

Estimating the Gravity Model When Zero Trade Flows Are Frequent

Will Martin and Cong S. Pham *

February 2008

Abstract In this paper we estimate the gravity model allowing for the pervasive issues of heteroscedasticity and zero bilateral trade flows identified in an influential recent paper by Santos Silva and Tenreyro. We use Monte Carlo simulations with data generated using a heteroscedastic, limited-dependent-variable process to investigate the extent to which different estimators can deal with the resulting parameter biases. While the Poisson Pseudo-Maximum Likelihood estimator recommended by Santos Silva and Tenreyro solves the heteroscedasticity-bias problem when this is the only problem, it appears to yield severely biased estimates when zero trade values are frequent. Standard threshold-Tobit estimators perform better as long as the heteroscedasticity problem is satisfactorily dealt with. The Heckman Maximum Likelihood estimators appear to perform well if true identifying restrictions are available.

JEL: C21, F10, F11, F12, F15

* Will Martin and Cong S. Pham are Lead Economist and Consultant, Development Economic Research Group at the World Bank, respectively. Cong S. Pham is also affiliated with the School of Accounting, Economics and Finance, Deakin University, Melbourne, Victoria, Australia. Their email addresses are wmartin1@worldbank.org. and cpham68@gmail.com . The authors would like to thank João Santos Silva and Silvana Tenreyro for their helpful comments and data. They also would like to thank Caroline Freund, Hiau Looi Kee, Daniel Lederman and participants in presentations at the World Bank and the European Trade Study Group 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 partly to its apparently good performance in representing trade flows and partly to the strong theoretical foundations provided in papers such as Anderson (1979) and Anderson and van Wincoop (2003). In an influential recent paper, Santos Silva and Tenreyro (2006) focused on econometric problems resulting from heteroscedastic residuals and the prevalence of zero bilateral trade flows. Using Monte Carlo simulations, they showed that traditional estimators are likely to yield severely-biased parameter estimates and identified a preferred estimator that appeared to overcome these problems. Applying this estimator to real-world data, they obtained results that raised serious questions about one of the key theoretical predictions of Anderson and van Wincoop (2003)—a unit coefficient on GDP—and about previous empirical estimates of distance effects. Given the widespread use the gravity model in modern empirical trade analysis, the questions raised by Santos Silva and Tenreyro have justly received a great deal of attention.

The first concern raised by Santos Silva and Tenreyro (2006) was a fundamental one—the fact that, by Jensen’s inequality, a log linear model cannot be expected to provide unbiased estimates of mean effects when the errors are heteroscedastic. The existence of the problem is non-controversial. What was striking was the magnitude of the apparent bias, with popular estimators frequently yielding estimates biased by 35 percent or more (positively or negatively) in large samples.

The second concern emphasized by Santos Silva and Tenreyro (2006) arises from the prevalence of zero values of the dependent variables, which are undefined when converted into logarithms for estimation using the popular log-linear specification. Santos

Silva and Tenreyro pointed out that these values are very common—almost half of the observations in their empirical application were zero. Hurd (1979) has shown that heteroscedasticity can lead to large biases in samples truncated by exclusion of zero values—as has been the case in most estimates of the gravity model—although Arabmazar and Schmidt (1981) found these problems to be less serious in the censored regression case where the zero values are retained.

The tractable and apparently robust alternative approach to estimation recommended by Santos Silva and Tenreyro—the Poisson Pseudo-Maximum-Likelihood (PPML) estimator—has already been widely adopted in estimation of gravity equations (see, for example, Westerlund and Wilhelmsson (2007); Xuepeng Liu (2007); Hebble, Shepherd and Wilson (2007)). Santos Silva and Tenreyro show why the normal equations used to solve for the PPML estimator should make it more robust, and show using Monte Carlo simulations that its coefficient estimates are much less subject to bias resulting from heteroscedasticity. Their case for the Poisson estimator seems extremely strong for analysis of nonlinear relationships in models where zero values of the dependent variable are infrequent.¹

Given the firm theoretical underpinnings for zero trade provided in recent papers such as Baldwin and Harrigan (2007); Hallak (2006); and Helpman, Melitz and Rubinstein (2007), it seems important to follow Santos Silva and Tenreyro's suggestions (2006, p642) to examine the performance of limited-dependent variable estimators that

¹ Heteroscedasticity in nonlinear models with relatively few zero observations is likely to arise in many empirical studies in economics, including estimation of consumer demand systems, firm cost and profit functions, and consumption, investment and money demand functions.

are robust to distributional assumptions and to examine whether their endorsement of the PPML estimator holds under data generating process that generate substantial numbers of true zero observations. This re-examination seems particularly important because, for their Monte-Carlo analyses, Santos Silva and Tenreyro used a data-generating-process that generates no zero values. They did create some zero values for sensitivity analyses by rounding observations but—as they note—this is a different data generating process from that underlying the zero values in modern theoretical models where firms choose whether to trade and, if so, how much to trade (see Helpman, Melitz, and Rubinstein (2007)).

If the PPML estimator does not prove robust to the joint problems of heteroscedasticity and limited-dependent variables, then a key challenge is to identify an approach that can deal with these problems. To address this challenge, we outline a strategy for choosing estimators in the limited-dependent variable case, and then provide evidence on the performance of different estimators.

The next section of the paper considers the potential estimation problems associated with samples containing zero values of the dependent variable, particularly when heteroscedasticity is present. After considering the strategy for selecting estimators, and outlining a range of potential estimators, we use Monte Carlo analysis to assess which approaches come closest to reliably recovering the true parameter values. Finally, we investigate the implications of the choice of estimator for the parameter estimates obtained in a real-world application of the gravity model.

II. Econometric Problems Associated with Zero Trade Flows

In recent years, it has become widely recognized that the level of trade—even in the aggregate—between any two countries is frequently zero. Around half the observations

on total bilateral trade in the data sets used by Santos Silva and Tenreyro (2006) and by Helpman, Melitz, and Rubinstein (2007) were of zero trade flows—and the share in datasets involving disaggregated trade flows is frequently much higher. Baldwin and Harrigan (2007) find that 92.6 percent of potential import flows to the USA at the finest (10-digit) level of disaggregation are zero. Some of the reports of zero trade reflect non-reporting or errors and omissions and a few may reflect rounding error associated with very small trade flows. However, it appears that most of the zero trade flows between country pairs in carefully-prepared datasets reflect a true absence of trade, rather than rounding errors. Hallak (2006), Helpman, Melitz, and Rubinstein (2007) and Baldwin and Harrigan (2007) attribute these zeros to failure to meet the fixed costs associated with establishing trade flows.

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

$$(1) \quad y_{1i}^* = f(x_{1i}, \beta_1) + u_{1i}$$

$$(2) \quad y_{2i}^* = f(x_{2i}, \beta_2) + u_{2i}$$

where x_{ji} is a vector of exogenous regressors, β_i is a $K_j \times 1$ vector 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 of interest and equation (2) is a sample selection rule that determines whether observations have a non-zero value, and results in:

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

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

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_i suffer from omitted variable bias because they omit relevant explanatory variables.

The Tobit model is a special case of the Heckman model with the right hand sides of (1) and (2) identical (Heckman 1979). In this case, we must distinguish between a latent variable, y^* , and the actual realized value of y . which equals zero when $y^* \leq 0$. In this situation, a simple diagram based on Tobin's Figure 1 provides important insights into the key problems arising in applying standard approaches to estimation. Our version of the figure includes the nonlinearity and heteroscedasticity of the gravity equation emphasized by Santos Silva and Tenreiro (2006) as well as a limited-dependent variable.

Figure 1. The nature of the limited-dependent variable bias

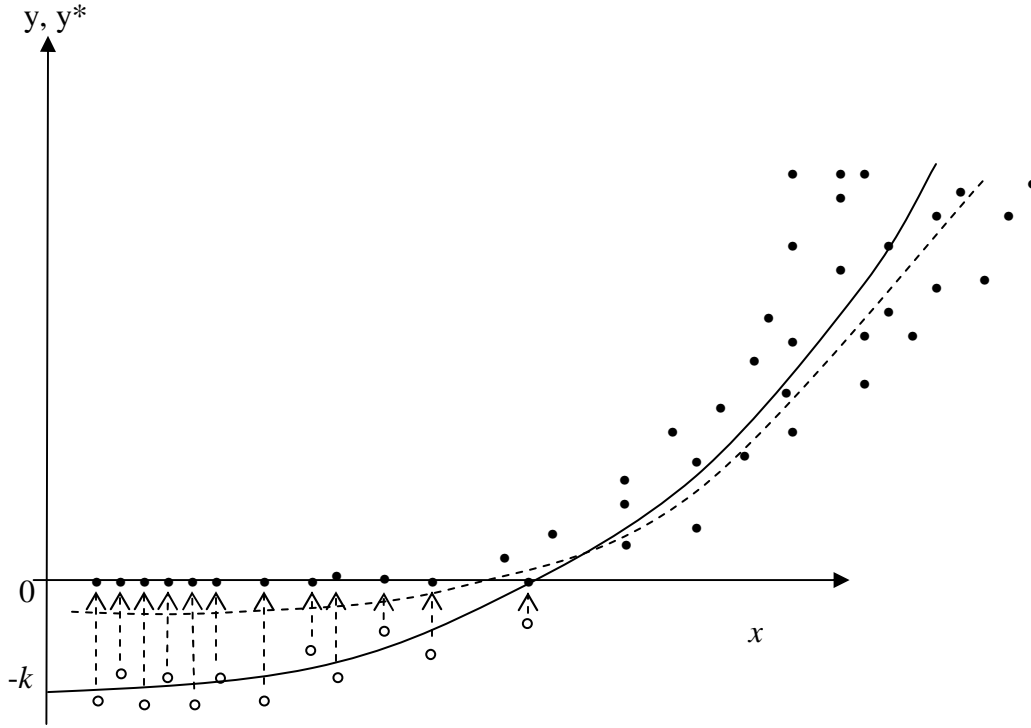


Figure 1 shows the relationship between a latent variable, y^* , and an explanatory variable, x , with the nonlinear non-stochastic relationship represented by the solid curve, and individual data points by the bullets. Observations corresponding to any value of y^* less than zero (white bullets) are observed as zero realizations of y . In this censored regression case, the residuals associated with low values of the latent variable are likely to be replaced by the 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 standard estimation procedures on a sample containing the zero values is likely to bias downward the slope of the relationship between x and y^* at all points, resulting in an estimate something like the dashed curve (Greene 1981). The diagram makes clear that a demonstration of the quality of an estimator based on its performance with non-censored data may provide little or no indication of its performance when the data generating process is characterized by a limited-dependent variable relationship.

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 of the limit observations are transformed from their initial values to zero. In the case of truncation, only those values of the error terms that lead to positive y^* values are retained. In the diagram, this leads to exclusion of the rightmost white bullet, and hence to a situation where $E(u) > 0$. 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 worse estimates.

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 on 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 (Santos Silva and Tenreiro 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 Santos Silva and Tenreiro. Observations at the zero limit reflect

the probability mass associated with all outcomes below zero, meaning that their likelihood is characterized by the distribution function, rather than a density function. If the variance of the error term for these observations is incorrectly specified—as it would be with a model assuming homoscedasticity applied to the data points represented in the diagram—this will clearly bias the realized values of the distribution function, potentially creating bias in any coefficients estimated from this sample. If the underlying data generating process involves heteroscedastic errors, this introduces a link between heteroscedasticity and bias in coefficient estimates that is quite different from the Jensen’s inequality problem.

III. 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 universal agreement that the gravity model involves relationships between variables that are nonlinear in levels, with the functional form for the underlying relationship between the explanatory variables in the levels being that used by Santos Silva and Tenreyro (2006, p644):

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

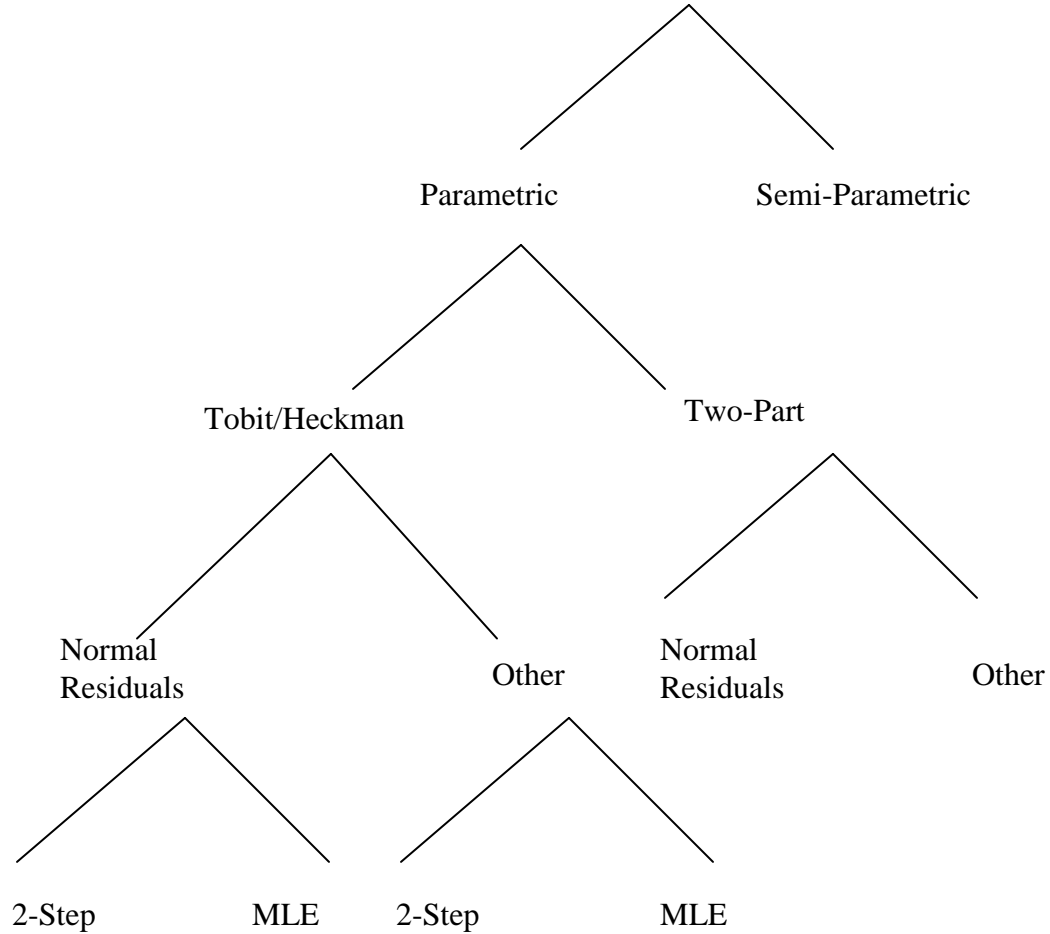
where y_i represents bilateral trade; x_i is a vector of explanatory variables (some of which may be linear, some in logarithms and some dummy variables) for observation i , β is a vector of coefficients, and ε_i is an error term whose variance, unlike those of equations (1) and (2), need not be constant across observations. As noted by Santos Silva and

Tenreyro, the gravity model has traditionally been estimated after taking logarithms, which allows estimation using linear regression techniques. However, econometric problems are likely unless $\varepsilon_i = \exp(x_i\beta) \cdot v_i$ where v_i is distributed independently of x_i with a zero mean and constant variance.

The use of the logarithmic transformation for the dependent variable creates an immediate difficulty when trade is zero, since the log of zero is undefined. The most common response to this problem is to truncate the sample by deleting the observations with zero trade. This is, in principle, inefficient, since it ignores the information in the limit observations. 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, which automatically allows 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 special features of the resulting model may lead to more 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 following decision tree:

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



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.

If the parametric approach is taken, the first decision required is whether to adopt a Tobit/Heckman 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 trade and how much to trade are taken by individual firms based on the profitability of trade in their products. 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. If predictions of actual trade levels conditional on positive trade are required, they can still be generated using the Tobit/Heckman approach.

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), since earlier studies included insufficient variation in the exogenous variables to mitigate the multicollinearity between these variables and the additional variables in the Heckman (1979) sample selection model.² 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

² 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.

multicollinearity problems than the other standard estimators. Clearly, these results suggest that the consistency of the data with the assumptions of the Heckman/Tobit 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 Santos Silva and Tenreyro (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 Santos Silva and Tenreyro's recommendation of the PPML estimator was that it is more robust to heteroscedasticity than one based on the normal distribution even 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); and the PPML and Gamma Pseudo-Maximum Likelihood (GPML) estimators discussed by Santos Silva and Tenreyro.

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)). The ingenious Heckman two-step estimator involves

adding a variable that adjusts for the probability of sample selection, and hence overcomes the omitted variable bias to which the model is subject without this addition. One concern is that this additional variable³ 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 Heckman/Tobit 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) found this estimator with some excluded variables in the behavioral equation outperformed other estimators for limited-dependent variables.

IV. Monte Carlo Simulations

In this section of the paper, we extend the procedures used by Santos Silva and Tenreyro (2006) for the case without limited-dependent variables to cases where zero observations of the dependent variable occur with frequencies similar to those observed in real-world data. We also follow the approach to dealing with heteroscedasticity taken by Santos Silva and Tenreyro and by Westerlund and Wilhelmsson that is, we posit a range of different types of heteroscedasticity, and test these implications of these for the performance of different estimators.

We adopted the Santos Silva and Tenreyro (2006, p647) specification in equations (14) and (15) of their paper.

³ Which is the inverse Mills ratio for the particular observation (Heckman 1979).

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

where x_{1i} is a standard normal variable designed to mimic the behavior of continuous explanatory variables such as distance or income levels; x_{2i} is a binary dummy designed to mimic variables such as border dummies that equals 1 with probability 0.4 and the data are randomly generated using $\beta_0=0$, $\beta_1= \beta_2 =1$. Following Santos Silva and Tenreyro, 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 Tenreyro'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 the equation is estimated in the levels; Case 3 is homoscedastic for estimation in logarithms; Case 2 is an intermediate case; 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 incorporate the true zero estimates, 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. This approach is the data-generating process underlying the Eaton and Tamura (1994) estimator and has the interpretation of introducing a threshold level of potential trade that must be exceeded before trade actually occurs. It differs fundamentally from the rounding approach used by

Santos Silva and Tenreyro and the approaches of setting observations to zero randomly or for particular groups used by Martinez-Zarzoso, Nowak-Lehmann and Vollmer (2007). It also differs from the data generating process for a model with different explanatory variables in the selection and behavioral models—a model that we investigate later in the paper. Our data generating process for these initial simulations was:

$$(7) \quad y_i^0 = \exp(x_i\beta) - k = \exp(\beta_0 + \beta_1x_{1i} + \beta_2x_{2i}).\eta_i - k$$

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

Within our sample, 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. A k value of 1.5 generated higher shares of zeros and a substantial increase in the mean trade level although the share of zero trade levels still falls somewhat below the 70 percent observed by Brenton (personal communication) in an analysis of bilateral trade at the 5-digit level of the SITC. Table 1 shows two measures of the extent of censoring in the sample generated using equation (4). The analysis was performed in Stata 9.2, using double precision to minimize numerical errors and we followed Santos Silva and Tenreyro in using samples of 1000 observations, replicated 10,000 times. .

Our first estimation task was to replicate the simulations of Santos Silva and Tenreyro to ensure that our approach gave the right results for a sample without censoring. The results of this replication are presented in Appendix Table 1. While our results are not exactly the same as Santos Silva and Tenreyro's because of the stochastic nature of the analysis, they are completely consistent.

With this validation accomplished, we began with a semi-parametric approach, the LAD model. We then turned to the traditional single-equation models that do not

explicitly allow for estimation of the limited-dependent nature of the data-generating process. Next, we considered single-equation models designed for situations with censored data. Finally, we examined sample-selection models where the selection equation includes variables that are excluded from the equation determining trade volumes.

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 applied to the estimation of gravity models, Paarsch (1984) found the censored version of this model to give satisfactory results in large samples with censored data. Because the version of this estimator in Stata requires the model to be linear, 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 2.

From the results in Table 2, 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 even in samples of 1000. The Censored LAD estimator performed extremely badly in cases (1) to (3), with bias typically in the order of 70 percent. Given this poor

performance, even with a censored data-generating-process, we did not pursue this estimator further.

Standard Single-Equation Estimators

In this section, we considered: (i) the traditional truncated OLS in logs regression, (ii) its censored counterpart with 0.1 added to the log of output, (iii) Truncated NLS in levels, (iv) Censored NLS, (v) a Gaussian Pseudo-Maximum Likelihood estimator, (vi) a Poisson Pseudo-Maximum Likelihood estimator, and (vii) Truncated Pseudo-Maximum Likelihood estimator. Results for $k=1$ and $k=1.5$ are presented in Tables 3 and 4 respectively.

An important feature of the results for the truncated OLS-in-logarithms model is its apparently strong sensitivity to the heteroscedasticity problems emphasized by Santos Silva and Tenreyro. In Case 3, where the error distribution is consistent with the log-linear model, this model produces estimates with very small bias for $k = 1.5$. Where $k = 1.0$, the bias is around 5 percent for both coefficients. However, when we move to cases, involving heteroscedastic errors in the log-linear equation, the bias changes markedly. Where $k = 1$, the estimated bias rises to around 20 percent in cases 1 and 2, and -20 percent in case 4. The response of the bias to changes in the heteroscedasticity reflects one of Santos Silva and Tenreyro's key findings. For this estimator, unlike many others, the biases in the coefficients are generally similar for the normally distributed explanatory variable, x_1 , and the dummy variable, x_2 .

The censored OLS model estimated in logarithms (with 0.1 added to overcome the log-of-zero problem) produces results that are almost always inferior to those obtained from the truncated OLS estimator discussed above. Except in Case 4, the biases were larger in absolute value, although the estimated standard errors were somewhat

smaller. The biases were also less consistent between x_1 and x_2 than was the case with the truncated OLS. This result confirms our prediction from Figure 1 that results obtained from a censored model would likely be inferior to those from a truncated model.

Truncated NLS is the levels counterpart to the traditional estimator—truncated OLS in logs. The NLS estimator has lower bias than its logarithmic counterpart only in case 1, and is distinctly inferior in all other cases. In Case 3, the bias of the NLS estimator for $k=1$ is 40 percent for β_1 , nearly eight times the bias of truncated OLS. 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. In most cases, the bias of the censored regression is between 25 and 30 percent higher than for the corresponding truncated model. The superiority of the truncated OLS and NLS models over their censored regression counterparts suggests that just solving the “zero problem” and adding the zero valued observations to the sample is quite an unhelpful strategy.

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. In this case, the bias in the estimate of β_1 was 0.25 for $k = 1$ and 0.36 for $k = 1.5$. For β_2 the corresponding biases were 0.28 and 0.4. In most other cases, the same pattern prevailed, with biases that were large, and higher with a greater degree of censoring. Consistent with Santos Silva and Tenreiro’s findings, the bias with this estimator appears to be much less affected by heteroscedasticity than other estimators. Our results, however, suggest this advantage needs to be weighed in the gravity model context against its apparently greater vulnerability to the sample selection bias associated with limited-dependent variables.

A feature of our results for the standard estimators—and one which recurs throughout our findings and is consistent with the findings of Santos Silva and Tenreyro (2006)—is the wide variation in the size and direction of the bias in parameter estimates. In contrast with the findings of Greene (2001) for a linear model, the bias in the estimate of the slope coefficient resulting from use of the truncated or censored estimators is not consistently negative, and nor does it appear to be consistently related to the sample proportion of non-limit observations. Clearly, these results mean that much more caution is needed in interpreting results than is the case with linear models.

Single Equation Limited-Dependent Variable Estimators

Next, we turned to some of the single-equation models designed specifically for estimation with limited-dependent variables. First, we estimated the Eaton-Tamura model with the dependent variable in levels. Next, we turned to this model estimated with the dependent variable in logarithms, as originally proposed by Eaton and Tamura (1994). Then, we turned to the logical counterpart to the Poisson model advocated by Santos Silva and Tenreyro, a Tobit-type pseudo-maximum likelihood estimator with the residuals specified to be distributed Poisson. This is essentially the Tobit-Poisson model of Terza (1985). To program the likelihood function in Stata, we needed to replace the factorial function with $\exp(\ln\Gamma(y+1))$ to allow evaluation with non-integer values of the dependent variable.

The last two estimators in Tables 4 and 6 are the Heckman (1979) estimators—in both two-stage and maximum likelihood versions (see Amemiya (1984) for the derivation of the likelihood function). While—in contrast with two-stage least squares for simultaneous models—exclusion of a first stage regression variable from the second stage is not necessary for identification, the presence of such a variable may help mitigate

the frequently serious problems of multicollinearity in the second-stage equation. Our initial examination of the Heckman estimators assesses their performance in the specific Tobit case where the regressors and the error terms in the behavioral and sample selection equations are the same and exclusion of a first-stage variable would be inappropriate.

Key results for cases with $k = 1$ and $k = 1.5$ are presented in Tables 5 and 6. Since the broad pattern of results is similar, we discuss them together, except where there are important differences.

The Eaton-Tamura Tobit estimator with the dependent variable in levels has quite low bias—around three or four percent—relative to other estimators in Case 1. The same model estimated in logarithms produces quite good estimates—around six percent bias for β_1 with $k=1$ and 1.3 percent for $k=1.5$ —in Case 3, when the underlying error structure is consistent with its assumptions. Importantly, however, the bias of this estimator increases sharply as the residuals become heteroscedastic relative to the assumed functional form. In Case 1 the results from the log-linear model are biased downwards by 25 percent for both coefficients, at both depths of censoring. The E-T Tobit in levels had larger bias than its counterpart Truncated NLS except in case (1). This result parallels Manning, Duan and Rogers' (1987) finding that truncated OLS (Part 2 of the Two-Part model) can outperform sample-selection models even when the data are generated using a sample-selection process.

The Poisson-Tobit estimator in levels had very substantial bias in almost every case. Even in Case 1, where the error structure is consistent with the levels specification, the bias was around 25 percent for both β_1 and β_2 with $k=1$ and 35 and 40 percent with $k=1.5$. Consistent with Santos Silva and Tenreiro's findings, the extent of the bias does not appear to be sensitive to the properties of the error term. In Case 3, the extent of the

bias with this estimator is in the same size range for both $k=1$ and 1.5. Comparing the Poisson-Tobit result with the PPML, we find very little difference in the bias of the corresponding pairs of coefficient estimates. Simply moving to a limited dependent variable formulation provides little benefit in this case.

The two Heckman estimators are presented only in logarithms, since only the log-linear version can be estimated in standard statistical packages, such as Stata. We attempted to estimate the linear-in-levels version by maximum likelihood, but the estimator failed to converge—a common problem with this type of estimator (Nawata 1994). The two-step estimator performed extremely badly in almost all cases, with biases of 70 to 80 percent in numerous cases. The performance of these estimators was better in case (3), where heteroscedasticity should not have been a problem, than in most other cases. But even in this case, both of the Heckman-type estimators were outperformed by the E-T Tobit in logarithms.

All of the limited-dependent variable estimators in Tables 5 and 6 are attempting a difficult challenge—to identify the sample selection using only information on the distribution of residuals, and to estimate the other parameters. One feature of the Monte Carlo simulations presented above that makes the challenge greater is an unused restriction. While we know that the β_0 in the data generating process (equation (6)) was set to zero, this restriction has not been imposed in estimation because we are unlikely to have this information in practical applications. However, it remains of interest to know whether this information makes a difference to the results. Appendix Table 2 reports the results from Table 5 re-estimated with this restriction imposed.

This restriction noticeably reduced the bias of the estimates associated with the two NLS estimators, PPML, the E-T Tobit estimators and the Poisson-Tobit. It did not

lead to consistently lower bias, and frequently led to higher bias, with the OLS estimates, the GPML and the Heckman estimators. The PPML and Poisson-Tobit results were very similar and generally, although not always, inferior to the NLS and E-T Tobit results. In many cases, the GPML estimator performed much worse than in the absence of the restriction. The overall performance of the NLS and E-T Tobit estimators was very similar, although one or other was quite biased in some specific cases, such as case 4, where the E-T Tobit in levels had a bias of 0.33. The similarity between the results for the E-T Tobit estimator and the NLS estimators is surprising given that the E-T Tobit estimator is based on precisely the model used to generate the data. For these estimators, these results echo the finding by Manning, Duan and Rogers (1987) of superior performance by a standard estimator over one designed to deal with sample selection even when the sample selection model was the true model.

One potential approach to improving the performance of the E-T Tobit model might be to adjust for heteroscedasticity. We did this using 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, and γ and δ are parameters estimated together with the behavioral parameters of interest using nonlinear least squares. For the linear-in-levels model, the specified error process is $\sigma_i^2 = (\gamma + \delta \exp(x_i \beta))^2$. The results from this estimation process are presented in Table 7.

The E-T Tobit model in logarithms performs much better in Cases 1 to 3 than the E-T Tobit model without correction for heteroscedasticity. However, in Case 4, where the form of heteroscedasticity considered is not nested within the error specification, the performance is not worse than in Table 5. A disturbing feature of these results, even in cases (1) to (3), is the very high standard errors associated with the use of this estimator.

These standard errors do decline with sample size. Using a sample of 10,000, which is more consistent with contemporary cross-sectional analyses, these errors generally fall by approximately 30%. However, they suggest a need for caution in using this estimator unless the sample is very large.

The estimates obtained using the linear-in-levels model are generally much less satisfactory than those for the model in logs. Only in case (1) is the bias of this estimator reasonably small. In cases (2) through (4) both the bias and the standard error of these estimates are large. An indication of the problem with this estimator is provided by the estimates of β_0 . These show very large bias relative to the true value of zero for this parameter. The performance of this estimator improves enormously if β_0 is restricted to its true value of zero. In all cases except the estimate on the dummy variable in case (4), the bias is relatively small. Unfortunately, it seems unlikely that the information needed to impose this restriction with confidence would be available in real-world applications.

Clearly, the formulations used to deal with the heteroscedasticity problem in this section are completely consistent with the data-generating process in cases (1) to (3), but do not fully represent this process in Case (4). Case (4) could be captured relatively easily in the log-linear case using a more general representation of heteroscedasticity, such as $\sigma_i^2 = (\gamma + \delta_1 x_i + \delta_2 x_2)^2$. This would involve estimation of only one additional parameter in this simulation, while reducing the nonlinearity of the problem. However, it would likely involve many more additional parameters in actual empirical studies, which frequently include 10 or more explanatory variables. In the linear-in-variables case, it is less obvious how this approach might be generalized.

Sample Selection Models with Exclusion Restrictions

In light of the poor performance of all of the single equation models, we examined sample selection models based on Heckman (1979) such as those used by Francois and Manchin (2007); Lederman and Ozden (2007) and Helpman, Melitz and Rubeinstein (2007). In contrast with the single-equation version of the Heckman model presented in Tables 5 and 6, we assume in this section 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 a sample selection equation:

$$(9) \quad y_{2i} = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + u_{2i}$$

which determines whether or not y_{1i} is included in the sample for estimation. Variables x_{1i} and x_{2i} are included in both equation (9) and (6) with coefficient values of unity, but their interpretation is, of course, different since equation (9) is linear. The error term in (9) is $u_{2i} \sim N(0,1)$ while the covariance between u_{2i} and η_{1i} is 0.5. In addition, we include an additional, independently-distributed variable x_{3i} which is a dummy variable⁴ with probability 0.5 of being equal to one. If the realization of equation (9) exceeds its threshold value, that observation takes a nonzero value in equation (6). Otherwise y_{1i} is zero. The thresholds are set at the 30th and 50th percentiles of the distribution of y_{2i} , so that roughly 30% and 50% of trade outcomes are zero.

Using datasets of 1000 observations, we apply the Heckman estimation procedure in logarithms 10,000 times to assess its performance. The results of this estimation for the behavioral equation are presented in Table 8. A striking feature of the results of this analysis is the dramatic improvement in the results relative to earlier simulations, both when estimated in logs and in levels. Even for the two-step estimator in logs, the degree

⁴ The dummy variable is used as the excluded variable since the excluded variable is typically a dummy variable (see Helpman, Melitz and Rubinstein 2007).

of bias is an order of magnitude lower than in the earlier experiments. For the maximum likelihood estimator in logs, the reduction in bias is extraordinarily large, going from the highest amongst the standard estimators in the single equation case to the lowest by far in this experiment. The Heckman ML estimator produces estimates with very little bias in seven out of eight cases, with the exception being the estimate of β_2 in case (4).

Estimation of the Heckman model in levels by the two-step procedure is straightforward, except that a degree of ambiguity is introduced by the linear functional form of equation (9) required to allow its estimation using standard Probit estimators. Estimation of this model by maximum likelihood in Stata presented challenges. The maximum likelihood estimator was solved eventually with the derivatives provided analytically, although we are concerned about whether the results represent a global maximum. Interestingly, when estimating in levels, the two-step estimator had smaller bias than the maximum likelihood estimator in almost all cases, and generally much lower standard errors.

The apparent success of the Heckman Maximum Likelihood estimators in logs is particularly striking given that they make no explicit adjustment for the heteroscedasticity problem. The fact that they are estimated in logs means that they are subject to the Jensen's inequality problem emphasized by Santos Silva and Tenreyro. Further, there is some degree of bias due to the replacement of zero values of trade by unitary ones that yield a log of zero. However, our result is consistent with the findings by Manning, Duan and Rogers (1987) and by Leung and Yu (1996) that the sample selection model performs better than the two-part model (in this case the comparator is the truncated logarithmic OLS estimator) when there are exclusion restrictions.

Clearly, these results have potentially important implications for applied work. A key question is whether we have data on truly independent variables that belong in the selection equation but not in the trade value equation. Given the derivation of such models by Helpman, Melitz and Rubinstein (2007) or Baldwin and Harrigan (2007), variables associated with the fixed costs of establishing trade flows would appear to qualify. Variables such as the common-religion dummy used by Helpman, Melitz and Rubinstein (2007); common-language dummies; or the “Doing Business” indicators on the costs of starting a business seem plausible as indicators of fixed, rather than variable, costs of exporting. The exclusion of each country’s GDP from the trade flow equation in Hallak (2006) seems less appealing from this perspective.

V. Empirical Implementation

To investigate the performance of our preferred estimators relative to alternatives such as the traditional truncated OLS model, and the E-T Tobit estimator, and PPML we used the Santos Silva and Tenreyro (2006, p649) dataset kindly provided by the authors—a cross-sectional dataset of 18360 observations for 136 exporters and 135 importers . We considered both traditional (including GDP’s as explanatory variables) and the Anderson-van Wincoop (using country fixed effects) specifications. The results for this single analysis are presented in Tables 9 and 10.

The results in Table 9 are very informative for the insights they provide into the coefficients on GDP and the log of distance. Our preferred estimators, the Heckman ML and the E-T Tobit with correction for autocorrelation, suggest that the coefficient on exporter GDP is close to unity. This finding seems to be generally robust to the choice of exogenous variables to exclude in the second stage. Even though the Heckman-ML estimator has a smaller coefficient than the E-T Tobit, the difference between the

Heckman-ML coefficient and unity (0.02) is economically unimportant. By contrast, the PPML estimate is 0.711, substantially below one. For all estimators, the variable for importer's GDP has a smaller coefficient (between 0.8 and 0.9 with our preferred specifications) than the unitary coefficient suggested by Anderson and van Wincoop (2003). The results also point to a substantial difference in the estimated effects of distance between the PPML estimator and our preferred estimator, with our preferred estimators yielding estimates around 1.2 while the PPML estimator is 0.76. If we accept these results as indicating the possibility that the PPML estimator is biased, then one surprising feature is that the PPML estimates are biased down relative to the other estimators, even though our Monte Carlo analysis found upward bias in the estimates on both continuous and dummy variables in all of the four cases we considered.

A key feature of Table 10 is that the Anderson-van Wincoop (2003) formulation yields a similar, but larger, difference between the estimated impacts of distance in the PPML estimator and other estimators. In this case, our preferred estimators yield estimates of -1.38 or -1.29 while the PPML estimator yields an estimate of -0.75. Again, the surprising feature of the PPML estimate is that it appears to be biased downwards.

VI. Conclusions

The purpose of this paper is to follow up on the challenge laid down by Santos Silva and Tenreyro—to consider estimation of the gravity model in situations where zero trade flows are prevalent, particularly when trade is considered at a disaggregate level. In doing this, we build on Santos Silva and Tenreyro (2006) heteroscedasticity is a potentially major source of bias in traditional log-dependent estimating models of the gravity equation.

Acknowledging recent theoretical developments that suggest the reason for many cases of zero trade is failure to meet the fixed costs of establishing such trade, we devise a strategy for choosing an estimator. With this as background, we identify a number of potential estimators and investigate whether they lead to unbiased parameter estimates in the presence of these econometric problems. We use Monte Carlo simulations based on the design of Santos Silva and Tenreyro, modified to include a threshold level of trade that must be surmounted before positive trade levels are observed. The threshold was set to generate frequencies of zero trade similar to that observed in studies of aggregate trade flows.

The Monte-Carlo simulations confirm the importance of heteroscedasticity as a source of bias with a number of estimators, and the lesser susceptibility of the Poisson pseudo maximum likelihood (PPML) estimator to this problem. However, their recommended PPML estimator is found to be strongly susceptible to limited-dependent variable bias when a substantial fraction of the observations are censored. This problem does not appear to be solved with a Tobit-type censoring regression based on the Poisson distribution. The bias in the resulting estimator is generally similar to that of the PPML estimator, and frequently around 25 percent. While the resulting bias with this estimator is apparently not greatly influenced by the pattern of heteroscedasticity, it remains large across all of the forms of heteroscedasticity considered.

The smallest biases amongst the single-equation models were found with Eaton-Tamura Tobit estimators, but only when the functional form is consistent with the form of heteroscedasticity, or an appropriate correction is made to deal with heteroscedasticity. With errors consistent with estimation in levels, the bias of the Eaton-Tamura model was around 3 or 4 percent, irrespective of the fraction of the sample censored. With errors

consistent with log-linear estimation, the E-T model in logarithms also had the lowest bias. These estimators were, however, very vulnerable to deviations from the assumed distribution of the residuals. The E-T Tobit in logs, for instance, was biased downwards by about 25 percent when the underlying data were consistent with estimation in levels.

Both the truncated and the censored Least Absolute Deviation estimators gave very poor results, with very large bias, suggesting that semi-parametric estimation may not provide a solution to the combined problems of heteroscedasticity and sample selection plaguing estimation of the gravity model. The censored model generally produced much worse results than the truncated model, even though the data were generated using a censoring process.

The Heckman sample-selection estimators—whether in two-step or maximum likelihood—gave very poor results when estimated for a single equation with the same variables in the selection and estimation equations. However, when the data-generating process for the Monte Carlo simulations was modified to include an additional variable in the determination of bilateral flows with positive trade, the performance of these estimators improved dramatically, and yielded a combination of small bias and small standard errors in seven out of eight cases even without directly addressing the heteroscedasticity problem. Fortunately, the new trade theories that attempt to explain zero trade flows through firm heterogeneity suggest that there are some variables—those related to the fixed costs of establishing trade flows—that are appropriately excluded from the equation for the level of trade. The maximum likelihood estimator substantially outperformed the traditionally-favored two-step instrumental variable estimator, suggesting that a move to this approach to estimation would be desirable.

Finally, our empirical application to the Santos Silva and Tenreyro dataset found that the PPML estimator reconfirmed their assessment that the PPML estimator yielded smaller estimates of GDP and distance effects than other estimators. Our preferred estimators yielded estimates much closer to findings from traditional models such as truncated OLS than those of the PPML estimator. Given the difficulties encountered by the PPML estimator in dealing with data generated from datasets with zero trade levels, this difference would seem more likely due to the vulnerability of the PPML estimator than of received theory and empirical evidence.

Reference List

- 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.
- Arabmazar,A. and P.Schmidt "Further evidence on the robustness of the Tobit estimator to heteroscedasticity." *Journal of Econometrics* 17(1981): 258.
- Baldwin,R. and J.Harrigan. Zeros, quality and space: trade theory and trade evidence. [NBER Working Paper 13214]. 2007. Cambridge, MA, National Bureau of Economic Research.
- 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).
- Chay,K.Y. and J.L.Powell "Semiparametric censored regression models." *Journal of Economic Perspectives* 15(2001): 29-42.
- Duan,N. et al. "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.
- Greene,W. "On the asymptotic bias of the Ordinary Least Squares Estimator of the Tobit model." *Econometrica* 49(1981): 505-13.
- Hallak,J.C. "Product quality and the direction of trade." *Journal of International Economics* 68(2006): 238-65.
- Hebble, M., Shepherd, B., and Wilson, J. S. 2007. "Trade Costs and International Production Networks: Lessons from the Asia-Pacific Experience." Presented at Paper presented to the Conference of the European Trade Study Group, Athens,
- 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." Unpublished, 2007.

- Hurd,M. "Estimation in truncated samples when there is heteroscedasticity." *Journal of Econometrics* 11(1979): 247-58.
- Jones,A. "Health econometrics." *Handbook of Health Economics, Volume 1*. A.Culyer and J.Newhouse, eds., 267-344. Amsterdam: Elsevier Science, 2000.
- 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.
- Martínez-Zarzoso, Nowak-Lehmann, Felicitas, and Vollmer, Sebastian. 2007. "The log of gravity revisited." Presented at European Trade Study Group Annual Conference , Athens, 9-13-2007.
- Nawata,K. "Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator." *Economics Letters* 45(1994): 33-40.
- Paarsch,H.J. "A Monte-Carlo comparison of estimators for censored regression models." *Journal of Econometrics* 24(1984).
- 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 S.Tenreiro "The log of gravity." *The Review of Economics and Statistics* 88(2006): 641-58.
- 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." Unpublished, 2007.
- Xuepeng Liu "GATT/WTO promotes trade strongly: sample selection and model specification." Unpublished, 2007.

Table 1. Indicators of the degree of censoring for different intercept values
Percentage of Zero Trade Values

	$k = 1.0$	$k = 1.5$
	%	%
Case 1	41	51
Case 2	44	54
Case 3	49	60
Case 4	55	64
Percentage change in Mean		
Case 1	15	38
Case 2	15	40
Case 3	16	43
Case 4	19	50

Table 2. Monte Carlo results from Least-Absolute-Deviations Estimators

Estimator	Dependent Variable	β_1		β_2	
	Form	Bias	Std Error.	Bias	Std Error.
<i>(k=1.0)</i>					
Case 1: $V[y_i x]=1$					
Truncated LAD	Log	0.2189	0.0331	0.2284	0.0480
Censored LAD	Log	0.4906	0.0496	0.5069	0.0672
Case 2: $V[y_i x]=\exp(x_i\beta)$					
Truncated LAD	Log	0.1898	0.0447	0.1946	0.0791
Censored LAD	Log	0.5307	0.0237	0.5606	0.0354
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
Truncated LAD	Log	0.0291	0.0754	0.0277	0.1271
Censored LAD	Log	0.5145	0.0424	0.5311	0.0627
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
Truncated LAD	Log	-0.1915	0.0917	-0.1836	0.1547
Censored LAD	Log	0.5749	0.1856	0.1249	0.2450
<i>(k=1.5)</i>					
Case 1: $V[y_i x]=1$					
Truncated LAD	Log	0.2825	0.0389	0.2906	0.0558
Censored LAD	Log	0.6703	0.0204	0.6899	0.0284
Case 2: $V[y_i x]=\exp(x_i\beta)$					
Truncated LAD	Log	0.2106	0.0527	0.2106	0.0922
Censored LAD	Log	0.7111	0.0271	0.7343	0.0434
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
Truncated LAD	Log	-0.0169	0.0901	-0.0127	0.1505
Censored LAD	Log	0.7032	0.0529	0.7152	0.0855
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
Truncated LAD	Log	-0.2501	0.1067	-0.1856	0.1773
Censored LAD	Log	0.0178	0.1783	0.0139	0.1389

Table 3. Monte Carlo results with standard estimators, ($k=1.0$)

Estimator	Dependent Variable	β_1		β_2	
	Form	Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
Truncated OLS	Log	0.2045	0.0589	0.2198	0.0845
OLS (ln(y+0.1))	Log	0.2682	0.0287	0.3541	0.0576
Truncated NLS	Level	0.0980	0.0216	0.1101	0.0269
Censored NLS	Level	0.1382	0.0341	0.1693	0.0358
GPML	Level	0.4236	0.2064	0.4164	0.2356
PPML	Level	0.2505	0.0356	0.2800	0.0447
Truncated PPML	Level	0.0964	0.0282	0.0984	0.0409
Case 2: $V[y_i x]=\exp(x_i\beta)$					
Truncated OLS	Log	0.1903	0.0578	0.1971	0.0929
OLS (ln(y+0.1))	Log	0.2327	0.0302	0.3154	0.0641
Truncated NLS	Level	0.0888	0.0434	0.0985	0.0739
Censored NLS	Level	0.1202	0.0458	0.1434	0.0759
GPML	Level	0.5925	0.1857	0.5705	0.2154
PPML	Level	0.2580	0.0387	0.2883	0.0616
Truncated PPML	Level	0.0819	0.0304	0.0812	0.0565
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
Truncated OLS	Log	0.0585	0.0669	0.0569	0.1147
OLS (ln(y+0.1))	Log	0.0961	0.0362	0.1621	0.0772
Truncated NLS	Level	0.3958	20.4693	0.0906	2.8135
Censored NLS	Level	0.5032	24.2816	0.1895	2.1603
GPML	Level	0.9216	0.1563	0.8640	0.2055
PPML	Level	0.2550	0.0896	0.2899	0.1435
Truncated PPML	Level	0.0275	0.0282	0.0294	0.1457
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
Truncated OLS	Log	-0.1885	0.0793	-0.1579	0.1366
OLS (ln(y+0.1))	Log	0.0115	0.0412	0.1520	0.0895
Truncated NLS	Level	0.6677	25.2023	0.2360	7.1123
Censored NLS	Level	0.8191	24.4597	0.3241	6.364
GPML	Level	0.4991	0.1793	0.5581	0.2474
PPML	Level	0.1978	0.1219	0.2580	0.1886
Truncated PPML	Level	-0.1265	0.1514	0.0318	0.1964

Table 4. Monte Carlo results with standard estimators, ($k=1.5$)

Estimator	Dependent Variable	β_1		β_2	
	Form	Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
Truncated OLS	Log	0.2274	0.0721	0.2365	0.1015
OLS (ln(y+0.1))	Log	0.2300	0.0312	0.3294	0.0623
Truncated NLS	Level	0.1430	0.0311	0.1596	0.0350
Censored NLS	Level	0.2099	0.0523	0.2621	0.0529
GPML	Level	0.5317	0.2694	0.5210	0.3155
PPML	Level	0.3565	0.0489	0.4004	0.0557
Truncated PPML	Level	0.1513	0.0351	0.1527	0.0489
Case 2: $V[y_i x]=\exp(x_i\beta)$					
Truncated OLS	Log	0.1875	0.0686	0.1877	0.1095
OLS (ln(y+0.1))	Log	0.1859	0.0332	0.2784	0.0691
Truncated NLS	Level	0.1238	0.0497	0.1354	0.0829
Censored NLS	Level	0.1727	0.0554	0.2051	0.0874
GPML	Level	0.7811	0.2655	0.7487	0.2655
PPML	Level	0.2581	0.0386	0.2883	0.0616
Truncated PPML	Level	0.1108	0.0304	0.1070	0.0652
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
Truncated OLS	Log	0.0150	0.0794	0.0078	0.1355
OLS (ln(y+0.1))	Log	0.0261	0.0399	0.0975	0.0820
Truncated NLS	Level	0.4323	22.8808	0.0897	3.0700
Censored NLS	Level	0.5591	23.8062	0.2499	2.1946
GPML	Level	1.2545	0.2476	1.1793	0.3106
PPML	Level	0.3367	0.0999	0.3842	0.1643
Truncated PPML	Level	0.0092	0.1182	0.0087	0.1659
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
Truncated OLS	Log	0.2477	0.0908	-0.1608	0.1558
OLS (ln(y+0.1))	Log	0.0907	0.0442	0.1892	0.0921
Truncated NLS	Level	0.6759	27.5164	0.2413	6.8576
Censored NLS	Level	0.9026	24.7490	0.3913	6.2925
GPML	Level	0.6443	0.2393	0.7508	0.3229
PPML	Level	0.2576	0.1319	0.3513	0.2083
Truncated PPML	Level	-0.1692	0.1669	0.0344	0.2166

Table 5. Monte Carlo results with limited dependent variable estimators, ($k=1.0$)

Estimator	Dependent Variable Form	β_1		β_2	
		Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
ET-Tobit	Level	0.0335	0.025	0.0339	0.0301
ET-Tobit	Log	-0.2510	0.0375	-0.2517	0.0446
Poisson-Tobit	Level	0.2510	0.0344	0.2795	0.0441
Heckman-ML	Log	0.2840	0.0484	0.3016	0.0772
Heckman-2SLS	Log	0.5979	0.0734	0.6241	0.0981
Case 2: $V[y_i x]=\exp(x_i\beta)$					
ET-Tobit	Level	0.1478	0.0762	0.1594	0.0846
ET-Tobit	Log	-0.1509	0.0430	-0.1603	0.0554
Poisson-Tobit	Level	0.2584	0.0378	0.2876	0.0619
Heckman-ML	Log	0.2782	0.0478	0.2878	0.0869
Heckman-2SLS	Log	0.5778	0.0851	0.5963	0.1158
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
ET-Tobit	Level	0.6566	0.1818	0.6683	0.1711
ET-Tobit	Log	0.0667	0.0624	0.0621	0.0888
Poisson-Tobit	Level	0.2415	0.0785	0.2811	0.1383
Heckman-ML	Log	0.1396	0.0742	0.1399	0.119
Heckman-2SLS	Log	0.2988	0.1450	0.3026	0.1751
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
ET-Tobit	Level	0.6846	0.0844	0.7269	0.3015
ET-Tobit	Log	0.0639	0.0753	0.0081	0.1142
Poisson-Tobit	Level	0.1629	0.0917	0.2031	0.1614
Heckman-ML	Log	-0.0294	0.1080	-0.1419	0.1472
Heckman-2SLS	Log	0.3576	0.2619	0.2391	0.2312

Table 6. Monte Carlo results with limited dependent variable estimators, ($k=1.5$)

Estimator	Dependent Variable Form	β_1		β_2	
		Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
ET-Tobit	Level	0.0368	0.0303	0.0414	0.0348
ET-Tobit	Log	-0.2401	0.0435	-0.2494	0.0507
Poisson-Tobit	Level	0.3570	0.0471	0.3998	0.0549
Heckman-ML	Log	0.3446	0.0568	0.3584	0.0906
Heckman-2SLS	Log	0.7749	0.096	0.8052	0.1242
Case 2: $V[y_i x]=\exp(x_i\beta)$					
ET-Tobit	Level	0.1847	0.0836	0.1980	0.0889
ET-Tobit	Log	-0.1884	0.0494	-0.1908	0.0613
Poisson-Tobit	Level	0.3562	0.0492	0.3983	0.0739
Heckman-ML	Log	0.3109	0.0548	0.3158	0.1012
Heckman-2SLS	Log	0.7081	0.1094	0.7274	0.1438
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
ET-Tobit	Level	0.5731	0.3302	0.6209	0.2696
ET-Tobit	Log	0.0131	0.1628	0.0741	0.077
Poisson-Tobit	Level	0.3279	0.0881	0.3752	0.1591
Heckman-ML	Log	0.1255	0.1041	0.1211	0.1518
Heckman-2SLS	Log	0.3477	0.1999	0.3492	0.2336
Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}\beta)(\exp(x_i\beta))^2$					
ET-Tobit	Level	0.7267	0.0919	0.7688	0.0894
ET-Tobit	Log	0.0809	0.0915	0.0525	0.1304
Poisson-Tobit	Level	0.2047	0.0880	0.2271	0.1239
Heckman-ML	Log	-0.0739	0.1779	-0.1264	0.1939
Heckman-2SLS	Log	0.4101	0.3653	0.3465	0.3172

Table 7: Monte Carlo Results: Eaton-Tamura Tobit Adjusted for Heteroscedasticity, (k=1)

	β_0		β_1		β_2	
	Bias	Std. Error.	Bias	Std. Error.	Bias	Std. Error.
Logs						
Case 1	0.0218	0.3373	-0.0267	0.1627	-0.0567	0.2085
Case 2	-0.0625	0.4037	-0.041	0.2378	-0.1221	0.3498
Case 3	-0.3332	0.4974	-0.0082	0.2644	-0.0656	0.3593
Case 4	-0.2574	0.4185	-0.1816	0.3898	-0.2748	0.3996
Levels						
Case 1	0.0832	0.3797	-0.0187	0.1202	-0.0211	0.1201
Case 2	0.5125	0.2509	-0.1656	0.1006	-0.1773	0.1198
Case 3	0.7819	0.6879	-0.2625	0.2654	-0.2713	0.3105
Case 4	1.3579	0.9614	-0.4723	0.317	-0.3661	0.487
Levels, $\beta_0 = 0$						
Case 1	-	-	-0.0051	0.1589	-0.0383	0.6146
Case 2	-	-	0.0073	0.0145	0.0371	0.0323
Case 3	-	-	0.0365	0.0470	0.0584	0.0780
Case 4	-	-	-0.0236	0.0648	0.3465	0.0836

Table 8: Monte Carlo Results: Heckman Model with exclusion restriction

Estimator	Approximate % Zero Values	(Samples of 1000)			
		β_1		β_2	
		Bias	Std. Error.	Bias	Std. Error.
Heckman-2SLS-Log					
Case 1	30%	0.1281	0.0420	0.1344	0.0554
	50%	0.0802	0.0121	0.0853	0.059
Case 2	30%	0.1157	0.0433	0.1203	0.0597
	50%	0.0855	0.0499	0.0892	0.0673
Case 3	30%	-0.00002	0.0107	-0.00009	0.0151
	50%	-0.000006	0.0136	-0.00006	0.0184
Case 4	30%	0.0538	0.0122	-0.1963	0.0144
	50%	0.0449	0.0139	-0.2082	0.0156
Heckman-Maximum Likelihood-Log					
Case 1	30%	0.0345	0.0303	0.0408	0.0454
	50%	-0.0136	0.0319	-0.0102	0.0468
Case 2	30%	0.0310	0.0297	0.0354	0.0495
	50%	0.00057	0.0334	0.0027	0.0543
Case 3	30%	-0.00009	0.0107	-0.00016	0.0151
	50%	-0.00006	0.0137	-0.00012	0.0184
Case 4	30%	0.0612	0.0104	-0.1889	0.0132
	50%	0.0467	0.0132	-0.2061	0.0151
Heckman-2SLS-Levels					
Case 1	30%	-0.0045	0.0216	-0.0036	0.0274
	50%	-0.0065	0.0187	-0.0053	0.0252
Case 2	30%	-0.0001	0.0739	0.0013	0.0915
	50%	-0.0033	0.0732	-0.0011	0.0898
Case 3	30%	0.0043	0.0799	0.0033	0.0868
	50%	0.0046	0.0803	0.0035	0.0864
Case 4	30%	-0.0030	0.0731	-0.2842	0.0606
	50%	-0.0020	0.0732	-0.2831	0.0599
Heckman-Maximum Likelihood-Levels					
Case 1	30%	0.0192	0.0600	0.0033	0.1207
	50%	0.0237	0.0624	0.0086	0.0573
Case 2	30%	-0.0260	0.6658	0.0337	0.1470
	50%	0.0746	0.2576	0.1300	0.6889
Case 3	30%	0.0098	0.1621	-0.0252	0.1527
	50%	-0.0106	0.3015	0.0641	0.4603
Case 4	30%	0.0272	0.1015	-0.3042	0.1462
	50%	-0.2533	2.6453	-0.1309	1.2949

Table 9: Traditional Gravity Equation

Independent Variables	Truncated OLS	PPML	Heckman 2SLS	Heckman ML	Heckman 2SLS	Heckman ML	Heckman 2SLS	Heckman ML	ET-Tobit
Log exporter's GDP	0.938 ** (0.012)	0.711 ** (0.027)	1.023 ** (0.017)	0.964 ** (0.014)	1.030 ** (0.020)	0.976 ** (0.016)	1.044 ** (0.020)	0.981 ** (0.016)	1.076 ** (0.011)
Log importer's GDP	0.798 ** (0.012)	0.720 ** (0.025)	0.854 ** (0.014)	0.813 ** (0.012)	0.893 ** (0.016)	0.856 ** (0.014)	0.902 ** (0.016)	0.859 ** (0.014)	0.902 ** (0.010)
Log exporter's GDPC	0.200 ** (0.017)	0.191 ** (0.056)	0.224 ** (0.017)	0.214 ** (0.017)	0.208 ** (0.019)	0.203 ** (0.019)	0.209 ** (0.019)	0.203 ** (0.019)	0.212 ** (0.016)
Log importer's GDPC	0.099 ** (0.018)	0.170 ** (0.045)	0.128 ** (0.017)	0.116 ** (0.017)	0.081 ** (0.019)	0.072 ** (0.019)	0.083 ** (0.019)	0.074 ** (0.019)	0.139 ** (0.015)
Log distance	-1.172 ** (0.034)	-0.756 ** (0.060)	-1.308 ** (0.036)	-1.255 ** (0.034)	-1.263 ** (0.038)	-1.232 ** (0.038)	-1.272 ** (0.038)	-1.235 ** (0.038)	-1.223 ** (0.031)
Contiguity	0.317 * (0.127)	0.170 (0.100)	0.191 (0.146)	0.272 * (0.143)	0.171 (0.155)	0.217 (0.154)	0.158 (0.156)	0.213 (0.154)	-0.304 * (0.115)
Common language	0.670 ** (0.067)	0.751 ** (0.129)			0.817 ** (0.071)	0.770 ** (0.070)	0.829 ** (0.071)	0.774 ** (0.070)	0.807 ** (0.061)
Colonial tie	0.389 ** (0.070)	0.013 (0.145)	0.902 ** (0.056)	0.850 ** (0.055)	0.363 ** (0.076)	0.351 ** (0.075)	0.365 ** (0.076)	0.352 ** (0.075)	0.391 ** (0.064)
Landlocked_exporter	-0.063 (0.062)	-0.888 ** (0.150)	-0.049 (0.065)	-0.051 (0.065)	-0.062 (0.068)	-0.067 (0.068)	-0.063 (0.068)	-0.067 (0.068)	-0.271 ** (0.057)
Landlocked_importer	-0.665 ** (0.060)	-0.721 ** (0.133)	-0.676 ** (0.064)	-0.667 ** (0.064)	-0.706 ** (0.067)	-0.695 ** (0.066)	-0.710 ** (0.067)	-0.695 ** (0.067)	-0.714 ** (0.053)
Exporter's remoteness	0.482 ** (0.08)	0.353 ** (0.127)	0.653 ** (0.078)	0.616 ** (0.080)	0.552 ** (0.084)	0.540 ** (0.083)	0.556 ** (0.084)	0.541 ** (0.084)	0.444 ** (0.070)
Importer's remoteness	-0.189 * (0.085)	0.249 ** (0.121)	-0.055 (0.081)	-0.069 (0.080)	-0.141 (0.087)	-0.135 (0.086)	-0.144 (0.088)	-0.136 (0.087)	-0.055 (0.076)

Table 9: Traditional Gravity Equation

Independent Variables	Truncated OLS	PPML	Heckman 2SLS	Heckman ML	Heckman 2SLS	Heckman ML	Heckman 2SLS	Heckman ML	ET-Tobit
FTA	0.487 ** (0.098)	0.120 (0.086)	0.412 ** (0.107)	0.438 ** (0.106)	0.442 ** (0.121)	0.458 ** (0.120)	0.438 ** (0.122)	0.457 ** (0.120)	-0.189 * (0.075)
Openess	-0.106 * (0.050)	-0.447 ** (0.083)	-0.031 (0.049)	-0.094 * (0.048)	-0.018 (0.051)	-0.068 (0.050)	-0.005 (0.051)	-0.064 (0.050)	0.089 * (0.045)
Number of observations	9613	18360	15500	15500	15500	15500	15500	15500	18360
Excluded variable:	No	No							
Common language			Yes	Yes					
Starting-business procedures					Yes	Yes			
Starting-business time							Yes	Yes	
Heteroscedasticity correction	No	No	No	No	No	No	No	No	Yes
γ									2.538 ** (0.041)
δ									-0.1027 ** (0.0042)

Table 10: The Anderson-van Wincoop Gravity Equation

	Truncated	PPML	Heckman	Heckman	Heckman	Heckman	Heckman	Heckman	ET-Tobit
Independent Variables	OLS		2SLS	ML	2SLS	ML	2SLS	ML	
Log distance	-1.347 ** (0.031)	-0.750 ** (0.041)	-1.383** (0.031)	-1.382 ** (0.030)	-1.383 ** (0.034)	-1.383 ** (0.034)	-1.380 ** (0.034)	-1.380 ** (0.034)	-1.290 ** (0.030)
Contiguity	0.174 (0.129)	0.370 ** (0.091)	0.184 (0.125)	0.185 (0.125)	0.151 (0.135)	0.152 (0.135)	0.155 (0.134)	0.155 (0.134)	-0.224 (0.126)
Common language	0.406 ** (0.067)	0.383 ** (0.093)			0.447 ** (0.070)	0.447 ** (0.070)	0.444 ** (0.070)	0.444 ** (0.070)	0.520 ** (0.059)
Colonial tie	0.666 ** (0.70)	0.079 (0.134)	0.917** (0.055)	0.915 ** (0.055)	0.615 ** (0.074)	0.614 ** (0.074)	0.613 ** (0.073)	0.613 ** (0.073)	0.656 ** (0.060)
FTA	0.310 ** (0.098)	0.376 ** (0.077)	0.289 ** (0.099)	0.291 ** (0.098)	0.283 ** (0.112)	0.283 ** (0.112)	0.288 ** (0.111)	0.288 ** (0.111)	-0.344 ** (0.095)
Exporter fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	9613	18360	18360	18360	15500	15500	15500	15500	18360
Excluded variable:	No	No							
Common language			Yes	Yes					
Start-business procedure					Yes	Yes			
Start-business time							Yes	Yes	
Heteroscedasticity cornn	No	No	No	No	No	No	No	No	Yes
γ									1.973 ** (0.049)
δ									-0.0684 ** (0.0053)

Appendix Table 1. Replicating Santos Silva and Tenreyro's "Log of Gravity" Results

Estimator	Dependent Variable Form	β_1		β_2	
		Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
PPML	Level	0.00021	0.016	0.0006	0.027
NLS	Level	-0.000063	0.008	0.00014	0.017
GPML	Level	0.01370	0.068	0.0086	0.083
OLS	Log	0.39001	0.390	0.35635	0.053
OLS(y>0.5)	Log	-0.16340	0.027	-0.15428	0.038
OLS(y+1)	Log	-0.40217	0.013	-0.37644	0.022
Case 2: $V[y_i x]=\mu(x_i\beta)$					
PPML	Level	-0.00006	0.019	0.00052	0.039
NLS	Level	0.00040	0.033	0.00122	0.057
GPML	Level	0.00440	0.043	0.0029	0.062
OLS	Log	0.21072	0.030	0.20032	0.048
OLS(y>0.5)	Log	-0.17817	0.026	-0.17158	0.042
OLS(y+1)	Log	-0.42357	0.014	-0.39894	0.025
Case 3: $V[y_i x]=\mu(x_i\beta)^2$					
PPML	Level	-0.00378	0.071	-0.00089	0.101
NLS	Level	0.34889	22.873	0.04003	2.030
GPML	Level	-0.00001	0.031	0.00036	0.064
OLS	Log	-0.00002	0.026	0.00074	0.053
OLS(y>0.5)	Log	-0.26688	0.033	-0.26648	0.055
OLS(y+1)	Log	-0.49065	0.019	-0.47112	0.034
Case 4: $V[y_i x]=\mu(x_i\beta)+\exp(x_{2i})\mu(x_i\beta)^2$					
PPML	Level	-0.00751	0.102	-0.00582	0.146
NLS	Level	0.58673	23.663	0.10991	3.206
GPML	Level	0.00410	0.057	-0.0009	0.109
OLS	Log	0.13249	0.039	-0.12444	0.075
OLS(y>0.5)	Log	-0.39215	0.042	-0.41328	0.072
OLS(y+1)	Log	-0.51437	0.021	-0.58055	0.041

Appendix Table 2. Monte Carlo results with standard estimators
($\beta_0=0$ and $k=1.0$)

Estimator	Dependent Variable Form	β_1		β_2	
		Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
Truncated OLS	Log	-0.0975	0.0357	-0.4425	0.0540
OLS (ln(y+0.1))	Log	0.2680	0.0276	-0.7498	0.0415
Truncated NLS	Level	0.0024	0.0114	-0.1923	0.0159
Censored NLS	Level	0.0079	0.0111	-0.1066	0.0164
GPML	Level	0.4107	0.1795	0.5727	0.0988
PPML	Level	0.0514	0.0110	-0.2906	0.0145
Truncated PPML	Level	-0.0250	0.0162	-0.1475	0.0238
Case 2: $V[y_i x]=\exp(x_i\beta)$					
Truncated OLS	Log	-0.1217	0.0419	-0.4934	0.0635
OLS (ln(y+0.1))	Log	0.2325	0.0299	-0.8597	0.0487
Truncated NLS	Level	0.0008	0.0254	-0.0870	0.0421
Censored NLS	Level	0.0074	0.0234	-0.1064	0.0407
GPML	Level	0.5348	0.1509	-0.5930	0.0898
PPML	Level	0.0337	0.0195	-0.2535	0.0359
Truncated PPML		-0.0239	0.0224	-0.1330	0.0369
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
Truncated OLS	Log	-0.2441	0.0585	-0.6173	0.0810
OLS (ln(y+0.1))	Log	0.0959	0.0392	-1.1442	0.0634
Truncated NLS	Level	-0.0185	0.1291	-0.0579	0.1948
Censored NLS	Level	-0.0051	0.1203	-0.1069	0.1802
GPML	Level	0.7518	0.1350	-0.6008	0.0977
PPML	Level	0.0318	0.0627	-0.2520	0.0853
Truncated PPML	Level	-0.0262	0.0730	-0.0823	0.0919
Case 4: $V[y_i x]=\exp(x_i \beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
Truncated OLS	Log	-0.3693	0.0650	-0.5617	0.1017
OLS (ln(y+0.1))	Log	-0.0116	0.0440	-1.4787	0.0762
Truncated NLS	Level	-0.0724	0.1786	0.0784	0.2739
Censored NLS	Level	-0.0228	0.1513	-0.0945	0.2338
GPML	Level	0.4298	0.1580	0.5478	0.1463
PPML	Level	0.0109	0.0843	-0.2130	0.1278
Truncated PPML	Level	-0.0859	0.1018	0.1198	0.1398

Appendix Table 3. Monte Carlo results with limited dependent variable estimators
($\beta_0=0$ and $k=1.0$)

Estimator	Dependent	β_1		β_2	
	Variable Form	Bias	Std Error.	Bias	Std Error.
Case 1: $V[y_i x]=1$					
ET-Tobit	Level	-0.0012	0.0687	0.0057	0.0145
ET-Tobit	Log	-0.0644	0.0057	-0.0318	0.0057
Poisson-Tobit	Level	0.0301	0.0145	-0.2513	0.0250
Heckman-ML	Log	-0.1421	0.0358	-0.3689	0.0546
Heckman-2SLS	Log	-0.1604	0.0355	-0.3326	0.0572
Case 2: $V[y_i x]=\exp(x_i\beta)$					
ET-Tobit	Level	-0.0027	0.0247	0.0375	0.0456
ET-Tobit	Log	-0.0670	0.0070	-0.0519	0.0111
Poisson-Tobit	Level	0.0337	0.0195	-0.2528	0.0364
Heckman-ML	Log	0.3878	-0.1830	0.0439	-0.3691
Heckman-2SLS	Log	-0.1947	0.0429	-0.3339	0.0682
Case 3: $V[y_i x]=(\exp(x_i\beta))^2$					
ET-Tobit	Level	-0.0294	0.1352	0.1957	0.2420
ET-Tobit	Log	-0.0998	0.0110	-0.1463	0.0166
Poisson-Tobit	Level	0.0212	0.0581	-0.2596	0.0826
Heckman-ML	Log	-0.3109	0.0607	-0.3636	0.0893
Heckman-2SLS	Log	-0.3038	0.0592	-0.3646	0.0900
Case 4: $V[y_i x]=\exp(x_i\beta)+\exp(x_{2i}) (\exp(x_i\beta))^2$					
ET-Tobit	Level	-0.0450	0.1695	0.3359	0.3277
ET-Tobit	Log	-0.1162	0.0132	-0.1150	0.0236
Poisson-Tobit	Level	-0.0182	0.0646	-0.2301	0.1138
Heckman-ML	Log	-0.4037	0.0660	-0.3398	0.1176
Heckman-2SLS	Log	-0.4010	0.0654	-0.3462	0.1162