091	-0.161	0.156	
	Log	0.026	0.040	0.097	0.082	0.091	0.044	0.189	0.092	
	Level	0.432	22.881	0.090	3.070	0.676	27.516	0.241	6.858	
	Level	0.559	23.806	0.250	2.195	0.903	24.749	0.391	6.292	
	Level	0.337	0.100	0.384	0.164	0.258	0.132	0.351	0.208	
	Level	0.009	0.118	0.009	0.166	-0.169	0.167	0.034	0.2167	
	Level	1.255	0.248	1.179	0.311	0.644	0.239	0.7508	0.3229	
	Level	0.635	0.075	0.6488	0.1413	0.4629	0.1123	0.5805	0.2041	

Appendix Table 3. Monte Carlo results with limited dependent variable estimators, ET-Tobit DGP with k=1.5

Dependent		β_1		β_2		β_1		β_2		
Estimator	Variable Form	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	Bias	Std Error.	
	Case 1: $V[y_i x]=1$					Case 2: $V[y_i x]=\exp(x_i\beta)$				
ET-Tobit	Level	0.0368	0.0303	0.0414	0.0348	0.1847	0.0836	0.1980	0.0889	
ET-Tobit	Log	-0.2401	0.0435	-0.2494	0.0507	-0.1884	0.0494	-0.1908	0.0613	
Poisson-Tobit	Level	0.3570	0.0471	0.3998	0.0549	0.3562	0.0492	0.3983	0.0739	
Heckman-ML	Log	0.3446	0.0568	0.3584	0.0906	0.3109	0.0548	0.3158	0.1012	
Heckman-2SLS	Log	0.7749	0.0960	0.8052	0.1242	0.7081	0.1094	0.7274	0.1438	
Heckman-ML	Level	0.1805	0.1270	0.2008	0.9312	0.1002	0.1287	0.405	4.6382	
Heckman-2SLS	Level	0.0523	0.0045	-0.2292	0.0126	0.0532	0.0226	-0.235	0.0378	
ZIP	Level	0.2620	0.0230	0.2920	0.0450	0.2290	0.0350	0.2460	0.0810	
	Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i})$ $(\exp(x_i\beta))^2$				
ET-Tobit	Level	0.5731	0.3302	0.6209	0.2696	0.7267	0.0919	0.7688	0.0894	
ET-Tobit	Log	0.0131	0.1628	0.0741	0.0770	0.0809	0.0915	0.0525	0.1304	
Poisson-Tobit	Level	0.3279	0.0881	0.3752	0.1591	0.2047	0.088	0.2271	0.1239	
Heckman-ML	Log	0.1255	0.1041	0.1211	0.1518	-0.0739	0.1779	-0.1264	0.1939	
Heckman-2SLS	Log	0.3477	0.1999	0.3492	0.2336	0.4101	0.3653	0.3465	0.3172	
Heckman-ML	Level	0.3870	3.5500	2.6472	72.202	0.4410	2.5477	1.1797	17.435	
Heckman-2SLS	Level	0.0329	0.1489	-0.2386	0.1501	0.0096	0.1932	-0.2572	0.2202	
ZIP	Level	0.0960	0.1310	0.1180	0.1860	-0.1130	0.2000	0.1447	0.2366	

Appendix Table 4 Average Absolute Bias and Standard Errors Across the Four Cases

		β_1	SE	β_2	SE	β_{Ave}	SE
ET DGP- Standard Estimators							
Truncated OLS	Log	0.16	0.11	0.16	0.07	0.16	0.09
OLS (ln(y+0.1))	Log	0.25	0.07	0.15	0.03	0.2	0.05
Truncated NLS	Level	0.13	2.51	0.31	11.43	0.22	6.97
Censored NLS	Level	0.21	2.16	0.4	12.21	0.3	7.18
PPML	Level	0.28	0.11	0.24	0.07	0.26	0.09
Truncated PPML	Level	0.06	0.11	0.08	0.06	0.07	0.08
GPML	Level	0.6	0.23	0.61	0.18	0.61	0.2
NBML	Level	0.17	0.08	0.16	0.04	0.16	0.06
ET DGP-Limited Dependent Estimators							
ET-Tobit	Level	0.4	0.15	0.38	0.09	0.39	0.12
ET-Tobit	Log	0.12	0.08	0.13	0.05	0.13	0.07
Poisson-Tobit	Level	0.26	0.1	0.23	0.06	0.25	0.08
Heckman-ML	Log	0.22	0.11	0.18	0.07	0.2	0.09
Heckman-2SLS	Log	0.44	0.16	0.46	0.14	0.45	0.15
Heckman-ML	Level	0.17	7.47	0.21	1.45	0.19	4.46
Heckman-2SLS	Level	0.18	0.11	0.03	0.09	0.1	0.1
ZIP	Level	0.2	0.13	0.14	0.09	0.17	0.11
Heckman DGP-Standard Estimators							
Truncated OLS	Log	0.2	0.04	0.1	0.06	0.15	0.05
OLS (ln(y+0.1))	Log	0.36	0.02	0.15	0.07	0.26	0.04
Truncated NLS	Level	0.11	0.93	0.16	0.41	0.13	0.67
Censored NLS	Level	0.14	0.92	0.16	0.95	0.15	0.93
PPML	Level	0.13	0.06	0.03	0.08	0.08	0.07
Truncated PPML	Level	0.09	0.08	0.03	0.08	0.06	0.08
GPML	Level	0.97	0.22	0.34	0.21	0.65	0.21
NBML	Level	0.19	0.04	0.2	0.1	0.19	0.07
Heckman DGP-Limited Dependent Estimators							
ET-Tobit	Level	0.17	1.11	0.38	0.63	0.27	0.87
ET-Tobit	Log	0.25	0.05	0.3	0.04	0.28	0.05
Poisson-Tobit	Level	0.1	0.05	0.03	0.09	0.07	0.07
Heckman-ML	Log	0.02	0.04	0.06	0.06	0.04	0.05
Heckman-2SLS	Log	0.06	0.05	0.08	0.06	0.07	0.06
Heckman-ML	Level	0.18	1.09	0.17	4	0.18	2.55
Heckman-2SLS	Level	0.21	1.69	0.16	0.96	0.19	1.32
ZIP	Level	0.06	0.08	0.01	0.08	0.03	0.08
HMR DGP-Standard Estimators							
Truncated OLS	Log	0.02	0.07	0.1	0.04	0.06	0.05
OLS (ln(y+0.1))	Log	0.03	0.04	0.14	0.02	0.09	0.03
Truncated NLS	Level	0.08	3.15	0.7	6.42	0.39	4.79
Censored NLS	Level	0.14	3.5	1.09	6.73	0.62	5.11

Appendix Table 4 Average Absolute Bias and Standard Errors Across the Four Cases

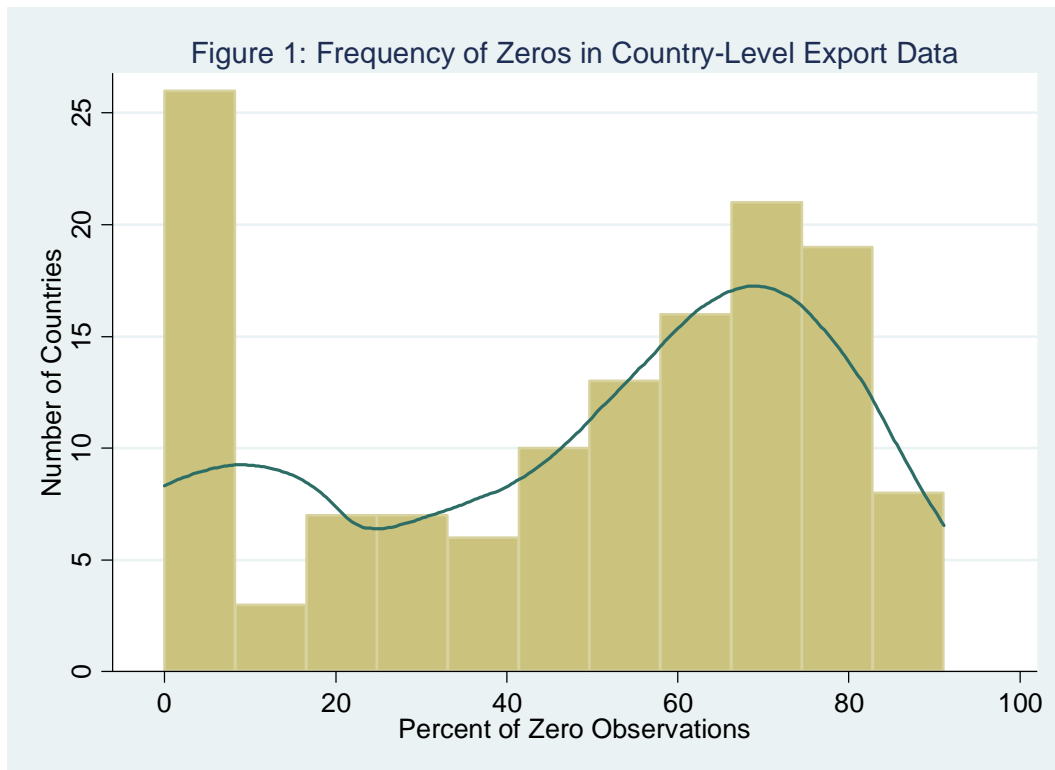
		β_1	SE	β_2	SE	β_{Ave}	SE
PPML	Level	0.01	0.17	0.4	0.09	0.2	0.13
Truncated PPML	Level	0	0.14	0.03	0.1	0.02	0.12
GPML	Level	0	0.14	0.62	0.07	0.31	0.1
NBML	Level	0.01	0.12	0.44	0.07	0.22	0.1
HMR DGP- Limited-Dependent Estimators							
ET-Tobit	Level	0.39	0.43	0.26	0.14	0.32	0.29
ET-Tobit	Log	0.23	0.07	0.08	0.05	0.16	0.06
Poisson-Tobit	Level	0.14	0.52	0.13	0.16	0.13	0.34
Heckman-ML	Log	0.02	0.07	0.1	0.04	0.06	0.05
Heckman-2SLS	Log	0.03	0.06	0.47	0.06	0.25	0.06
Heckman-ML	Level	0.28	3.87	0.85	6.68	0.57	5.28
Heckman-2SLS	Level	0.13	3.39	1.04	6.57	0.59	4.98
ZIP	Level	0.01	0.2	0.02	0.13	0.02	0.16
ZIP with x_3	Level	0	0.13	0.01	0.12	0.01	0.12
HMR	Level	0.02	0.07	0.12	0.04	0.07	0.06

Annex A

Patterns of Zeros in Trade Data

Since the purpose of this paper is to investigate the performance of different estimators when trade is characterized by frequent zero flows it is important to know what patterns of zeros are observed in real trade data. In particular, we are interested in how frequently zeros appear and whether they appear to be distributed independently of the variables that determine levels of nonzero trade flows. We look at the frequency of zeros using two samples of trade data at different levels of disaggregation. The first is the SST (2006) data set of 18360 observations on bilateral aggregate trade flows for 136 exporters and 135 importers in 1990. The second data set, developed under the leadership of Robert Feenstra and available from the NBER website, is for product-level US exports to more than 100 countries for more than 9000 ten-digit HS classifications.²⁰ We use the average of U.S. exports from 2002 to 2006 to identify consistently zero trade flows.

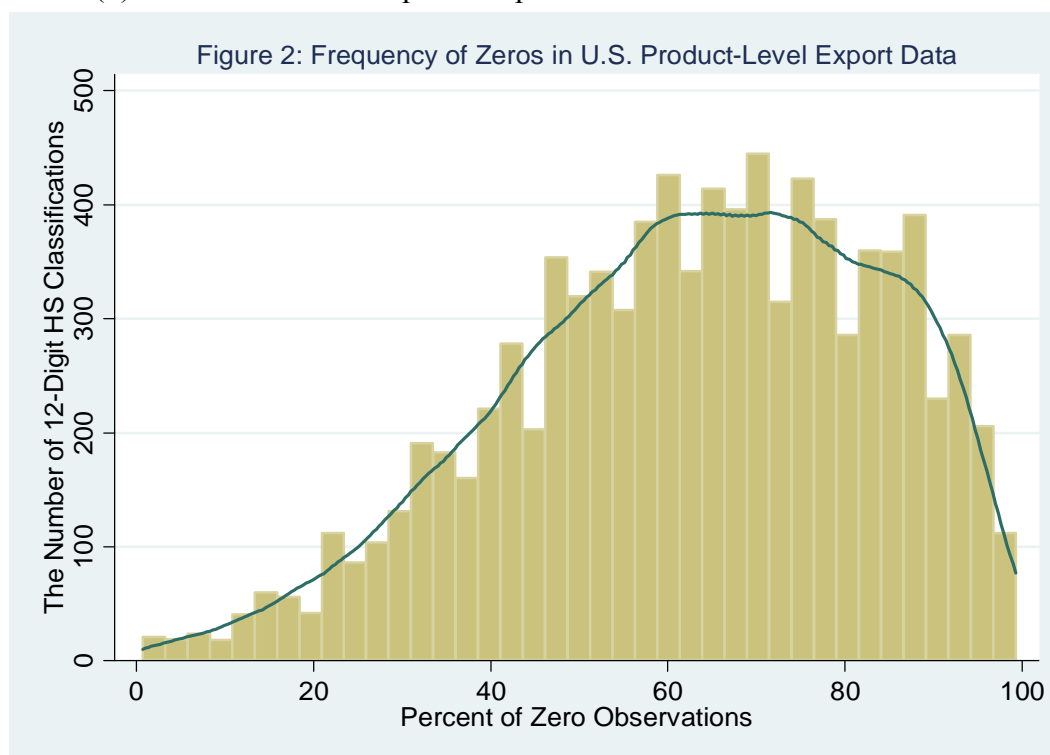
The percentage of zero trade flows in the SST (2006) data set is 47.6%. Only 40% of the sample involves trade in both directions while 7.4% of the sample involves trade flows in only one direction. Table A.1, which shows in detail the features of the country-level trade data, reveals that smaller and poorer economies export to a much smaller number of destinations than other countries. Twenty-six such countries have zero trade with over 100 countries. Figure 1 presents a histogram and a kernel density for the percentages of zeros in the exports of the 136 exporters. It shows that 26 countries of the



Notes: (1) The kernel density estimate shown by the solid line is calculated using the Epanechnikov kernel function with a band width equal to 0.0822.

²⁰ See for example Feenstra, Romalis and Schott (2002). The link to the data is: <http://cid.econ.ucdavis.edu/data/sasstata/usxss.html>.

(2) The total number of possible positive trade flows is 135.



Notes: (1) The kernel density estimate is calculated using the Epanechnikov kernel function with a band width equal to 0.025.

(2) The total number of 10-digit products the U.S. exports is 9037.

sample—all of which are high-income economies—have zero exports to under 8% of their possible foreign markets. However, many other countries have zero exports to between 40% and 90% of their potential trading partners.

In U.S. product-level exports at the 10 digit level, zero trade flows are also very frequent. This is despite the fact that the U.S. is the largest economy in the world and one of the largest exporters, so the frequency of zeros is likely much lower than in the exports of other countries at this level of disaggregation.²¹ Zero observations represent 63% of the total 1346513 (149 destinations * 9037 products) possible export flows. Figure 2 provides a histogram and a kernel density for the percentages of zeros in U.S. exports in the 9037 10-digit H.S. classifications. It shows that for 7702 of the 9037 H.S. classifications (i.e. 85%), the U.S. had zero exports to 40% or more of the 149 foreign destinations for the period 2002 - 2006.

Table A.2 presents the patterns of zero trade flows in both samples by groups of exporters' GDP, by importers' GDP, and by groups of bilateral distances. These are standard explanatory variables for bilateral trade in gravity regressions. For both samples, the GDP of exporters and importers and the bilateral distance are clearly important determinants of the likelihood of zero bilateral trade. On average, larger exporters and larger importers have positive exports to a larger number of foreign markets. Countries are

²¹ Note that the average presence of zero exports for the period 2002-2006 is less than the presence of zero exports we would expect to observe using annual export data.

on average less likely to trade with distant markets as evidenced by the fact that the percentages of zero trade flows are increasing in distance.

To shed more light on the determinants of positive bilateral trade flows we turned to regression techniques. For the aggregate country-level data set we ran a Probit regression to predict nonzero trade flows. For US exports, we used a linear probability model (LPM) because STATA does not allow us to include the 9037 product-specific dummies needed to control for product-specific factors in the disaggregated model.²² Table 3 shows that variables such as GDP, GDPC (per capita GDP) and bilateral distance are important and statistically significant determinants of the likelihood of positive bilateral trade. A common border is found to reduce the probability of positive trade. This surprising result was also obtained by HMR (2008) in their first-stage Probit regressions. Potential explanations include: (a) that countries sharing borders are more likely to be involved in conflicts that disrupt their bilateral trade, and (b) that the dummy variable on contiguity is strongly negatively correlated with bilateral distance, and may be confounding the effect of sharing a common border. As a rough tie-breaker, we ran the same regressions excluding the bilateral distance variable, and found that the contiguity variable became positive and statistically significant.²³

The detailed analysis of the patterns of zeros in both aggregate and very disaggregated trade data suggests that—as might be expected from economic theory—zero trade flows are not randomly determined. They are the result of fundamental economic determinants such as the inability of any exporter from a particular country to meet the cost threshold for positive trade. The resulting relationships between the decision to trade and the resulting volume of trade are likely to have implications for estimation strategies.

²² We also ran a Probit regression for each of the 9037 10-digit products and on average the results are quantitatively and qualitatively the same. These results are available upon request.

²³ The strong correlation between bilateral distance and common border dummy may also explain why the common border dummy is not found to have a robust and significant positive effect on the volume of trade once trade takes place in the empirical results presented in Section VII. Results for both the Probit and the LPM models excluding the bilateral distance variables involved positive and statistically significant effects of a common border on the probability of positive bilateral trade. These results are available from the authors upon request.

Table A.1: Frequency of Zero Trade Flows

	Santos Silva & Tenreyro's Sample of Country-Level Trade Data	Sample of U.S. Product-Level Export Data
Distance		
1 st to 33 rd percentile	41%	46%
34 th to 66 th percentile	45%	71%
67 th to 100 th percentile	57%	71%
Exporter GDP		
1 st to 33 rd percentile	66%	
34 th to 66 th percentile	50%	
67 th to 100 th percentile	25%	
Importer GDP		
1 st to 33 rd percentile	62%	85%
34 th to 66 th percentile	50%	69%
67 th to 100 th percentile	23%	34%

Notes: (1) The numbers are the percentages of zero trade flows in the total number of possible trade/export flows.

Table A.2: Determinants of Positive Trade Flows

Independent Variables	Probit Regression (<i>Sample of Country-Level Bilateral Trade</i>)	LPM (<i>Sample of U.S. Product-Level Exports</i>)
Log exporter's GDP	0.458 *** (51.49)	
Log importer's GDP	0.330 *** (42.13)	0.0945 *** (15.66)
Log exporter's GDPC	0.097 *** (9.73)	
Log importer's GDPC	0.104 *** (10.25)	0.0389 *** (4.20)
Log distance	-0.461 *** (-18.14)	-0.232 *** (-10.07)
Contiguity	-0.490 *** (-4.40)	-0.315 *** (-3.64)
Common language	0.310 *** (7.99)	0.032 (0.87)
Colonial tie	0.185 *** (4.54)	0.045 (1.20)
Landlocked_exporter	0.043 (1.31)	
Landlocked_importer	-0.076 ** (-2.23)	-0.021 (-0.83)
Exporter's remoteness	0.080 (1.56)	
Importer's remoteness	-0.093 ** (-1.80)	0.234 *** (6.58)
FTA	1.210 *** (7.25)	0.035 (0.44)
Openess	0.322 *** (11.83)	-0.036 (-1.27)
Product dummies		Yes
Constant	-15.714 *** (-21.54)	-2.102 *** (-4.75)
N. of obs.	18360	1102514

Notes: (1) Probit regression results report the coefficient estimates and z-statistics in the parentheses

(2) LPM is the Linear Probability Model. T-statistics computed using the robust standard error with allowance for clustering on the exporter-importer pairs are in parentheses.

(3) Note that the number of observations of the LPM regression is not 1346513 (i.e. 149 destinations * 9037 products) because the data on GDP, GDP per capita and importer's remoteness are not available for all countries.

Annex B

Generation of the HMR DGP for Monte Carlo Simulations

Step 1

Generate a uniformly distributed random variable x_3 with 1000 observations.

Sort x_3 from lowest to highest value.

Create an indicator variable name **order** that is equal to 1, 2...1000. So order is strongly correlated with x_3 . This indicator variable will be used to merge x_3 with the data set of y_2 and other variables.

Save this data file under file name Uniform.

Step 2

Generate random variables x_1 , x_2 , y_2 with correlation matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$. In other words, the correlation between x_1 and x_2 and x_1 and y_2 is zero while the correlation between x_2 and y_2 , which is the dependent variable of the equation that determines the sample selection rule, is 0.5. By construction, all of these variables are normally distributed with mean zero and variance 1.

x_1 then is transformed into a dummy variable with 40% of zeros. Specifically, for x_1 with values below the 40th percentile x_1 is replaced by 0 and for x_1 with values above the 40th percentile x_1 is replaced by 1.

Sort the data by y_2 from lowest to highest values. The order of the other variables changes accordingly and their correlations remain intact.

Create an indicator variable **order** that is equal to 1, 2...1000. So variable **order** is strongly correlated with y_2 . This indicator variable will be used to merge with the data set of x_3 in the Uniform file we saved in Step 1.

Step 3

Merge those variables x_1 , x_2 , y_2 and x_6 above with the data file Uniform including x_3 by indicator variable **order**. **Note that the correlation between x_2 and y_2 remains intact.**

Transform the normally distributed variable $y_2 \sim N(0,1)$ so that y_2 has 25% of its values less than zero. Note that all the correlations remain intact as the transformation only shifts the distribution of y_2 to the right.

Replace the values of x_3 below the 25th percentile by order with zeros. This means that the observations with zero trade always have zero firms.

Step 4

Generate $\exp(x_i\beta)=\exp(1+x_1+x_2+x_3)$ and multiplicative error term (η) with its variance a function of $\exp(x_i\beta)$. Note that $y_1=\exp(x_i\beta) \cdot \eta$ and that the log of the multiplicative error term is $\ln(\eta)$ in the paper.

Step 5

Generate a random variable $\ln(\eta^*)$ that has a correlation of 0.5 with $\ln(\eta)$.

Step 6

Generate the excluded restriction variable x_4 : $x_4 = y_2 - [1 + x_1 + x_2 + \ln(\eta^*)]$.

Step 7

Apply the selection rule to generate zeros of y_1 as follows: for y_2 with values less than zero, replace y_1 with zero. For y_2 with values greater than zero, y_1 remains the same.

Appendix C

Non-Nested Test for the Nature of the Data Generating Process

Because we have found that the best estimator for the gravity model may depend upon the nature of the data-generating process, it may be important to try to identify the nature of the process giving rise to the observed data before proceeding to estimation. Our first step in applying the estimators to real-world data is therefore to try to assess whether real-world trade data such as those used in the SST (2006, p649) study are more consistent with the ET-Tobit model or the Heckman model.²⁴ Once we have completed this step, we turn to a comparison—along the lines of that reported by SST--of the empirical estimates arising from the different models.

Since the ET-Tobit and the Heckman models are non-nested we use the Vuong test—based on comparison of the log-likelihoods of alternative models—to discriminate between them.²⁵ The distance (or the closeness) between the true and unknown model and any model is measured using the Kullback-Leibler information criteria (KLIC). This criterion is defined as follows:

$$KLIC \equiv E[L^{True}] - E[L^*]$$

where L^{True} is the log of the conditional density of the dependent variable y given \mathbf{X} , the vector of explanatory variables (i.e. the true but unknown model) and L^* is the log of the conditional density of the dependent variable y given \mathbf{X} when an untrue but known model is applied to estimate y . Thus, minimization of the $KLIC$ is equivalent to maximization of $E[L^*]$. Vuong (1989) showed that determining which of the two models is closer to the true model is asymptotically equivalent to determining which model has an average log-likelihood statistically greater than its rival. Specifically, Vuong's test involves the likelihood-ratio statistic:

$$LR_n(\hat{\beta}_n, \hat{\theta}_n)/(\hat{\omega}_n \sqrt{N}) \equiv [L_n^{ET-Tobit}(\hat{\beta}_n) - L_n^{Heckman}(\hat{\theta}_n)]/(\hat{\omega}_n \sqrt{N})(*)$$

²⁴ We follow Helpman et al. (2008) in using common religion as our excluded restriction variable for the Heckman model.

²⁵ See Vuong (1989) for a more technical and detailed discussion of this test.

where L_n stands for the log-likelihood of the ET-Tobit and the Heckman model, while $\hat{\omega}_n^2$ is the estimated variance of the pointwise log-likelihood ratio. $\hat{\omega}_n^2$, the estimated variance of the likelihood-ratio statistic is computed as follows:

$$\hat{\omega}_n^2 \equiv \frac{1}{n} \sum_1^n \left[L_n \frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)} \right]^2 - \left[\frac{1}{n} \sum_1^n L_n \frac{f(\hat{\beta}_n)}{f(\hat{\theta}_n)} \right]^2 \quad (**)$$

where $f(\hat{\beta}_n)$ and $f(\hat{\theta}_n)$ are the individual log likelihoods of the ET-Tobit model and the Heckman model, respectively.

The numerator of the likelihood-ratio statistic needs correction for degrees of freedom. Following Clark (2001)²⁶ we define K_n , the degrees of freedom correction, as follows:

$$K_n = \left[\frac{p}{2} L_n(N) - \frac{q}{2} L_n(N) \right] = \left[\frac{16}{2} L_n(18360) - \frac{14}{2} L_n(18360) \right] = 9.81793 \quad (***)$$

where $p=16$, $q=14$ and $N=18360$ are the number of estimated coefficients in the ET-Tobit model and the Heckman model and the number of observations, respectively.

Given the results for the ET-Tobit and Heckman ML model, the likelihood-ratio statistic of the Vuong test is: $\frac{-97237.926 - (-26780.961) - 9.81793}{7.818637 * \sqrt{18360}} = -66.496$ where the values of $L_n^{ET-Tobit}$, $L_n^{Heckman}$ and $\hat{\omega}_n$ are -97237.926, -26790.961 and 7.818637, respectively. Since the z-statistic is significantly less than zero we conclude that it is more likely that this data set was generated by the Heckman sample-selection model than by a threshold-Tobit model. This suggests that we should place more emphasis on the ZIP models found in the Monte-Carlo simulations to be better estimators for Heckman-based data than on other models. We are, however, conscious of the more specific test result from Santos Silva, Tenreiro and Windmeijer (2014) that a single-index model such as the ET-Tobit model might better represent the characteristics of this particular dataset.

²⁶ Clark's paper is available at: <http://www.rochester.edu/college/psc/clarke/450315.pdf>.