New Approach to Estimating Gravity Models with Heteroscedasticity and Zero Trade Values

Mnasri, Ayman and Nechi, Salem

Qatar University, Qatar University

6 April 2019

New Approach to Estimating Gravity Models with Heteroscedasticity and Zero Trade Values

Ayman Mnasri[1]
Qatar University
Salem Nechi[2]
Qatar University

Abstract

This paper proposes new estimation techniques for gravity models with zero trade values and heteroscedasticity. We revisit the standard PPML estimator and we propose an improved version. We also propose various Heckman estimators with different distributions of the residuals, nonlinear forms of both selection and measure equations, and various process of the variance. We add to the existent literature alternative estimation methods taking into account the non-linearity of both the variance and the selection equation. Moreover, because of the unavailability of pre-set package in the econometrics software (Stata, Eviews, Matlab, etc.) to perform the estimation of the above-mentioned Heckman versions, we had to code it in Matlab using a combination of fminsearch and fminunc functions. Using numerical gradient matrix G, we report standard errors based on the BHHH technique. The proposed new Heckman version could be used in other applications. Our results suggest that previous empirical studies might be overestimating the contribution of the GDP of both import and export countries in determining the bilateral trade.

JEL classification: F10, F14, C01, C10, C13, C15, C63

Key words: Gravity model, Heteroscedasticity; Zero Trade values; New Heckman; New PPML

[1] Corresponding author. Assistant Professor of Economics, Department of Finance and Economics, College of Business and Economics, Qatar University, Doha, Qatar, P.Box. 2713. amnasri@qu.edu.qa, +974 44036756.
[2] Assistant Professor of Economics, Department of Finance and Economics, College of Business and Economics, Qatar University, Doha, Qatar, P.Box. 2713. snechi@qu.edu.qa, +974 44036497.

1. Introduction

A typical gravity model of international trade predicts that the trade flows between two countries depend mainly on their economic sizes (measured by the GDP) and the distance between them.[3] Broadly speaking, the gravity equation is built around a nonlinear relationship between the trade variable and a set of explanatory variables. The strong theoretical foundations of the gravity model (i.e., Anderson, 1979; Anderson and van Wincoop, 2003) resulted in its good performance in explaining trade flows between countries. However, despite its empirical success in predicting accurately trade flows, some estimation practices have been subject to criticisms. In fact, the way zero values of bilateral trade were treated and the approaches considered to handle the heteroscedasticity issues represent the main estimation problems of the gravity model. For instance, for estimation purposes, the gravity equation is usually log-linearized. This technique suffers from two major issues, though. The first issue is related to the way the log-linearized models treated the zero bilateral trade values and the second issue is related to the econometric applications of Jensen's inequality.

For the first problem, if the recorded trade between two countries has a value of zero, the log-linearization of the model leads to an estimation problem associated with the undefined cases of the dependent variable. Consequently, two methods have been widely used in the estimations of the log-linearized gravity models to deal with the zero trade flows: (i) estimate the model using only observations for which trade values are non-zeros (censored data); or (ii) augment the dependent variable (i.e., trade) by 1 to avoid undefined cases of log of zero. These proposed solutions could yield biased and inefficient estimators. In fact, the inconvenience of using censored data is that the reduction in the sample size is significant reaching more than 50% in some cases, especially in the disaggregated data. If the zero bilateral trade is not correlated with the explanatory variables in the gravity equation, then estimating gravity model without the zero trade observations (censored data) will not be biased. Yet, it is rational to think that the chance of having zero bilateral trade increases when the potential for bilateral trade between the two countries is low (so, it is obviously correlated with the explanatory variables). Hence, estimating the model without taking into account the zero observations could generate a bias in the estimated coefficients. Augmenting the dependent variable by 1, on the other hand, resolves the issue related to the reduction of the sample size discussed above, but it could present an important source a miss-specification in the estimated model as it interprets zero bilateral trade flows as absence of potential for trade between the two countries. This interpretation may not match the expectations of comparative advantage theories (based on relative factor abundance). In other words, the zero bilateral trade between two countries does not necessarily mean that the potential bilateral trade is a real zeros, but it could be explained by other factors. Moreover, as explained below, adding 1 to the dependent variable avoids undefined cases of log of zero but not the issue related to the econometric applications of the Jensen's inequality.

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

The second issue associated with the estimation of the log-linearized gravity equation is related to the Jensen's inequality, which implies that the expected value of the logarithm of a random variable will not be equal to the logarithm of its expected value, yet, it could depend on the mean and the variance of that random variable. Hence, when estimating the gravity model with the standard estimation techniques that use log-transformed data assuming constant elasticity, the intercept coefficient estimation will be biased. Moreover, the whole estimation could yield biased results for the other coefficients when the heteroscedasticity is correlated with the explanatory variables, which, indeed, could be the case as we will explain later.

Santos Silva & Tenreyro (2006) (SST (2006), hereafter) criticized conventional estimation practices of the log-linearized gravity trade models and proposed solutions to deal with the heteroscedasticity issue and the zero trade values. Broadly speaking, they argue that the gravity equation should be estimated in its multiplicative form using a Poisson Pseudo-Maximum-Likelihood (PPML, hereafter) estimation technique. The PPML is a special case of the Generalized Nonlinear Linear Model (GNLM) framework in which the variance is assumed proportional to the mean. SST (2006) show that this method is robust to different patterns of heteroscedasticity, deals with the Jensn's inequality and resolves the inefficiency problem. The work of SST (2006) is particularly striking given that their results raised important questions about the findings of many seminal studies in the trade literature (e.g., Anderson and van Wincoop, 2003) who predicted the coefficient on GDP close to one. The improvements that the PPML method has brought to the estimation of gravity models made it tractable in the international trade literature. In fact, it is has been used extensively in estimation of gravity equations (Bosquet and Boulhol, 2015; Egger and Tarlea, 2015; Dai et al., 2014; Lin, 2013; Yotov, 2012; de Sousa, 2012; Egger & Larch, 2011; Head et al., 2010; Shepherd, 2010; Fitzgerald, 2008; Tenreyro, 2007; among others).[4]

However, despite the proven robustness of the PPML, some issues related to the zero trade values and heteroscedasticity persist. The first issue is related to the way the zeros were dealt with. When we estimate the gravity equation with PPML and non-zero data, the estimation will suffer from the censoring bias discussed above (i.e., selection bias). When we include all data, the technical problem we use to have with log-transformation disappears; yet, in this case, we will force the estimated gravity model to predict a trade level that should be as close as possible to zero. However, it is obvious that the zero trade does not necessarily mean that the potential trade between the two countries is exactly zero. It is true that the chances of having zero trade between two countries should increase when their potential bilateral trade is small, but it does not necessarily imply it. At the same time, we can have zero trade between two countries where the potential trade that should be predicted by the gravity model is not necessarily close to zero. Moreover, the fact that gravity equation is an exponential function that, technically, can never be equal to zero, makes it invalid for dealing with zero trade' values.

The second issue related to the PPML estimator is technical and related to the assumption that the variance is proportional to the trade mean (i.e., $\exp(x\beta)$), which may raise questions about the optimality of the PPML estimator. In section 3, we show that we can obtain a slightly different estimator from the PPML. Moreover, when we estimate the nonlinear form of the gravity equation

[4] According to Google scholar, SST (2006) is cited 4097 times, as of February 2019.

(e.g, PPML, nonlinear least squares, etc.), we can avoid the Jensen inequality bias but estimators will not be efficient, given that variance process is usually non-constant. In other words, while the PPML estimator resolves the Jensen's inequality issue, the inefficiency problem persists (as well as the zero trade problem).

Subsequent attempts in the literature to deal with heteroscedasticity and the prevalence zero trade values did not fully address these issues. For instance, the concerns related to the prevalence of zero trade values do not seem to be resolved. In fact, the appeal to the standard Heckman correction for sample selection to deal with the zero trade flows has some limitations. As a general rule, when trade data is characterized with frequent zero values, selection equations are traditionally estimated on the basis of the same set of explanatory variables. Although Martin and Pham (2008) include additional variables in the selection equation, their approach has limitations too. First, they have no flexibility in modelling the selection equation as they assume a linear selection equation. They log-linearize the gravity model, most probably, to be able to estimate it with the standard statistical packages (Stata, Eviews, etc.), which will result in the Jensen inequality bias (heteroscedasticity). Second, the maximum likelihood estimation they propose assumes normal distribution of the error terms (while it could be lognormally distributed). Third, the Heckman model used in the conventional statistical packages (Stata and other software) assume constant variance, which leads to efficiency problem.

We consider the above approach has shortcomings. Our contribution consists in dealing with these shortcomings by developing a new Heckman model to improve the existent estimators (in addition to the contribution related to the improvement of the PPML). First, we propose various Heckman estimators with different distributions of the residuals, (normal & log normal). Second, our new proposed Heckman allows for nonlinear forms of the selection equation, which would provide more flexibility in modelling the zero trade values. Indeed, zero bilateral trade could be due to a low potential trade or/and to other variables that can affect the non-trading decisions and not necessarily the potential bilateral trade between the two countries. Third, we explicitly model the conditional variance process as function of $\exp(x\beta)$ and $\exp(x\beta)^2$. So, our new proposed estimator allows for constant variance, variance proportional to trade level, variance proportional to the squared of the trade level, or for any combination of those variance processes. To do so, we estimate the conditional variance process simultaneously within the log-likelihood function of the New Heckman model.

Other important contributions of this paper are related to the computational challenges we face when estimating our augmented model and reporting the standard errors. The biggest challenge we have encountered is how to perform the estimation of the new Heckman versions (log-normal distribution, nonlinearity, conditional variance estimation). No pre-set package in the econometrics software (Stata, Eviews, Matlab, etc.) can deal with this estimation. We had to code in Matlab and use a combination of *fminsearch* and *fminunc* functions. Another challenge is related to how to report standard errors of the estimated coefficients. Using numerical gradient matrix G (Davidsson and Mackinnon, 2003), we report standard errors based on the BHHH technique. The proposed new Heckman version could be used in other applications.

The remainder of the paper is organized as follows. Section 2 provides a brief discussion of the literature while Section 3 highlights the theoretical and empirical applications of the gravity model. In section 4, we present our proposal to deal with the zero trade values and heteroscedasticity. Section 5 presents the simulation results and compare our estimator with the PPML, and the commonly used estimators in the literature under different specifications of the variance process and the Heckman selection equation. Section 6 presents the new estimates of the gravity equation. Section 7 concludes.

2. Literature Review

In this section, we review the literature on gravity models; discuss the proposed estimation methods, the main findings, and the drawbacks. The most problematic issues related to the estimation of gravity models are the heteroscedasticity and zero trade values. Note that in dealing with the issue of zero trade values, while most studies considered truncated or censored data, a trend in the literature had attempted to estimate the gravity equation without deleting zero trade values. Parametric and semi-parametric approaches were used in this trend of the literature.

Parametric approaches that estimate gravity models that include limit observations adopted two types of models: Two-Part model or a Tobit/Heckman model. Broadly speaking, the Two-Part model suggests that we first use a qualitative-dependent model such as Probit to determine whether a particular bilateral trade flow will be zero then estimate the relationship between trade values and explanatory variables using only truncated data (Leung and Yu, 1996).[5] The Two-Part model allows the sample selection and the behavioral equations to be estimated independently (see for example, Duan et al., 1983), which is 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. The Tobit/Heckman model, on the other hand, proposes to either use two-step estimators such as Heckman (1979) or a maximum likelihood approach such as Tobin (1958), Puhani (2000) or Jones (2000). Puhani (2000) reviewed the literature on the Heckman correction for sample selection bias and concluded 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. Consequently, the Tobit/Heckman approach has been used more often in the literature.[6] Note also that some semi-parametric models (such as Chay and Powell, 2001) had attempted to estimate the gravity equation without deleting the limit observations. Such applications, however, have been infrequent because of the computational efficiency problems related to them.

In an influential paper, SST (2006) addressed the heteroscedasticity and zero trade values issues using the Pseudo Poisson Maximum Likelihood (PPML) estimator. Although this estimator

[5] Potential estimators include OLS, NLS (Frankel and Wei, 1993), PPML and GPML (SST, 2006).

[6] Regardless of the type of parametric model adopted, the most common distribution about the residuals is that they are normally distributed. Exceptions exist though (Poisson and Gamma distributions highlighted by SST, 2006).

outperformed those obtained with traditional methods (OLS, NLS, etc.) it was subject to some criticisms, and many alternative estimation methods were proposed (Gamma Pseudo Maximum-Likelihood, Feasible Generalized Least Squares, ET-Tobit, etc.). Martin and Pham (2008) revisit the work of SST (2006) and investigate to what extent their PPML estimator can deal with the biases of estimators that resulted from the heteroscedasticity and the prevalence of zero trade values. They conclude that the recommended PPML estimator successfully deals with the heteroscedasticity bias, only when the zero trade flows issue is not significant. In fact, they show that if the data is characterized by high frequency of zero trade values the PPML estimator is not bias free. Therefore, the PPML could be applied to a range of multiplicative (but not gravity) models such us consumer-demand systems, Cobb-Douglas type production functions and the stochastic impacts by regression on population, affluence and technology model, etc., when the number of zero values is relatively small. When the zero trade values are frequent, however, Martin and Pham (2008) argue that a standard threshold Tobit estimators perform better if the heteroscedasticity problem is satisfactory dealt with. Moreover, they show that if true identifying restrictions are available, the Heckman Maximum Likelihood estimators perform well. Santos Silva and Tenreyro (2011) argued that Martin and Pham (2008) results were based on misspecified models and showed that the PPML remains the best estimator even in the presence of high frequency data points of zero trade values.

Based on Monte Carlo simulations, Martínez-Zarzoso (2013) compares the PPML estimator with the Gamma Pseudo-Maximum-Likelihood (GPML), a nonlinear least square (NLS) estimator and a feasible generalized least squares (FGLS) estimator using three different data sets. She finds that the PPML is less affected by the heteroscedasticity but it does not outperform the other estimators in terms of bias and standard errors. Simulations without zero values, however, show that GPML presents the lowest bias and standard errors. She concludes that, for any application, the selection of the most appropriate estimator requires a number of tests and depending on the characteristics of each dataset. The focus of Martinez-Zarzoso (2013), however, is on cases with no zero trade values or with relatively low frequency of zero values for the dependent variable.

Kareem (2013) estimates a gravity equation in the presence of zero trade using trade data for 47 African countries over the period 1980-2002. She considered a large set of models including the log linear model, pooled regression model, fixed effects model, random effects model, multiplicative models, the generalized linear models (GLM), the Poisson Pseudo Maximum Likelihood (PPML) estimator, Negative Binomial Poisson Maximum Likelihood (NBPML) estimator. His results show that there is no one general best performing model, although most of the linear estimators outperform the GLM estimators in many of the robust checks performed. Herrera and Baleix (2010) provide a survey of the literature concerning the specification and estimation methods of the gravity equation and the several problems related its empirical application. For this purpose, they use a gravity equation based on Anderson and van Wincoop (2003) model and discuss the fit of different estimation procedures (OLS, panel regression with fixed and random effect and simple and panel Poisson methodology) applied to a large dataset of bilateral exports for 47 countries (80% of world trade) over the period 1980-2002. Their findings suggest that none of the estimators outperform the others in all aspects.

Xiong and Chen (2014) estimate gravity equation in the presence of sample selection and heteroscedasticity, using a two-step method of moments (TS-MM) estimator. Their Monte Carlo experiment shows that the TS-MM estimates are resistant to various combinations of sample selection and heteroscedasticity. Moreover, the TS-MM estimator performs reasonably well even when data-generating process deviates from the TS-MM assumptions.

Sukanuntathum (2012) performs two steps estimation of gravity models under heteroscedasticity and data censoring. He considers different estimators including OLS, NLS, PPML, GPML and NBPML (Negative binomial pseudo-maximum likelihood estimator). He recommends the use of the NBPML in the second step when both heteroscedasticity and zero flows occur because it gives consistent parameter, robust to different forms of heteroscedasticity and greatly deal with zero flows. Burger et al., (2009) use modified Poisson fixed effects method to estimate a gravity model of trade with excess zeros. Their Zero-Inflated estimation technique provides viable alternatives to both the lognormal and standard Poisson specification of the gravity model of trade in the presence of high frequency of zero trade values and failure of homoscedasticity assumption. Herrera (2013) compares alternative methods to estimate the Anderson and van Wincoop (2003) model. The estimators considered are truncated OLS, OLS (1+X), Tobit, Probit, Heckman, Panel fixed, Panel random, and the PPML. The findings suggest that the ad hoc methods are not appropriate for estimating the gravity equation since they provide biased and inefficient estimates. Although several authors have proposed the use of PPML, it does not behave so well for an aggregated dataset in the presence of unobserved heterogeneity. He suggests that Heckman sample selection model is preferred estimation method within nonlinear techniques in the presence of heteroscedasticity and significant proportion of zero observations.

Joakim and Fredrik (2009) examine the effects of zero trade on the estimation of the gravity model using both simulated and real data with a panel structure, which is different from the more conventional cross-sectional structure. They suggest an alternative approach to the usual log-linear estimation method, which can result in highly deceptive inference when some observations are zero. Their proposal consists in using the Poisson fixed effects estimator, which eliminates the problems of zero trade and is shown to perform well in small samples. Arvis and Shepherd (2011) argue that in addition to dealing with heteroscedasticity and zero trade flows, the Poisson estimator also solves the adding up problem. They also argue that it is the only quasi-maximum likelihood estimator that preserves total flows between the actual and estimated bilateral trade matrices. The theoretical and empirical findings strengthen the case for using Poisson as the workhorse gravity model estimator.

Krisztin and Fischer (2015) argue that estimating the gravity equation by means of PPML lead to consistent, but biased parameter estimates if spatial dependence between origin-destination flows is ignored. To overcome this problem, they suggest eigenvector spatial filtering variants of the Poisson gravity model (with or without zero inflation) along with pseudo maximum likelihood estimation. Assane and Chiang (2014) find that OLS and Heckman models produce coefficient estimates for distance and other trade costs parameters that are higher than those of PPML estimator, which is consistent with SST (2006) findings. This result could be driven by the large fraction of zero trade flows as well as by heteroscedasticity problems. They conclude that while Heckman and PPML are the appropriate estimation procedures to handle zero trade flows, the PPML estimator is the only estimator that deals with heteroscedasticity problems.

3. The Theoretical and Empirical Applications of the Gravity Model

Consider the econometric formulation of the traditional gravity equation

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

where y_{ij} represents the trade flows from country i to country j; x_i and x_j represent the GDP for countries i and j, respectively; D_{ij} is the distance between countries i and j; and $\beta_1, \beta_2,$ and β_3 are unknown parameters. η_{ij} is an error term with expectation, $E(\eta_{ij}|x_i, x_j, D_{ij})$, of one and assumed to be statistically independent of the regressors which means that

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

To estimate this model, traditional approaches in trade literature start by log-linearizing equation (1) then estimating the parameters of interest by least squares using the following equation:

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

Two major issues are associated with this approach. First, if data includes values of zero, estimating the gravity models in their log-linearized forms such as equation (2) lead to complications associated with observations for the dependent variable with zero trade values. Many methods have been considered in the literature to deal with this issue (simple deletion of the observations of zero trade flows from the data or substitute the observations of zero trade values with a very small number, typically 1). Criticisms to these proposed methods argue that the estimators of the parameters based on these two approaches will be inconsistent (we discussed these issues in the Introduction and literature review sections). Second, estimating equation (2) with least squares can produce biased estimators if the $var(\eta_{ij})$ is correlated with the regressors. In fact, there is evidence to believe that that variance of bilateral trade data is not constant (SST, 2006). Moreover, as emphasized by SST (2006), the trade variance might be correlated with the trade level. In fact, if the trade level between two countries is significantly high, we should expect the variability of trade values to be also high, and vice-versa. In other words, the mean of the error terms ($\ln \eta_{ij}$) and regressors ($x_i, x_j, or\ D_{ij}$) of equation (2) are correlated implying that the OLS estimation will be biased. Unfortunately, many studies neglect this fact leading to a bias related to Jensen's inequality.

3.1 Estimation in the Presence of Heteroscedasticity

Economic theory suggests that if y and x are linked by constant-eleasticity model, we can write the model as[7]

$$y_i = \exp(x_i\beta) + \varepsilon_i \qquad (3)$$

where y_i represents the bilateral trade, x_i is a vector of explanatory variables, $\beta_i$is a vector of coefficients, and ε_i is an error term. With $y_i \geq 0 \;\; and\; E[\varepsilon_i|x] = 0$, (3) can also be written as

$$y_i = \exp(x_i\beta)\,\eta_i$$

with $\eta_i = 1 + \frac{\varepsilon_i}{\exp(x_i\beta)} \;\; and\; E[\eta_i|x] = 1$. Taking the logarithms of both sides (assuming that $y_i \geq 0$) leads to

$$\ln y_i = x_i\beta + \ln \eta_i \qquad (4)$$

The estimation of the above log linear representation of the constant-elasticity model is useful under very specific conditions on the error term. Note that when η_i is independent of x_i , the conditional variance of $y_i(and\; \varepsilon_i)$ is proportional to $\exp(2x_i\beta)$.

3.1.1 Proposed estimation solutions in the literature, SST (2006)

SST (2006) argue that the conditional variance will depend on $\exp(x_i\beta)$. Therefore, OLS estimators will be inconsistent. Moreover, SST (2006) argue that the heteroscedasticity issue leads to estimates biased by 35% or more. To deal with these problems, SST (2006) propose an alternative approach consisting at estimating the multiplicative form of equation (2). That is, estimating

$$y_{ij} = \exp\left[ln\beta_0 + \beta_1 ln\, x_i + \beta_2 ln\, x_j - \beta_3 ln\, D_{ij}\right]\eta_{ij} \qquad (5)$$

using the Poison Pseudo-Maximum Likelihood (PPML) estimator.[8] η_{ij} is a log normal random variable with mean 1 and variance σ_i^2.

The results of SST (2006) are based on simulation experiment to assess the performance of a set of simple pseudo-maximum likelihood (PML) estimators of many constant-elasticity models in the presence of heteroscedasticity (i.e., different specifications of the process generating σ_i^2) and zero trade values, focusing on the PPML. The PML estimators considered in their exercise are

[7] The general form for a constant-elasticity model is $y_i = \exp(x_i\beta)$, the function $\exp(x_i\beta)$ is interpreted as the conditional expectation of y_i given x , denoted $E[y_i|x]$. Since the relation between y and x holds on average but not for every i, an error term associated with each observation can be defined as $\varepsilon_i = y_i - E[y_i|x]$.

[8] This approach is based on McCullagh and Nelder (1989) model which estimates the parameters of interest using a Pseudo-Maximum Likelihood estimator.

those of the NLS, gamma PPM, OLS of the log linear, a truncated OLS, and ET-Tobit. SST (2006) aimed measuring the magnitude of the biases resulting from two experiments. Their results show that despite the different specifications of the heteroscedasticity, almost all estimators are badly biased. Moreover, except under very special circumstances, estimation based on the log-linearized model cannot be recommended. The results of the PPML estimator are very encouraging though. In fact, its performance is reasonably good for all specifications of the heteroscedasticity. Therefore, the PPML method seems to be the best approach to estimate constant-elasticity models.

It is true that when we use PPML, estimation will not suffer from the Jenson inequality bias, yet, the estimations will not be efficient, given the high degree of volatility in the variance process, especially when combined with the presence of zero observations. Despite its popularity and its superiority over many other estimators, we believe that an improved version of the PPML can be obtained. To see how this is the case, we re-write the minimization problem and show that the FOC should be different form the PPML FOC.

The objective function of the nonlinear least squares (NLS) estimator can be formulate as follow:

$$\hat{\beta} = \arg\min_{b} \sum_{i=1}^{n} [y_i - \exp(x_i\beta)]^2,$$

The FOC derived from the above objective function (equation (8) in SST, 2006) can be written as

$$\sum_{i=1}^{n} \left[y_i - \exp\left(x_i\hat{\beta}\right)\right] \exp\left(x_i\hat{\beta}\right) x_i = 0$$

SST (2006) argue that we can get a more efficient estimator by following McCullagh and Nelder (1989) and estimate the parameters of interest using a PML estimator based on the assumption on the functional form of the conditional variance is proportional to $\exp(x_i\beta)$. The proposed estimator by SST (2006) consists in dividing the NLS FOC by the conditional variance, which yields the following FOC:

$$\sum_{i=1}^{n} \left[y_i - \exp(x_i\tilde{\beta})\right] x_i = 0$$

The advantages of the PPML can be summarized as follow. First, it takes into account the heteroscedasticity by not overweighting noisier observations (large $\exp(x_i\beta)$). PPML also avoids the Jenson inequality bias as it estimates the nonlinear form and allows to include zeros observations. Moreover, the fact that PPML is already available in many econometrics software (SATA for instance), makes it very attractive and practical. It is also important to note that simulations exercises of SST (2006) and subsequent empirical investigations (including the present study) show that PPML is relatively robust to different variance process scenarios. However, one could argue that if the variance process form is known (proportional to $\exp(x_i\beta)$), it would be more efficient to include it in the objective function rather than dividing the NLS FOC by $\exp(x_i\beta)$ as SST(2006) did.

3.1.2 Our proposed approach, "Optimal PPML"

In the present study, we start from the same assumption made by SST (2006) which assumes that the conditional variance process is proportional to $\exp(x_i\beta)$, then we derive a new FOC from a modified version of the objective function where errors are divided by their corresponding standard deviations, $\exp\left(\frac{1}{2}x_i\beta\right)$. The new objective function can be written as:

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} \left[\frac{y_i - \exp(x_i\beta)}{\exp\left(\frac{1}{2}x_i\beta\right)}\right]^2, \quad (6)$$

The FOCs are then:

$$\sum_{i=1}^{n} \left[\left(\left(y_i - exp(x_i\hat{\beta})\ \right).x_i\right).\left(\frac{(y_i + exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}\right)\right] = 0\ , \quad (7)$$

which are different from the FOCs of the PPML estimator. It can be shown that the expectation of the FOCs (7) given x are equal to $\sum_{i=1}^{n} x_i$.[9] To ensure that FOCs expectation are zeros, we can simply use demeaned variables in the estimation. Even if we do not, the only biased coefficient will be the constant term. It is also worthy to note that our FOCs are, to some extend, asymptotically equivalent to the standard PPML FOCs[10].

An important step in our approach is the estimation of the variance of estimators**.** FOC (7) is equivalent to a vector of moment conditions: $W'(y - \exp(x\beta)) = 0$, where each element of W is

$$x_i \frac{(y_i + exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}.$$ [11]

Following Davidsson and Mackinnon (2003), we can show that a reasonable way to estimate a Heteroscedasticity-Consistent Covariance Matrix Estimator (HCCME) of β is to use the sandwich covariance matrix:

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

---

[9] $E\left\{\sum_{i=1}^{n}\left[\left(\left(y_i - exp(x_i\hat{\beta})\right).x_i\right).\left(\frac{(y_i + exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}\right)\right] \mid x_i\right\} =$
$E\left\{\sum_{i=1}^{n}\left[(\varepsilon_i.x_i).\left(\frac{\varepsilon_i + 2\,exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}\right)\right] \mid x_i\right\} = E\left\{\sum_{i=1}^{n}\left[\left(\frac{\varepsilon_i{}^2 x_i + 2\,\varepsilon_i\,x_i\,exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}\right)\right] \mid x_i\right\} =$
$\sum_{i=1}^{n}\left(\frac{E(\varepsilon_i{}^2.x_i|x_i)}{exp(x_i\hat{\beta})}\right) = \sum_{i=1}^{n} x_i$.

[10] Since $E(y) = exp(x_i\beta)$, $\frac{(y_i + exp(x_i\hat{\beta}))}{exp(x_i\hat{\beta})}$ will vanish and (7) coincides with PPML FOCs.

[11] For the standard PPML, the i^{th} element of W is simply x_i

Where $\hat{X}$ with typical element corresponding to an estimate of the derivative of $exp(x_i\beta)$ with respect to b which is: $x_i\, exp(x_i\hat{\beta})$. $\widehat{\Omega}$ is $n \times n$ diagonal matrix with the squared residual $\hat{\varepsilon}_i{}^2$as the t^{th} diagonal element.
If we assume that the variance is constant, the covariance matrix becomes:

$$\left(\hat{X}'\,\widehat{W}\,\left(\widehat{W}'\widehat{W}\right)^{-1}\widehat{W}'\hat{X}\right)^{-1}\,\hat{\sigma}^2$$

SST(2006) reported standard errors based on the last version of the covariance matrix which assumes constant variance of the residuals.

3.2 Problems Associated with the Prevalence of zero values of the Dependent Variable

Zero trade values became frequent in most trade datasets. For instance, SST (2006) data contains almost 50% zero trade values. Similarly, about 50% of the observations on bilateral trade in Helpman et al. (2007) dataset were of zero values. Baldwin and Harrigan (2007) argue that more than 90% of potential import flows to the USA are zero. According to Martin and Pham (2008), 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. Although the reported zero values in datasets may reflect errors, omissions or non-reporting, most of zero trade values in carefully prepared datasets reflect true absence of trade (Martin & Pham, 2008). A significant trend in the literature attributes the absence of trade flows (hence, zero values) to the high fixed costs associated with bilateral trade.

Hence, including zeros in the estimation of the gravity model could lead to substantial biases as this would force the estimated gravity model ($\exp(X\beta)\,.\,\eta$)) to predict a trade level that should be as close as possible to zero. However, it is obvious that the zero trade does not necessarily mean that the potential trade between the two countries is exactly zero. It is true that the chances of having zero trade between two countries should increase when their potential bilateral trade is small, but it does not necessarily imply it. At the same time, we can have zero trade between two countries where the potential trade that should be predicted by the gravity model is not necessarily very close to zero. Excluding zero values would be a reasonable solution if the censoring process is random and is not correlated with the explanatory variables. However, it is rational to believe that the chance of having zeros trade between countries increases when the potential bilateral trade decreases. Thus, excluding zeros and estimating truncated data would suffer from selection bias.

Tobin (1958) had shown that zero values of the dependent variable could create potentially large biases in parameter estimates, even in linear models, if the estimator used does not allow for this feature of the data generating process. Hurd (1979) shows that estimations of gravity models with truncated data would result in large biases. Arabmazar and Schmidt (1981) find these problems less serious in the censored regression case where the zero values are retained.

3.2.1 Theoretical Illustration of the problem

The inclusion of the zero bilateral trade can lead to two types of bias, misspecification (i.e., systematic) and correlation biases.

Misspecification bias: For zero trade observations, the multiplicative gravity equation is obviously misspecified as $E(y|x) \neq \exp(x\beta)$ given that exponential function is strictly positive. Without loss of generality, let's assume that $x \geq 0$. In this case, the estimation of gravity model including zero values would underestimate β (negative bias) as the model will tend to assign unrealistically large negative value to β for $y = 0$. So, even when the zero occurrence is not correlated with explanatory variables, including zeros would yield, in principle, biased coefficients.

Correlation bias: In case where zeros occurrence is correlated with x, the sign of the bias caused by including zero values in the data, will depend on the sign of the true β.

a) If $\beta > 0$, then the estimated $\hat{\beta}$ will be overestimated. In this case, the overall bias (systematic bias + correlation bias) will be ambiguous.
b) If $\beta < 0$, then the magnitude of $\hat{\beta}$ will be amplified (negative bias), and the overall bias well be negative.

3.2.2 Proposed estimation solutions in the literature

Two techniques have been used in the literature to deal with prevalence of the zero trade values: single equations and two-equation techniques. Single equation techniques consist in estimating linear and nonlinear models. The oldest and mostly used linear single equation single equation estimation is the OLS method used to estimate linear regressions while logit, probit and Tobit are among the widely used nonlinear models. Although many estimators have been used to estimate the gravity equation, the PPML is the among the most efficient and influential single equation estimators. The empirical application of SST (2006) minimized the bias of the estimators and dealt with the zero values by avoiding the log-linear specification. Moreover, SST (2006) results raised important questions related to the main findings of the widely used model of Anderson and van Wincoop (2003), namely, a coefficient of almost one on GDP.[12] The PPML and other nonlinear estimators implicitly assume that there is nothing special about the zeros. The problem is just to get them in the estimation sample. However, as we discussed above, zero trade values do not necessarily imply the absence of potential of trade between two countries.

For the two-equation techniques, since Tobin (1958), many studies attempted to deal with the prevalence of zero values. However, the most influential and significant work was conducted by Heckman (1979). Broadly speaking, Heckman (1979) generalized Tobin's (1958) approach to estimation in the presence of this problem based on non-random sample selection. According to the selection framework proposed by Heckman, the dependent variable y_i is observed for a part of the data only. In general, the linear form of the Heckman selection model is specified as follow:

---

[12] The tractability and robustness of SST (2006) model attracted a significant literature (Westerlund & Whilhelmsson, 2007; Xuepend Liu, 2007; Hebble et al., 2007, among others).

$$\begin{cases} y_i = X_i\beta + \varepsilon_i & (8) \\ z_i = W_i\gamma + u_i & (9) \end{cases}$$

where z_i is a binary variable $\{0,1\}$. y_i is observed, only when $z_i = 1$. The error terms $\varepsilon_i and\ u_i$ follow a bivariate normal distribution with correlation ρ. Equation (8) is the response equation and (9) is the selection equation, which determines whether observations have a non-zero value. Estimating equation (8) in the presence of sample selection would result in biased estimates of the coeficients $\beta_i s$ because of the omitted relevant explanatory variables.[13] In our case, the dependent variable is the trade value between two countries.

As discussed earlier, when estimating data without zero trade values, we can face a selection bias, if the selection process is not random, which is more likely to be the case in the gravity equation. That is, the chance of having zeros trade between country increases when the potential bilateral trade decreases. Even if we do estimate the data including zero observations, we may also face another issue: specification problem. Indeed, in this case, the model would interpret zeros as absence of any potential trade, which is not necessarily the case (overestimating coefficients). That is why we need to use an estimation technique that accurately models the zeros occurrence process.

3.2.3 Our proposed approach: "New Heckman"

Model specification and selection seem to be very challenging tasks when estimating Gravity models with many zero values of the trade flows. Although many functional forms have been used in the trade literature, the nonlinear relationship between the variables (in levels) seems to be more tractable and widely used. In fact, the estimation of the log linear representation would work only with non-zero observations (truncated data) which would lead to bias if there is correlation between the zero occurrence and explanatory variables. Besides the issues related to the prevalence of zero trade values, there is evidence that variance of bilateral trade data is not constant. Note, however, that although the nonlinear estimators (NLS, PPML, etc.) help avoiding the Jenson inequality bias, yet, the estimators will not be efficient, especially when combined with the presence of zero observations, which could lead to other type of bias as zero trade value does not necessarily mean that the model should predict zero potential bilateral trade.

Heckman seems to be a rational remedy. Yet, it does not deal with the heteroscedasticity and assumes normal errors. Moreover, since econometrics software do not allow the estimation of non-linear Heckman Models, previous attempts to work with Heckman (or Tobit) had to log-transform the data. It is true that this approach is consistent with the lognormal errors (since the log of lognormal distribution is a normal distribution), yet, we will then encounter the Jenson 'inequality bias as those estimation techniques assume constant variance.

To overcome these shortcomings, we propose NEW HECKMAN estimators based on the nonlinear model, a new selection equation, and an estimation of the variance process.

---

13 See Martin and Pham (2008) for a graphical and intuitive explanation.

The selection equation we consider is given by:

$$y_1 = \exp(x\beta)\eta$$
$$y_2 = \exp(x\beta) + z\alpha - k + \mu \qquad (10), \qquad \binom{\eta}{\mu} \sim \left(0, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right)$$

$$\begin{cases} y = y_1, \; y_2 > 0 \\ y = 0, \; y_2 \leq 0 \end{cases} \qquad (11)$$

Where ρ is the correlation between μ and η, and σ is the standard deviation of the measure equation. k is a constant term. As will be explained later, the term k has two interpretation. First, it ensures that the trade between two countries takes place only if a minimum potential trade or a threshold is met. Second, for simulation purposes, k is introduced to ensure the prevalence of a significant number of zero trade values. The selection equation determines whether a bilateral trade can occur or not.

The log-likelihood function is[14]

$$\sum_{y_2 \leq 0} log\, \Phi(-\exp(x\beta) - z\alpha + k) + \sum_{y_2 > 0} log\, \frac{1}{\sigma} f\left(\frac{\mathrm{y} - \exp(x\beta)}{\sigma}\right) +$$
$$\sum_{y_2 > 0} log\, \Phi\left(\frac{\exp(x\beta) + z\alpha - k + \rho(\mathrm{y} - \exp(x\beta))/\sigma}{(1-\rho^2)^{\frac{1}{2}}}\right) \quad (12)$$

$where: \sigma^2 = |\omega_0| + |\omega_1|\hat{y} + |\omega_2|\hat{y}^2$, $\hat{y} = \exp(x\beta)$. $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. f(.) is the probability density function (Normal or Lognormal distribution)

4. Estimation Techniques

In this section, we revisit SST (2006) and the different Heckman models specifications (standard Heckman and Marin and Pham, 2008) and extend them to cases where the selection equation and heteroscedasticity take nonlinear forms, and the distribution of the error terms is not restricted to the normal case.

4.1 Parameters estimations

We estimated the parameters using two estimation techniques: single equations and two equations.

4.1.1 Single equations

[14] When the ML estimation includes calculating the probability of a function of y (log(y) or $\frac{y}{\exp(x\beta)}$) , we need to adjust the loglikelihood function accordingly. For instance, for the case of $\frac{y}{\exp(x\beta)}$, we need to add the following term the loglikelihood function: - log(exp($x\beta$)) = - $x\beta$. In fact, if m and n are two dependent variables and f(.) is a probability density function, then: $f(m) = f(n)\left|\frac{\partial n}{\partial m}\right|$

Our single equations techniques are based on the traditional methods used in the literature to estimate the gravity equation (linear using the log linear forms and nonlinear using the multiplicative forms). For the log-linearized forms, we consider censored data only while for the multiplicative forms we consider both complete and censored data. Our estimators include OLS, Heckman 2SLS, NLS, Censored NLS ($y > 0$), PPML, Censored PPML ($y > 0$), and the a new PPML estimator, 'Optimal_PPML', defined by equation (7) (derived form the modified version of the objective function, equation (6)). We also estimate generalized versions of Heckman 2SLS, NLS, and a Censored NLS. Generalized versions are obtained by running the following loop:

i. Estimate constant variance model
ii. Use squared residuals ($\hat{\varepsilon}^2$) to estimate the variance process:

$$Var = \gamma_0 + \gamma_1 \hat{\varepsilon}^2 + \gamma_2 [\hat{\varepsilon}^2]^2 \qquad (13)$$

iii. Estimate the equation with new variance process
iv. Continue until convergence

The estimation of the generalized versions constitute one of our contributions in this paper. Table 1 below describes our single equation estimators.

Table 1. Single equations estimators

Estimator	Notation
Ordinary least Squares	OLS
Heckman Two Stages least Squares	Heckman 2SLS
Generalized Heckman Two Stages Least Squares	GHeckman 2SLS
Log Heckman Two Stages Least Squares	heckman2sls_Log
Log Generalized Heckman Two Stages Least Squares	Gheckman2sls_Log
Nonlinear lease Squares	NNLS
Generalized Nonlinear Least Squares	Gnnls
Nonlinear lease Squares Censored	nnls_C
Generalized Nonlinear least Squares Censored	Gnnls_C
Poisson Pseudo-Maximum Likelihood	PPML
Poisson Pseudo-Maximum Likelihood Censored	PPML_C
Optimal Poisson Pseudo-Maximum Likelihood	Optimal_PPML
Optimal Poisson Pseudo-Maximum Likelihood Censored	Optimal_PPML_C

4.1.2 Two-equation estimators

Our two equations techniques are based on various specifications of the mesure and the selection equations. For the measure equation, although our theoretical specification is aligned with the literature, our estimation approach is a novel contribution. In fact, while standard two equations models log linearize the measure equation for estimation purposes, we estimate the nonlinear form, which would avoid the Jenson inequality bias. Our contribution consists also in considering different processes of the error terms (normal vs. lognormal) and endogenously estimating the variance process (constant, nonlinear of order one, and of order two, or a combination of them). As for the determination of the selection equation, we consider three methods: the ET-Tobit model, the standard Heckman model and our proposed New Heckman model discussed above.

Our first exercise in the two-equation framework consists in revisit the standard Heckman approach, which is based on the following model

$$\begin{aligned} y_1 &= \exp(x\beta)\eta \\ y_2 &= \quad x\alpha + \nu \end{aligned} \qquad (14)$$

And it follows that

$$\begin{cases} y = y_1, \ y_2 > 0 \\ y = 0, \ y_2 \leq 0 \end{cases} \qquad (15)$$

For estimation purposes, the above model is log-linearized.

Next, we propose a new selection equation where the occurrence of zero depends on the level of the potential trade and other variables not included in the measure equation.

$$y_2 = \exp(x\beta) - k + \alpha z + \mu \qquad (16)$$

where k is a constant term to ensure that a minimum potential trade or a threshold that should be met for a trade to take place. The selection equation determines whether a bilateral trade can occur or not. If the selection equation (i.e. the error term μ) is correlated with the measure equation (i.e. the error term η_i), then Heckman estimation is efficient and the selection bias will be avoided. However, when the error terms in the two equations are not correlated, the Heckman estimation becomes equivalent to a standard Maximum Likelihood (ML) estimation of the censored data. In the case a gravity model, it is rational to believe that factors determining whether or not a bilateral trade between two countries occurs depend on two factors: (i) the potential trade predicted by the gravity model. In fact, we can think of a minimum potential trade or a threshold (k) that should be met for a trade to take place, (ii) variables possibly omitted in y_1 or external factors affecting the occurrence of trade itself which we try to capture by αz. Obviously, in this case, the two errors terms (η and μ) are correlated, which makes the Heckman technique plausible.

This last selection equation represents one of our main contributions in this paper. In fact, the selection equation we propose estimates significantly less coefficients compared to standard Heckman selection equation, where we need to estimate new coefficients for all explanatory variables of the measure equations (i.e., αs). Clearly, by using the new Heckman method we substantially decrease the computational burden. At the same time, the assumptions made in this method might not be very restrictive. In fact, it is rational to think that the x variables affect the selection process through the implied potential trade ($x\beta$). Therefore, putting $x\beta$ in the selection equation y_2 instead of estimating new coefficients ($x\alpha$) seems to be very intuitive. Obviously, we need to maximize the log-likelihood function. Estimating these nonlinear, non-smooth with a relatively high number of coefficient is very challenging. Another contribution of this paper is the MATLAB code we use to run this maximization problem.

We estimate our two equations models (ETobit, standard Heckman and New Heckman) using different specifications of the distribution of the error terms (Normal Vs Lognormal) and different variance process scenarios. In total, four (4) scenarios were considered:[15]

- Scenario 1: $\eta \sim N(1, \frac{1}{\alpha[\exp(x\beta)]^2})$ which is equivalent to $y \sim N(\exp(x\beta), \frac{1}{\alpha})$ where α is a constant .
- Scenario 2: $\eta \sim N(1, \alpha)$ which is equivalent to $\frac{y}{\exp(x\beta)} \sim N(1, \alpha)$.
- Scenario 3: $\eta \sim logN(1, \frac{1}{\alpha[\exp(x\beta)]^2})$ which is equivalent to $y \sim logN(\exp(x\beta), \frac{1}{\alpha})$.
- Scenario 4: $\eta \sim LogN(1, \alpha)$ which is equivalent to $log\frac{y}{\exp(x\beta)} \sim N(\mathrm{m}, \mathrm{v})$, with m and v are given below.

$$m = log\left(\frac{1}{\sqrt{\alpha+1}}\right), and\ v = log\left(\frac{\alpha}{\sqrt{2}}\right), \quad (17)$$

For each of those scenarios, we propose an estimator that would be the most efficient given the error distribution and the variance process assumption. In fact, the estimators of scenarios 1 and 3 would be the first choice if y has constant variance (i.e η has an inversely proportional variance to the square of $\exp(x\beta)$). While, estimators of scenarios 2 and 4 would the best choice if η ($i.e. \frac{y}{\exp(x\beta)}$) has a constant variance.

Given that the previous four estimators assume constant variance, we propose a new set of estimators, where the variance process is endogenously estimated in the probability density function. To do so, we assume that the variance has the following functional form: [16]

$$\sigma^2 = |\omega_0| + |\omega_1| \exp(x\beta) + |\omega_2|[\exp(x\beta)]^2 \quad (18)$$

Therefore, β and ω are simultaneously estimated. To the best of our knowledge, this is a novel contribution.

Table 2 describes the two-equation estimators (ET-Tobit, standard and New Heckman estimators).

Table 2. Two-equation estimators

	Notation
Model	
$y = \begin{cases} \exp(x\beta)\,\eta - V, if\ \ y + V > 0 \\ 0, if\ \ y + V \leq 0 \end{cases}$	ETtobit (all versions)
$\begin{cases} y = y_1,\ y_2 > 0 & y_1 = \exp(x\beta)\eta \\ y = 0,\ y_2 \leq 0 & y_2 = \quad x\alpha + \nu \end{cases}$	Heckman (all versions)
$\begin{cases} y = y_1,\ y_2 > 0 & y_1 = exp(x\beta)\eta \\ y = 0,\ y_2 \leq 0 & y_2 = exp(x\beta) - k + \alpha z + \mu \end{cases}$	New Heckman (all versions)

[15] For ETobit case, we need to add the threshold the parameter V. i.e: ($(y + V) \sim N(\exp(x\beta), \frac{1}{\alpha})$)

[16] For scenarios 2 and 4, $Var = 1/|\omega_0| + 1/[|\omega_1| \exp(x\beta)] + 1/[|\omega_2|[\exp(x\beta)]^2]$

Estimated Probability Function[17]	
$Normal\left(\frac{\mathrm{y}-\exp(x\beta)}{\sigma}\right)$	heckmanml, heckmanml_New, ETtobit
$Nomral\left(\frac{\frac{y}{\exp(x\beta)}-1}{\sigma}\right)$	n_heckmanml, n_heckmanml_New, n_ETtobit
$Lognormal\left(\frac{\mathrm{y}-\exp(x\beta)}{\sigma}\right)$	heckmanml_L, heckmanml_L_New, ETtobit_L
$Normal\left(\frac{\mathrm{Log(y)}-x\beta-m}{\sigma}\right)$	heckmanml_LOG, heckmanml_LOG_New, ETtobit_LOG

Each estimator has a Generalized version (G_heckman, G_heckmanml_New…) where the variance is endogenously estimated (eq(18)).

4.2 Computational challenges

Note that it was challenging to estimate those versions since standard econometrics software do not allow to estimate nonlinear versions whether in the measure equation or selection equation. Moreover, the estimation of the variance process makes the computations more challenging. Our simulations are conducted using MATLAB. The *fminsearch* and *fminunc* commands were particularly helpful. *fminsearch* is very efficient in getting good starting values for the estimations when minimizing a non-smooth nonlinear objective function. Once, we get good starting values (especially for the variance process and selection equation: αs and γs), we use *fminunc* to reach the global minimum. Then we use a combination of *fminsearch* and *fminunc* to make sure that we reached the global minimum (usually, we do not need to run the loop more than one time).

Another challenge is how to estimate the standard errors. We use the scoring method proposed by Berndt, Hall, Hall, and Hausman's in 1974 (BHHH), for the maximization of log-likelihood functions. This method consists in using the cross product of the matrix of first derivatives to estimate the Hessian matrix.

$$VAR = (G^T G)^{-1} \qquad (19)$$

when G is the gradient matrix (numerically computed). When using this method to estimate the standard errors for standard ET-Tobit, we get very similar results (standard errors) to the ones reports in STATA, which increases the credibility of the BHHH technique used in this paper to report standards errors of the estimated coefficients

[17] For the ET_tobit estimators versions, we need to add the threshold parameter V to y. For instance, for ETtobit, the estimated probability function is:

$$Normal\left(\frac{\mathrm{y}+\mathrm{V}-\exp(x\beta)}{\sigma}\right)$$

5. Simulations

5.1 Data Generating Processes

We conduct different sets of experiments to allow for different data generating processes (DGP). In fact, we consider three dimensions in our data generating process, the distribution of the errors (normal vs. lognormal) and the censuring frequency (different values of k), and different patterns of the heteroscedasticity (5 cases). The multiplicative model adopted by SST (2006) which consists in the following specification:

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

was used for all experiments. x_{1i} is a standard normal variable designed to capture the behavior of continuous explanatory variables such as income levels and distance while x_{2i} is a binary dummy that equals 1 with probability of 0.4 and designed to capture dummy variables such as boarder and free trade agreements. Any set of observations of all variables is generated in each replication using $\beta_0 = 0$, $\beta_1 = \beta_2 = 1$. Data on bilateral trade, y , are generated as

$$y_i = \mu(x_i\beta)\eta_i, \qquad (21)$$

where η_i is a lognormal random (and normal) error with mean 1 and variance σ_i^2.

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

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

5.1.1 First Data Generating

In the First Data Generating (FDG) process we incorporated a negative constant term, $-k$, to ensure the prevalence of a significant number of zero trade values. That is,

$$y_i = \exp(x_i\beta)\,\eta_i - k \,, \quad (22)$$

This is the data generating process underlying the Eaton and Tamura (1994) estimator, which ensures that the trade actually occurs only if a threshold level of potential trade is exceeded. In our simulations, we consider all possible combinations between the values of $k = (0, 1)$ and the distribution of the error terms (normal vs. lognormal).

The equation (22) is equivalent to

$$y_i = \exp(x\beta) - k + \varepsilon_i$$

5.1.2 Second Data Generating

In the Second Data Generating (SDG) process, data is generated using the following specifications.

$$y_1 = \exp(x_1\beta)\eta$$

$$y_2 = \exp(x_1\beta) - k + \mu$$

We also allow for correlation between the error terms of the selection and correction equations. That is, $corr(\eta, \varepsilon) = \rho$. Depending on the model specification (single equations, ETobit, Standard Heckman and New Heckman), different values for k are considered; $k = 0, 0.5, 1, 1.25$.

5.1.3 Third Data Generating

In the Third Data Generating (TDG) process we have

$$y_1 = \exp(x_1\beta)\eta$$

$$y_2 = \exp(x_1\beta) - k + z\alpha + \mu$$

As in SDG, we allow for correlation between the error terms of the selection and correction equations. That is, $corr(\eta, \varepsilon) = \rho$. We consider only one value of $k = 1$. We conduct simulations for the newly introduced Hackman equations only.

5.2 Simulation Results

5.2.1 FDG

Single Equations

Table A1-1 and Table A1-2 in the Appendix summarize the simulation results of the FDG process of single equations when the estimation errors follow normal and lognormal distributions, respectively. For every case of the variance process we report the bias and the standard error of the estimators of the two parameters of interest (β_1 and β_2) obtained using the different estimation techniques.

Our first estimation task was to match the results of cases 1 to 4 of the experiments conducted by SST (2006) and to confirm their findings in relation to the superiority of PPML. Our findings are indeed consistent with SST (2006). When $k = 1$, all estimators are biased, except the PPML for lognormal errors (all cases of the variance process), and the new PPML that we introduced, Optimal_PPML for normal errors (all cases of the variance process). Note, however, that although these two estimators are relatively good, they are not very efficient. The poor performance of the other estimators (i.e., high bias) is due to their misspecifications. In fact, the models used to generate those estimators (OLS, NLS, Heckman_2SLS, etc.) do not assume any k in the gravity equation. Note that this issue becomes more pronounced when the estimation requires a log transformation (OLS, Heckman_2sls_Log, etc.) as the biases reach high levels.

When $k = 0$, many estimators do very well, especially in case of lognormal (Gheckmans2sls, Gnnls, Gnnls_C, PPML, PPML_C). Moreover, for the lognormal case, when $k = 0$, the censored and non-censored NLS and PPML (i.e. NLS vs. NLS_C and PPML vs. PPML_C) are exactly the same. For the case of normal errors, generalized estimators (Gheckman2sls, heckman2sls_Log, and Gheckman2sls_Log) are the best estimators. It is worth noting that in this case (normal distribution), even for $k = 0$, a degree of censoring is needed as $y = \exp(x\beta) + \varepsilon$ can be negative (in this case, y =0).[18]

Three takeaways from the single equations experiments. First, none of the estimators is robust to the censoring. Second, OLS yields consistent results only in the case of constant variance ($[y_i|x] = 1$) when $k = 0$. Moreover, since $\log(\eta)$ has a constant variance, the only bias will be on β_0 (not on $\beta_1$and β_2). Third, and most importantly, the new PPML estimator we introduced, Optimal_PPML, outperforms SST (2006) PPML estimator for normal errors where $k = 1$.

Two-equation (Maximum likelihood) Techniques

As discussed earlier, our two-equation techniques are based on different methods to determine the measure and the selection equations.

Eaton-Tamura Tobit (ET-Tobit)

We estimate the Eaton-Tamura Tobit (ET-Tobit) model using different specifications of the distribution of error term. Table A2-1 and Table A2-2 in the Appendix report the results of the ET-Tobit estimators with the different variance scenarios described above. As can be seen, the E_ET-Tobit is perfectly consistent with the FDG. In particular, it yields consistent estimators for both $k = 0$ and $k = 1$ in the case of constant variance (case 1) and errors normally distributed. When error terms are normally distributed, GET-Tobit and Gn_ET-Tobit are the most efficient estimators for both $k = 0$ and $k = 1$, and they are robust for the other specifications of the variance process (cases 2-5). Table 3 below indicates the efficient ET_Toibit estimators for the different patterns of heteroscedasticity.

Table 3: Most Efficient Estimators of the FDG ET-Tobit Models ($k = 0$, $k = 1$)

	Case 1	Case 2	Case 3	Case 4	Case 5

[18] This is because if the error term is a large negative number with absolute value bigger than $\exp(x\beta)$ then y will be negative.

Normal	E_ET-Tobit, GETtobit & Gn_ETtobit	GET-Tobit & Gn_ET-Tobit	GET-Tobit & Gn_ET-Tobit	GET-Tobit* & Gn_ET-Tobit*	GET-Tobit & Gn_ET-Tobit
Log-normal	GET-Tobit_L & GET-Tobit_LOG	G_ET-Tobit_L & GET-Tobit_LOG	ETtobit_LOG, GET-Tobit_L & GET-Tobit_LOG	GET-Tobit_L* & G_ET-Tobit_LOG*	GET-Tobit_L & GET-Tobit_LOG

* Efficient for β_1 $but\ not\ for\ \beta_2$.

When the errors are log-normally distributed, G_ET-Tobit_L and G_ET-Tobit_LOG yield the most efficient estimators for all specifications of the variance process. Note that for the 4th case, β_2 estimator is more efficient using the ET-Tobit and GETtobit model, for the normal and lognormal distributions, respectively.

Standard Heckman Model

As highlighted above, we estimate two versions of the Heckman model, standard and new. In both versions we have $\mathrm{y}_1 = \exp(\mathrm{x}\beta)\eta$ (or $\mathrm{y}_1 = \exp((\mathrm{x}_1\beta) + \mathrm{u})$.

In the standard Heckman model, $y_2 = x_1\alpha + \epsilon$; and

$$y = \begin{cases} y_1, & y_2 > 0 \\ 0, & y_2 \leq 0 \end{cases}$$

Table A3-1 and Table A3-2 in the Appendix report the simulation results for the standard Heckman model. When $k = 0$, Gheckmanml and Gn_heckmanml yield very efficient estimators for the normal error case while Gn_heckmanml, Gheckmanml_L and Ghecmanml_LOG are the most efficient estimators when the errors follow lognormal distribution. Two exceptions are to highlight in each type of distributions of the error terms. For the normal errors, in the 4th specification of the heteroscedasticity the Gheckmanml and Gn_heckmanml are not the best estimators for β_2 as the bias is quite significant. Similarly, Gn_heckmanml, Gheckmanml_L and Ghecmanml_LOG are not the best estimators for β_2 in the 4th specification of the heteroscedasticity. Table 4 below summarizes the most efficient estimators of the standard Heckman equations in the FDG exercise.

Table 4: Most Efficient Estimators of the FDG Heckman Models ($k = 0$)

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**
Normal	Gheckmanml & Gn_heckmanml	Gheckmanml & Gn_heckmanml & heckmanml	Gheckmanml & Gn_heckmanml & n_heckmanml & heckmanml_LOG	heckmanml & Gheckmanml*, Gn_heckmanml* & Gheckmanml_L, Gheckmanml_LOG	Gheckmanml & Gn_heckmanml
Log-normal	Gn_heckmanml, Gheckmanml_L & Ghecmanml_LOG & heckmanml	Gn_heckmanml, Gheckmanml_L & Ghecmanml_LOG & heckmanml	Gn_heckmanml, Gheckmanml_L & Ghecmanml_LOG, & heckmanml	heckmanml, & Gheckmanml*, Gn_heckmanml*, Gheckmanml_L* & Ghecmanml_LOG*	Gn_heckmanml, Gheckmanml_L & Ghecmanml_LOG & heckmanml

* Efficient for β_1 *but not for* β_2 .

When $k = 1$, for all patterns of heteroscedasticity, none of the estimators is efficient under the assumption of normally distributed errors. But, when the errors are lognormally distributed, heckmanml_L, Gheckmanml_L and Gheckmanml_LOG for case 2 and n_heckmanml for case 3 are relatively efficient.

Like E_ET-Tobit, generalized versions of the Heckman model of the same error type yield the same estimators. More generally, generalized versions of the maximum likelihood estimators are very successful in dealing with the heteroscedasticity in its different forms (except for β_2 with the 4th pattern of the variance process).

New Heckman Model

In the new Heckman model $y_2 = \exp(x_1\beta) - k + z\alpha + \mu$, and,

$$y = \begin{cases} y_1, & y_2 > 0 \\ 0, & y_2 \leq 0 \end{cases}$$

Table A4 in the Appendix reports the simulation results of the New Heckman model. Independently of the distribution of the error terms or the heteroscedasticity process, none of the Heckman models yield good estimator. However, G_nheckmanml_New yields estimator with relatively acceptable bias for the lognormal errors case.

5.2.2 Second data generating Process

In the SDG process we have

$$y_1 = \exp(x_1\beta)\eta$$

$$y_2 = \exp(x_1\beta) - k + \mu$$

In the SDG process we allow for correlation between the error terms of the selection and correction equations. That is, $corr(\eta, \mu) = \rho$.

We also consider different values of $k = 0.5$ and $k = 1.25$

Single Equations

Depending on the pattern of heteroscedasticity and on the distribution of the error terms, some estimators seem to be more efficient than others are. For the normally distributed errors (Table A5-1 in Appendix), Heckman_2SLS is the most efficient for all patterns of heteroscedasticity with small bias and robust to the degree of censoring. For lognormally distributed errors (Table A5-2 in Appendix), however, Heckman_2SLS is the most efficient for the first and second patterns of heteroscedasticity only. For heteroscedasticity patterns 3, 4 and 5, NNLS estimators are reasonably good (small bias) but have large standard deviations. Note that the generalized versions do not improve the results. In contrary, it could worsen them. Therefore, when the censuring process is not well captured, considering different specifications of the variance process does not seem to ameliorate the quality of the estimators. Note also that heckman2SLS_LOG and Gheckman2SLS_LOG yield smaller variance for cases 2, 3 & 5. It is important to note as well that the PPML introduced by SST (2006) and the new PPML we introduced in this paper yield biased estimators. However, the censored PPML yields a slightly better results than the uncensored (complete data) counterpart while the new PPML is much better for the uncensored data than for the censored case. In other words, when working with censored data SST (2006) PPML outperforms the newly introduced PPML but when complete data is considered, the new PPML yields better estimation results.

Two-equation (maximum Likelihood) Techniques

ET-Tobit Equations

Results are reported in Table A6 in Appendix. As expected, the E_ET-Tobit estimators yield substantial bias. Except for the 3rd specification of the heteroscedasticity (constant variance), GETobit and Gn_ET-Tobit (for both type of error distributions, normal and lognormal) where the estimators are quite efficient, all other estimators yield very bad results.

Standard Heckman Equations

Results are reported in Table A7 in Appendix. Many estimators did reasonably well in terms of efficiency and standard deviations. When the errors are normally distributed, Gn_heckmanml is the best estimator in cases 1, 2 and 3 but none of the estimators is efficient for the variance processes described by cases 4 and 5. For the lognormal errors, however, Gheckmanml_L is the best estimator for case 1 and heckmanml_LOG is the best estimator for case 2, but no efficient estimators can be obtained for cases 3, 4 and 5. Note that although heckman_LOG model yields the best results, it has relatively high variance. It is important to note, however, that despite its superiority relative to other estimators, heckman_Log deteriorates the estimations, like the FDG case.[19]

Table 5: Efficient Standard Heckman Estimators

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**
Normal	Gn_heckmanml	Gn_heckmanml	Gn_heckmanml	None	none
Log-normal	Gheckmanml_L	heckmanml_LOG			

[19] In fact, in our simulations results of FDG process, when $k = 1$, heckman_Log yields widely biased estimators.

To deal with the shortcomings of this estimator, we introduce a new version of the Heckman estimator.

New Heckman Equations

Under the new Heckman specification, data is generated as $y_2 = \exp(x\beta) - k + z\alpha + \mu$. When errors are lognormal (Table 8B), regardless of the value of k (i.e., degree of censoring), Gheckmanml_L_New and Gheckmanml_LOG_New yield the best estimators for all patterns of heteroscedasticity, except for case 1 where heckmanml_New and Gn_heckmanml_New yield the best estimators. With high degree of censoring (k=1.25), however, heckmanml_New and heckmanml_LOG_New yield even better estimators. Again Heckman_Log gives very good results for high censoring data (k=1.25), but less accurate results for k = 0.5. Results are reported in Table A8-1 & Table A8-2 in the Appendix. Table 6 below summarizes the most efficient estimators.

Table 6: Efficient New Heckman Estimators

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**
Normal (k=0.5)	Gheckmanml_New, Gn_heckmanml_New, heckmanml_New	Gheckmanml_New, Gn_heckmanml_New	Gheckmanml_New	Gheckmanml_New	Gheckmanml_New, Gn_heckmanml_New
Log-normal	heckmanml_New & Gn_heckmanml_New	Gheckmanml_L_New & Gheckmanml_LOG_New	Gheckmanml_L_New & Gheckmanml_LOG_New	Gheckmanml_L_New & Gheckmanml_LOG_New	Gheckmanml_L_New & Gheckmanml_LOG_New

For normal errors (Table A8-1), with low degree of censoring (k=0.5), Gheckmanml_New is the only successful estimator for all cases of heteroscedasticity. Gn_heckmanml_New is efficient for cases 1, 2 & 5 while heckmanml_New is relatively efficient for case 1 only. For high censoring (k=1.25), however, only the 1st and 2nd cases of the variance process result in successful estimators. Gn_hechamnml_New are not very successful with high censoring (k=1.25) as compared to low censoring (k=0.5) especially for the 3rd, 4th and 5th specifications of the variance process. Similarly, Gheckmanml_New yield relatively good results with low degree of censoring (k=0.5), yet relatively higher bias for the 3rd, 4th and 5th cases of heteroscedasticity.

It is interesting to note that G_heckmanml_New and Gn_heckmanml_New do not always coincide. This could be explained by the convergence problems. In fact, it seems that when data is simulated from normal errors, convergence issue can occur when using G_heckmanml_New and Gn_heckmanml_New. Yet, this problem does not exist for Gheckmanml_L_New and Ghemanml_LOG_New even for the normal error case.

According to FDG and SDG results Heckmanml_Log efficiency is not robust to the censoring rate. Hence, we cannot really rely on Heckmanml_Log results. Indeed, when there is a significant difference between Heckmanml_Log and generalized versions, we should take the latter ones.

5.2.3 Third data generating Process

Results are reported in Table A9 in the Appendix. In the TDG process we run the simulations for the New Heckman specification only. Data is generated according to the following measure and selection equations:

$$y_1 = \exp(x_1\beta)\eta$$

$$y_2 = \exp(x_1\beta) + x_2\alpha + \mu$$

As in SDG, we allow for correlation between the error terms of the selection and correction equations. That is, $corr(\eta, \mu) = \rho$. We consider only one value of $k = 1$.

For lognormal errors, Gheckmanml_L_New and Ghemanml_LOG_New give the best (and very similar) results, for the 2nd, 3rd and 5th patterns of heteroscedasticity. Moreover, both estimators give relatively good results for case 1 but they are not the most efficient (as compared to heckmanml_New, Gheckmanml_New and Gn_heckmanml_New). For the 4th pattern of heteroscedasticity, heckmanml_LOG_New is the most efficient while Gheckmanml_L_New and Ghemanml_LOG_New are reasonably good estimators for β_1 but not β_2.

Table 7: TDG process - Efficient New Heckman Estimators

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**
Normal	G_heckmanml_New & Gn_hechamnml_New	G_heckmanml_New & Gn_hechamnml_New	G_heckmanml_New & Gn_hechamnml_New	G_heckmanml_New* & Gn_hechamnml_New*	G_heckmanml_New & Gn_hechamnml_New
Log-normal	heckmanml_New, Gheckmanml_New & Gn_heckmanml_New	Gheckmanml_L_New & Ghemanml_LOG_New	Gheckmanml_L_New & Ghemanml_LOG_New	heckmanml_LOG_New	Gheckmanml_L_New & Ghemanml_LOG_New

* Estimator of β_1 only.

For normal errors, G_heckmanml_New and Gn_hechamnml_New yield good results for all patterns of heteroscedasticity (only estimator of β_1 for case 4), but not as good as Heckmanml_L_New and Gchemanml_LOG_New for the lognromal errors case. Note also that the problem of convergence arises here again for the 3rd and 4th specifications of the variance process. Another important remark here is related to the estimator heckmanml_New characterized with unstable efficiency with high bias for the 3rd and 5th patterns of heteroscedasticity.

Based on the above simulations results, five important remarks worth highlighting. ***First***, for both the SDG and TDG processes, Gheckman_L_New and Gheman_LOG_New are efficient for lognormal errors. Yet, it is obvious that those results are very good for heteroscedasticity of type 1 and 2 (regardless of the censoring rate). For heteroscedasticity of type 3, 4 and 5, they still yield good results but less efficient than the heteroscedasticity of type 1 & 2. ***Second***, for all data generating processes, we notice that for almost all maximum likelihood estimators (i.e. two equation models), there is a high bias for the experiments where the heteroscedasticity is of type 4. ***Third***, as mentioned by, the variance is most likely to be proportional to the trade level; that is, type 2. ***Fourth***, for the single equations estimators, when the data is non-censored (k=0) and lognormal errors, the PPML estimator is very outstandingly efficient, which is consistent with the finding of SST (2006). Which is also the case for GNLS and Gheckman_2SLS. But, the non-censored data with normal errors, PPML yield large bias. ***Fifth***, overall, Gheckmanml_L_New and Gheckmanml_LOG_New seem to be the best estimators when errors are lognormally distributed.

6. New Estimates of the Gravity Equation (Results)

In this section, we use the most efficient estimators identified in the simulation exercise to quantify the effect of the conventional variables used in gravity models on the bilateral trade flows between countries. We compare our estimators with the predictions related to the same variables and the main contributions in the literature, namely SST (2006). Two estimation techniques are performed: single equations and two equations. In each technique, we consider the various specifications discussed earlier. This allows us to compare the predictions of models (ie., our estimates) with the predictions of other models who used different approaches to deal with heteroscedasticity and the prevalence of zero trade values. We structure this section into three subsections. We start with a brief description of the data, then we discuss the single equations results and we conclude with the two equations findings.

6.1. Data

As our model builds on the findings of important contributions in the literature and proposes refinements of existent estimators, we thought that using the database of SST (2006), the most significant contribution in the literature dealing with heteroscedasticity and the prevalence of zero trade values, is essential to validate our results and our comparative analysis. The data consists of a cross section bilateral export flows of 136 countries in 1990. A detailed description of the data, including the list of countries, variables, sources, etc. can be found in SST (2006). Briefly, the dependent variable consists of the 'bilateral exports' while explanatory variables are 'real GDP per capita', 'population', 'location', 'distance', 'remoteness', 'preferential trade agreements'. Moreover, the explanatory variables includes a set of dummy variables constructed to capture 'contiguity', 'common language', 'colonial ties', and 'access to water'.[20]

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

6.2 Single equations results

Table 8 presents the results of the traditional estimators widely used in the literature and discussed in the methodology and simulation sections. For instance, we report the estimates of the OLS model excluding the pairs of countries with zero bilateral trade. For the NLS, GNLS, and PPML are estimated using the whole sample (including the pairs of countries with zero bilateral trade values) while NLS_C, GNLS_C, and PPML_C represent the censored version of the corresponding models.

It is important to note that our findings are perfectly consistent with SST (2006) results. In particular, our estimates match both the whole sample and subsample PPML estimators. We were also able to find the same results for the other linear and nonlinear estimators. The same conclusions of SST (2006) are valid. In fact, compared with other estimators, the PPML estimator deals better with the heteroscedasticity. Moreover, the PPML estimates indicate that the importer's and exporter's GDP explain just above 70 percent of trade between countries, as opposed to higher values that other models suggest (e.g., OLS GDP estimated elasticities are above 90% and around 80% for the exporter's GDP and importer's GDP, respectively). Estimates of the other variables such as distance, remoteness, access to water and preferential-trade agreements are also consistent with the findings of SST (2006) and same conclusions apply. That is, the PPML estimates are different from other models as it deals with the heteroscedasticity and zero trade values differently.

[Table 8-1 here]

Our second exercise related to the single equations techniques consists of the estimation of the new set of estimators that we introduced in this paper. Table 8-2 reports the outcome of the estimates of the improved versions (whole sample and subsample of positive trade values only) of the Poisson model (Optimal_PPML and Optoimal_PPML_C), two versions of the standard two-stage least square regressions (2SLS) of Heckman model (one in level, Heckman_2SLS, and one in log, Heckman_Log_2SLS) and two generalized versions of the 2SLS Heckman model (GHeckman_2SLS and GHeckman_Log_2SLS).

[Table 8-2 here]

The first important observation is that, except for the two log versions of the two-stage Heckman model (Heckman_Log_2SLS and GHeckman_Log_2SLS), the estimates of the key parameters are relatively close. For instance, estimates of the log of exporter's GDP ranges from 0.6326 to .7376 and the estimates of the log of importer's GDP are between .6619 and .8617. The second observation is that the estimated coefficients of the improved Poisson models (Optimal_PPML and Optimal_PPML_C) that we introduced, although close to the PPML estimators of SST (2006) are lower for the key parameters. For example, the estimated coefficients of the log exporter's GDP and log importer's GDP are around 73% in the standard PPML while they are around 65% in the PPML improved versions. This suggests that exporter's and importer's GDP in the improved Poisson model have less weight than in the standard Poisson coefficients. Other variables such as the distance, exporter's and importer's per capita GDP, common language, contiguity, landlocked exporter and importer, remoteness, and preferential-trade agreements continue to be significant and with relatively same importance in the improved versions of the PPML estimators. Despite the differences in the weights of the coefficients estimates, our findings confirm the power, robustness and superiority of the Poisson estimator proposed by SST (2006) when the bilateral trade is estimated using single equations techniques.

Finally, we notice that the estimations of Heckman_Log _2SLS yield higher coefficients compared to the Heckman_level_2SLS. This may be due to the Jenson inequality bias (variance depending on the explanatory variables) faced when using Heckman_Log_2SLS.

Explanation of the results similarity between PPML and Truncated_PPML

If we look at the data used in SST(2006) and in the present paper, we will notice that the real trade values, once sorted, have an exponential shape (where half of non zeros trade observations are less than 1% of the mean value). It is also obvious to notice that the Log exporter's GDP and Log importer's GDP are the main two variables that contribute the most in explaining this exponential shape of trade data, as they are the variables with highest mean values and the highest variances out of all significant variables. The mean values of two key variables are around 22 for the zero trade observations, while they are around 28 for the right upper tail of the distribution (sorted as function of trade). This means that zeros are highly correlated with the main variables that can match the exponential distribution of trade. The fact that estimating the gravity equation would consist in mainly matching the upper part of the trade exponential distribution, including zeros in the estimation will not change much the numerical results, and the biases discussed in subsection 3.2 (systematic and correlation biases) will be significantly attenuated. This explains the similarity of estimations results of truncated and non-truncated data. It is worthy to note that this would not be the case if we have to deal with a data set where zeros trade values are not positively correlated with x (zeros trade explained by other omitted variables), or if the minimum non-zero trade value is high.

We can conclude that zeros trade values should be included in the estimation of the gravity model, but we should be careful when dealing with them. In fact, we cannot be assured that we will always be avoiding the biases discusses above.

6.3 Two-equations results

Our two equations techniques consisted in estimating the bilateral trade between countries using three methods: two conventional approaches (the ET-Tobit & the standard Heckman models) and a new approach of the Heckman model that we introduced early. In each of these three methods, we add to the body of the existent literature with new specifications of the selection equation, the variance process and the distribution of the error terms. In fact, we consider nonlinear form of the selection equation, lognormally distributed errors and we estimate the variance process simultaneously. In what follows, we discuss the results of standard and new Heckman models.[21]

In the Heckman model, we used three selection equation specifications:

i. Standard, where same set of variables used in measure equation are used in the selection equation in a linear manner. That is, $y_2 = x\alpha + \mu$,
ii. New Heckman: $y_2 = \exp(x_1\beta) - k + z\alpha + \mu$, where β is endogenously estimated in the measure equation, and
iii. New Heckman without common language variable in the measure equation but it is in the selection equation. [22]

Table 9 reports the results related to the efficient estimates of the standard Heckman model. The log likelihood functions confirm the findings of the simulations exercise and suggest that Gehckmanml_Log and Gehckmanml_L are the most efficient. These two maximum likelihood estimators suggest that the coefficients on exporter GDP and importer GDP are 0.50 and 0.44, respectively, estimates that are significantly lower that their counterparts in most studies in the literature, which are close to one.

Note that it is not surprising that these two estimators yield very similar coefficients estimates, as they are asymptotically equivalent, suggesting that our heteroscedasticity modelling is consistent in both cases. [23]

[Table 9 here]

Compared with the PPML estimates for the key variables, which are substantially below one (0.733 and 0.741 for exporter and importer GDP, respectively), the standard Heckman models we estimate provide lower coefficients (0.503 and 0.44). Furthermore, considering our improved PPML estimates (0.64 and 0.66 for the exporter and importer GDP, respectively), even though the differences with the standard Heckman estimates are smaller, they remain significant. Similarly, for all other explanatory variables, standard Heckman model provides lower estimates, compared to PPML and the improved PPML. The more significant differences, though, are related to (exporter's and importer's) remoteness, landlocked (exporter and importer) and common language variables. These results indicate that the bias of PPML estimator is significant. The improved

[21] Results of all versions of the ET-Tobit model are available upon request. Note that for the standard ET-Tobit model (i.e., estimation of the $\ln(a + T_{ij})$ model), we get the same results as SST (2006), especially for the key variables (with estimated coefficients of 1.0587 and 0.8475 for the log of exporter's GDP and log of importer's GDP, respectively), but our generalized versions give less weight for these variables (with estimated coefficients of .8186 and 0.6637 for the log of exporter's GDP and log of importer's GDP, respectively).

[22] Estimates of the New Heckman model without the common-language variable are available upon request.

[23] The differences between the coefficients of the key variables (i.e., exporter GDP and importer GDP) are insignificant (0.0012 for exporter GDP and -0.0001 for importer GDP).

PPML estimator, even though has smaller difference with the standard Heckman estimator, its bias remains important.

Table 10-1 and Table 10-2 present the results of the regular and generalized versions of the New Heckman model, respectively. The likelihood ratio test confirms the superiority of the generalized versions of the new Heckman model. In fact, based on LR tests, for all of the estimators considered, the generalized versions yield the most efficient results, more specifically Gheckmanml_Log_new and Gheckmanml_L_new. Note, however, that despite the differences in performance of the different estimators, the results are very close and report significantly lower coefficients of the key variables, namely exporter and importer GDP (around 0.5 and 0.4, respectively). It is important to note here that, even though the new Heckman models are estimated in log, our proposed sample selection equation corrects for the bias caused by the heteroscedasticity issue (i.e., Jensen's inequality problem). This represents, indeed, one of our main contributions and consists in a novel finding in the literature as previous empirical studies results might be overestimating the contribution of the GDPs of both import and export countries in determining the bilateral trade.

[Table 10-1 here]

[Table 10-2 here]

7. Conclusion

Estimation of trade flows between countries using gravity equations has been used extensively in the international trade literature. The prevalence of zero trade values and the presence of heteroscedasticity, however, raised many questions related to many specifications and estimation methods. Indeed, the way the log-linearized models treated the zero trade values and the approaches considered to handle the heteroscedasticity problems have led to efficiency and optimality issues. This paper proposes corrections for biases identified in some of the estimators used in the literature. We revisit the theoretical foundations and estimation practices of many specifications of the gravity equation and show that improving the consistency and efficiency of the estimates is possible.. Our contribution is threefold. First, we derive an improved optimal version of the well known the Poisson Pseudo Maximum Likelihood (PPML) estimator. Although the PPML is robust to different patterns of heteroscedasticity, deals with the Jensen's inequality, resolves the inefficiency problem, and tractability in the international trade literature, some issues related to the zero trade values and heteroscedasticity persist. In fact, using all data (i.e., including the zeros) in the estimation sample forces the estimated gravity model to predict a trade level that should be as close as possible to zero. But, zero trade values does not necessarily mean that the absence of potential for trade. Moreover, the derivation of the PPML estimator assumes that the variance is proportional to the mean, which may raise the questions about its optimality. While we start from the same assumption made by SST (2006), namely, the conditional variance propose is proportional to the mean, we derive a new FOC from a modified version of the objective function where error terms are divided by the corresponding standard errors, resulting in a new estimator. Second, we propose various Heckman estimators based on a new specification of both the measure and selection equations, different distributions of the error terms and various processes of the variance. We add to the existent literature alternative estimation methods taking into account the non-linearity of both the variance and selection equations, allow for lognormal distribution of the errors terms, and specify the selection equation in way it determines whether a bilateral trade can occur or not and treat the trade frictions differently. The considerations we included in our estimations led us to conclude that previous empirical studies might be overestimating the contribution of the GDP of both import and export countries in determining the bilateral trade. In fact, the estimates of the coefficients on the income of exporting and importing countries are more than 0.7 in the existent literature while our model predicts suggests that those estimates are in the range of 0.4 to 0.5.. Third, to perform the estimation of the above-mentioned Heckman versions (non-linear forms, lognormal distribution and simultaneous variance estimation) computational and technical challenges arise. In fact, no pre-set package in the econometric software (Stata, Eviews, etc.) can deal with these three dimension to estimate our augmented model. We had to code in Matlab using a combination of *fminsearch* and *fminunc* functions. Moreover, using numerical gradient matrix G, we report standard errors based on the BHHH technique.

Table 8-1: Single Equations Traditional Estimators

	OLS_C		NLS			GNLS		NLS_C			GNLS_C		PPML		PPML_C	
	coeficient	*s.e*	*coeficient*	*s.e*	*robust*	*coeficient*	*s.e*	*coeficient*	*s.e*	*robust*	*coeficient*	*s.e*	*coef*	*s.e*	*coef*	*s.e*
Constant	-28.4920	1.0880	-45.0989	0.2391	3.3792	-31.1494	0.4411	-45.0850	0.3306	3.3832	-23.5902	0.7634	-32.3261	2.0595	-31.5296	2.1610
Log exporter's GDP	0.9378	0.0116	0.7378	0.0044	0.0384	0.7204	0.0060	0.7376	0.0060	0.0384	0.8045	0.0101	0.7325	0.0268	0.7213	0.0268
Log importer's GDP	0.7978	0.0111	0.8619	0.0045	0.0410	0.7619	0.0060	0.8617	0.0062	0.0410	0.6708	0.0099	0.7411	0.0274	0.7319	0.0279
Log exporter's per capita GDP	0.2073	0.0166	0.3957	0.0097	0.1157	0.1996	0.0088	0.3953	0.0134	0.1158	0.0805	0.0097	0.1567	0.0533	0.1544	0.0527
Log importer's per capita GDP	0.1061	0.0167	-0.0325	0.0067	0.0619	0.1241	0.0075	-0.0325	0.0092	0.0619	0.3543	0.0125	0.1350	0.0449	0.1327	0.0445
Log of distance	-1.1660	0.0339	-0.9237	0.0085	0.0725	-0.8022	0.0122	-0.9234	0.0117	0.0725	-0.7099	0.0206	-0.7838	0.0546	-0.7763	0.0553
Contiguity dummy	0.3140	0.1425	-0.0813	0.0099	0.1004	0.0769	0.0389	-0.0811	0.0136	0.1005	0.8606	0.0379	0.1929	0.1043	0.2024	0.1052
Common-language dummy	0.6780	0.0640	0.6894	0.0159	0.0850	0.5789	0.0294	0.6894	0.0220	0.0851	1.5307	0.0312	0.7460	0.1347	0.7513	0.1342
Colonial-tie dummy	0.3968	0.0681	0.0358	0.0178	0.1254	0.1283	0.0266	0.0355	0.0246	0.1255	-0.8086	0.0470	0.0250	0.1498	0.0200	0.1500
Landlocked exporter dummy	-0.0620	0.0646	-1.3671	0.0305	0.2022	-0.6364	0.0264	-1.3670	0.0422	0.2023	-1.2966	0.0918	-0.8635	0.1572	-0.8724	0.1573
Landlocked importer dummy	-0.6645	0.0631	-0.4715	0.0223	0.1838	-0.6247	0.0283	-0.4716	0.0308	0.1839	-1.5345	0.0820	-0.6964	0.1408	-0.7035	0.1409
Exporter's remoteness	0.4671	0.0778	1.1878	0.0183	0.1821	0.6070	0.0311	1.1876	0.0252	0.1822	-0.1093	0.0517	0.6598	0.1338	0.6472	0.1352
Importer's remoteness	-0.2050	0.0808	1.0097	0.0179	0.1541	0.4505	0.0329	1.0094	0.0247	0.1542	0.1732	0.0534	0.5615	0.1185	0.5493	0.1197
Preferential-trade agreement dummy	0.4908	0.1053	0.4425	0.0137	0.1090	0.1410	0.0211	0.4426	0.0190	0.1091	-0.6770	0.0539	0.1811	0.0886	0.1794	0.0903
Openness dummy	-0.1696	0.0490	0.9280	0.0238	0.1912	-0.1956	0.0274	0.9270	0.0329	0.1915	-0.8274	0.0378	-0.1068	0.1312	-0.1394	0.1329

Table 8-2: Single Equations New Estimators

	Optimal_PPML			Optimal_PPML_C			Heckman_Log_2s2		GHeckman_Log_2s2		Heckman_2s2			GHeckman_2s2	
	coef	*s.e*	*robust*	*coef*	*s.e*	*robust*	*coef*	*s.e*	*coef*	*s.e*	*coef*	*s.e*	*robust*	*coef*	*s.e*
Constant	-27.4317	2.4322	2.4734	-26.9171	2.9409	2.5984	-33.6901	1.2430	-33.3712	1.2253	-45.0810	0.3307	3.3851	-30.8733	0.6044
Log of exporter's GDP	0.6409	0.0511	0.0404	0.6326	0.0625	0.0399	1.0537	0.0174	1.0485	0.0172	0.7376	0.0060	0.0384	0.7259	0.0083
Log of importer's GDP	0.6682	0.0636	0.0392	0.6619	0.0788	0.0392	0.8781	0.0143	0.8748	0.0141	0.8617	0.0062	0.0410	0.7624	0.0082
Log of exporter's per capita GDP	0.1337	0.0450	0.0801	0.1321	0.0534	0.0785	0.2281	0.0169	0.2269	0.0167	0.3950	0.0134	0.1158	0.1792	0.0118
Log of importer's per capita GDP	0.1594	0.0523	0.0676	0.1562	0.0632	0.0658	0.1309	0.0171	0.1292	0.0169	-0.0325	0.0092	0.0619	0.1142	0.0102
Log of distance	-0.6546	0.0649	0.0618	-0.6492	0.0780	0.0617	-1.2641	0.0360	-1.2603	0.0354	-0.9233	0.0117	0.0725	-0.7996	0.0172
Contiguity dummy	0.3525	0.1065	0.1499	0.3548	0.1288	0.1474	0.1548	0.1455	0.1613	0.1429	-0.0810	0.0136	0.1005	0.0971	0.0537
Common-language dummy	0.9090	0.1465	0.1945	0.9060	0.1750	0.1912	0.7673	0.0653	0.7661	0.0644	0.6894	0.0220	0.0851	0.5193	0.0407
Colonial-tie dummy	-0.2045	0.1591	0.2062	-0.2042	0.1927	0.2035	0.4384	0.0688	0.4345	0.0680	0.0355	0.0246	0.1255	0.1863	0.0356
Landlocked exporter dummy	-0.8249	0.3104	0.1861	-0.8297	0.3674	0.1845	-0.0607	0.0650	-0.0589	0.0643	-1.3667	0.0422	0.2023	-0.6004	0.0366
Landlocked importer dummy	-0.7550	0.2936	0.1842	-0.7561	0.3454	0.1824	-0.6866	0.0636	-0.6876	0.0629	-0.4716	0.0308	0.1839	-0.6064	0.0394
Exporter's remoteness	0.5045	0.2621	0.1725	0.5002	0.3230	0.1728	0.5187	0.0790	0.5143	0.0776	1.1875	0.0252	0.1823	0.5999	0.0423
Emporter's remoteness	0.6126	0.2515	0.1450	0.6061	0.3100	0.1452	-0.2004	0.0817	-0.2062	0.0804	1.0093	0.0247	0.1542	0.4361	0.0445
Preferential-trade agreement dummy	0.1195	0.1578	0.1019	0.1257	0.1923	0.1022	0.4565	0.1076	0.4496	0.1050	0.4426	0.0190	0.1091	0.1556	0.0302
Openness dummy	-0.4671	0.1950	0.1535	-0.4807	0.2362	0.1537	-0.0585	0.0508	-0.0716	0.0504	0.9268	0.0329	0.1915	-0.1828	0.0388

Table 9. Standard Heckman estimators

	Gheckmanml		Gehckmanml_Log		Gehckmanml_L		Gn_Hekcmanml	
	coeficient	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*
Constant	-22.7696	0.1962	-9.8237	0.4992	-9.9481	0.4938	-22.7696	0.1964
Log exporter's GDP	0.5922	0.0025	0.5029	0.0045	0.5017	0.0045	0.5922	0.0025
Log importer's GDP	0.6720	0.0028	0.4411	0.0043	0.4412	0.0042	0.6720	0.0028
Log exporter's per capita GDP	0.2073	0.0019	0.1179	0.0061	0.1172	0.0060	0.2073	0.0019
Log importer's per capita GDP	0.2430	0.0017	0.0895	0.0062	0.0900	0.0061	0.2430	0.0017
Log of distance	-0.8068	0.0054	-0.5935	0.0158	-0.5895	0.0155	-0.8068	0.0054
Contiguity dummy	0.4011	0.0179	0.4682	0.0359	0.4861	0.0351	0.4011	0.0179
Common-language dummy	0.8959	0.0078	0.5033	0.0278	0.5000	0.0274	0.8959	0.0078
Colonial-tie dummy	-0.0015	0.0084	0.0188	0.0290	0.0230	0.0287	-0.0015	0.0084
Landlocked exporter dummy	-0.7246	0.0080	-0.0658	0.0512	-0.0650	0.0511	-0.7246	0.0080
Landlocked importer dummy	-0.6082	0.0080	-0.3100	0.0500	-0.3077	0.0500	-0.6082	0.0080
Exporter's remoteness	0.1947	0.0118	0.2109	0.0408	0.2144	0.0402	0.1947	0.0118
Importer's remoteness	0.4584	0.0098	0.0389	0.0391	0.0496	0.0389	0.4584	0.0098
Preferential-trade agreement dummy	0.0774	0.0161	0.1694	0.0538	0.1746	0.0538	0.0774	0.0161
Openness dummy	-0.6613	0.0057	-0.1386	0.0211	-0.1366	0.0208	-0.6613	0.0056
Log likelihood	**123864.71**		**106948.97**		**106947.95**		**123864.71**	

Table 10-1: Two-equation estimators - New Heckman, regular

	Heckmanml_New		Hekcmanml_L_New		Heckmanml_Log_New		n_Heckman_New	
	coeficient	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*
Constant	-36.2899	0.0837	-7.2588	0.5083	-10.1691	0.6197	-20.3133	0.1328
Log of exporter's GDP	0.6561	0.0012	0.4480	0.0048	0.5391	0.0061	0.5111	0.0012
Log of importer's GDP	0.6826	0.0013	0.3564	0.0047	0.4247	0.0055	0.5190	0.0011
Log of exporter's per capita GDP	0.1994	0.0026	0.0934	0.0064	0.1182	0.0075	0.1685	0.0013
Log of importer's per capita GDP	0.0612	0.0018	0.0830	0.0066	0.0972	0.0076	0.2132	0.0013
Log of distance	-0.7135	0.0025	-0.5018	0.0128	-0.6041	0.0146	-0.6959	0.0035
Contiguity dummy	0.0961	0.0024	0.0834	0.0327	-0.0459	0.0603	0.4635	0.0094
Common-language dummy	0.0130	0.0047	0.2891	0.0233	0.3581	0.0275	0.5878	0.0056
Colonial-tie dummy	0.5438	0.0054	0.1948	0.0256	0.2271	0.0292	0.0543	0.0065
Landlocked exporter dummy	-0.3532	0.0059	-0.0566	0.0233	0.0126	0.0296	-0.4542	0.0056
Landlocked importer dummy	-0.0626	0.0057	-0.1952	0.0250	-0.2236	0.0293	-0.3400	0.0065
Exporter's remoteness	1.0375	0.0057	0.2397	0.0365	0.2355	0.0427	0.4819	0.0087
Emporter's remoteness	0.9609	0.0055	0.0624	0.0355	-0.0826	0.0408	0.4765	0.0071
Preferential-trade agreement dummy	0.5450	0.0045	0.3863	0.0337	1.3830	0.0795	0.5620	0.0099
Openness dummy	0.5480	0.0058	0.1319	0.0198	0.1500	0.0219	-0.2636	0.0036
Log likelihood	**153419.54**		**107706.72**		**108922.00**		**125491.48**	

Table 10-2: Two-equation estimators: New Heckman, generalized.

	Gheckmanml_New		**Gehckmanml_Log_New**		**Gehckmanml_L_New**		**Gn_Hekcmanml_New**	
	coeficient	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*	*coeficient*	*s.e*
Constant	-20.3128	0.1335	-9.2200	0.3747	-9.3684	0.3756	-20.3129	0.1088
Log of exporter's GDP	0.5111	0.0013	0.4760	0.0034	0.4770	0.0034	0.5111	0.0008
Log of importer's GDP	0.5190	0.0012	0.3841	0.0033	0.3848	0.0034	0.5190	0.0007
Log of exporter's per capita GDP	0.1685	0.0013	0.1052	0.0052	0.1049	0.0052	0.1685	0.0013
Log of importer's per capita GDP	0.2132	0.0013	0.0931	0.0052	0.0929	0.0052	0.2132	0.0012
Log of distance	-0.6959	0.0036	-0.5290	0.0118	-0.5324	0.0118	-0.6959	0.0034
Contiguity dummy	0.4635	0.0096	0.2755	0.0294	0.2710	0.0295	0.4635	0.0093
Common-language dummy	0.5878	0.0057	0.3797	0.0215	0.3785	0.0215	0.5878	0.0054
Colonial-tie dummy	0.0543	0.0065	0.1239	0.0226	0.1237	0.0226	0.0543	0.0065
Landlocked exporter dummy	-0.4542	0.0057	-0.0236	0.0268	-0.0237	0.0268	-0.4542	0.0056
Landlocked importer dummy	-0.3401	0.0066	-0.1989	0.0259	-0.1983	0.0259	-0.3401	0.0065
Exporter's remoteness	0.4819	0.0092	0.2402	0.0313	0.2471	0.0313	0.4819	0.0084
Emporter's remoteness	0.4764	0.0072	0.0688	0.0297	0.0776	0.0297	0.4764	0.0070
Preferential-trade agreement dummy	0.5620	0.0135	0.4023	0.0359	0.4017	0.0359	0.5620	0.0096
Openness dummy	-0.2636	0.0037	0.1015	0.0167	0.1035	0.0167	-0.2636	0.0035
Log likelihood	**125491.48**		**107433.23**		**107433.15**		**125491.48**	

References

Anderson, J., 1979. A Theoretical Foundation for the Gravity Equation," American Economic Review volume 69, pp.106–116.

Anderson, J., van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review, volume 93, pp. 170–192.

Arabmazar, A., Schmidt, P. 1981. Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity. Journal of Econometrics, volume 17, issue 2, pp. 253–58.

Arvis, J-F., Shepherd, B., 2013. The Poisson quasi-maximum likelihood estimator: A solution to the "adding up" problem in gravity models. Applied Economics Letters, volume 20, issue 6, pp. 515–519.

Assane, D., Chiang, E. P., 2014. Trade, structural reform, and institutions in Sub-Saharan Africa. Contemporary Economic Policy, volume 32, issue 1, pp. 20–29.

Baldwin, R., Harrigan, J., 2011. Zeros, Quality, and Space: Trade Theory and Trade Evidence. American Economic Journal: Microeconomics, volume 3, issue 2, pages 60-88.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. "Estimation and Inference in Nonlinear Structural Models", Annals of Social Measurement, volume 3, pp. 653-665.

Bosquet, C., Boulhol, H., 2015. What is really puzzling about the "distance puzzle". Review of World Economy, volume 151, issue 1, pp. 1-21.

Burger, M., van Oort F., Linders G-J., 2009. On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. Spatial Economic Analysis, volume 4, issue 2, pp. 167-190.

Chay, K. Y., Powell, J. L., 2001. Semiparametric Censored Regression Models. Journal of Economic Perspectives, Volume 15, issue 4, pp. 29–42.

Dai, M., Yotov, Y. V., Zylkin, T. 2014. On the Trade-diversion Effects of Free Trade Agreements. Economics Letters, volume 122, issue 2, pp. 321–325.

de Sousa, J., 2012 . The currency union effect on trade is decreasing over time. Economics Letters, Volume 117, pp. 917–920.

Duan, N., Manning, W. G., Morris, C. N., Newhouse, J. P., Duan, N., Manning, W. G., Newhouse, J. P., 1983. A Comparison of Alternative Models for the Demand for Medical Care. Journal of Business & Economic Statistics, Vol. 1, No. 2, pp. 115–126.

Egger, P. H., Tarlea, F., 2015. Multi-way clustering estimation of standard errors in gravity models. Economics Letters, Volume 134, pp. 144-147.

Egger, P., Larch, M., 2011. An assessment of the Europe agreements' effects on bilateral trade, GDP, and welfare. European Economic Review, volume 55, pp. 263–279.

Frankel, J. A., Wei, S-J., 1993. Emerging Currency Blocs. NBER Working Paper No. w4335. Available at SSRN: https://ssrn.com/abstract=227026.

Gómez-Herrera, E., 2013. Comparing alternative methods to estimate gravity models of bilateral trade. Empirical Economics, volume 44, Issue 3, pp. 1087–1111.

Gómez-Herrera, E., Baleix, J. M., 2010. Are estimation techniques neutral to estimate gravity equations? An application to the impact of EMU on third countries' exports. The Department of Theory and Economic History, Faculty of Economics and Business. University of Granada, http://www.etsg.org/ETSG2009/papers/gomez.pdf

Head, K., Mayer, T. and Ries, J., 2010. The Erosion of Colonial Trade Linkages After Independence. Journal of International Economics, volume 81, issue 1, pp. 1-14.

Heckman, J. J., 1979. Sample Selection Bias as a Specification Error. Econometric, volume 47, issue 1, pp.153–161.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. The Quarterly Journal of Economics, MIT Press, volume 123, issue 2, pp. 441-487.

https://ideas.repec.org/p/ags/eaa140/163341.html

Hurd, M., 1979. Estimation in truncated samples when there is heteroscedasticity. Journal of Econometrics, Volume 11, Issues 2–3, pp. 247-258.

Jones, A. M., 2000. Health Econometrics. In Culyer A. J., Newhouse, J. P., eds. Handbook of Health Economics, vol. 1, Part A, chapter 6, pp. 265-344. Amsterdam: North- Holland, 2000.

Kareem, F. O., 2014. Modeling and Estimation of Gravity Equation in the Presence of Zero Trade: A Validation of Hypotheses Using Africa's Trade Data.140th Seminar, December 13-15, 2013, Perugia, Italy 163341, European Association of Agricultural Economists.

Krisztin, T., Fischer, M. M., 2015. The Gravity Model for International Trade: Specification and Estimation Issues. Spatial Economic Analysis, volume 10, issue 4, pp. 451–470.

Leung, S. F., Yu, S., 1996. On the choice between sample selection and two-part models. Journal of Econometrics, volume 72 pp. 197-229.

Lin, F., 2013. Are distance effect really a puzzle? Economic Modelling, volume. 31, pp. 684-689.

Liu, X., 2009. GATT/WTO Promotes Trade Strongly: Sample Selection and Model Specification. Review of International Economics, volume 7, issue 3, pp. 428–446.

Martin, W., Pham, C.S., 2015. Estimating the Gravity Model When Zero Trade Flows are Frequent. World Bank Policy Research Working Paper 7308. http://documents.worldbank.org/curated/en/docsearch/document-type/620265

Martínez-Zarzoso, I., 2013. The log of gravity revisited. Applied Economics, volume 45, issue 3, pp. 311-327.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. 2nd ed., London, Chapman and Hall Publisher.

Puhani, P. A., 2000. The Heckman Correction For Sample Selection And Its Critique. Journal of Economic Surveys, volume 14, issue 1, pp. 53-68.

Russell Davidson, R., MacKinnon, J.G., (2003). Econometric Theory and Methods. Oxford University Press (New York), October 2003, with a 2004 copyright. The ISBN is 0-19-512372-7.

Santos Silva, J. M. C., Tenreyro, S., 2011. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. Economics Letters, volume 112, issue 2, pp. 220–222.

Santos Silva, J.M.C., Tenreyro, S., 2006. The Log of Gravity. The review of economics and Statsitics 88 (4), 641-658.

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

Shepherd, B., 2010. Geographic Diversification of Developing Country Exports. World Development, Volume 38, issue 9, pp. 1217-1228.

Sukanuntathum, A., 2012. Robust Estimation of Gravity Models under Heteroskedasticity and Data Censoring. Procedia - Social and Behavioral Sciences, volume 40, pp. 731-735.

Tenreyro, S., 2007. On the trade impact of nominal exchange rate volatility. Journal of Development Economics, volume 82, pp. 485–508.

Tobin, J., 1958. Estimation of Relationships for Limited Dependent Variables. Econometrica, volume 26, issue 1, pp. 24–36.

Westerlund, J., Wilhelmsson, F., 2011. Estimating the gravity model without gravity using panel data. Applied Economics, volume 43, issue 6, pp.641–649.

Xiong, B., Chen, S., 2014. Estimating gravity equation models in the presence of sample selection and heteroscedasticity. Applied Economics, volume 46, issue 24, pp. 2993-3003.

Yotov, Y. V., 2012. A simple solution to the distance puzzle in international trade. Economics Letters, Volume 117, pp 794–798.

Appendix

1) FDG Single equations

Table A1-1

Errors Normally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG Single equations									
case 1	OLS	-0.196099	0.028879	-0.161366	0.049473	0.023479	0.041546	0.058334	0.070431
	Heckman 2SLS	0.000261	0.010139	-0.000137	0.020185	0.127080	0.011912	0.127514	0.023972
	GHeckman 2SLS	-0.000121	0.010249	-0.000645	0.020465	0.147301	0.014717	0.147528	0.025587
	heckman2sls_Log	-0.019114	0.033089	0.014148	0.055113	0.379170	0.046678	0.425799	0.075712
	Gheckman2sls_Log	-0.018797	0.033062	0.014366	0.055117	0.374869	0.091425	0.415272	0.115795
	NNLS	-0.022986	0.009802	-0.026749	0.019325	0.117217	0.011373	0.116009	0.023531
	Gnnls	-0.028699	0.010706	-0.032876	0.020077	0.128466	0.013116	0.126289	0.024757
	nnls_C	-0.022986	0.009802	-0.026749	0.019325	0.149097	0.011213	0.153677	0.024106
	Gnnls_C	-0.028699	0.010706	-0.032876	0.020077	0.193703	0.021461	0.198924	0.031888
	PPML	-0.124846	0.012473	-0.123438	0.022290	0.086472	0.015449	0.085241	0.029410
	PPML_C	-0.124846	0.012473	-0.123438	0.022290	0.231669	0.016629	0.245720	0.032690
	Optimal_PPML	-0.216352	0.015935	-0.211279	0.022181	-0.030102	0.024559	-0.031896	0.032000
	Optimal_PPML_C	-0.189830	0.016785	-0.182467	0.023258	0.040586	0.027320	0.048911	0.034303
case 2	OLS	-0.105356	0.025890	-0.094027	0.057883	0.168961	0.041443	0.172169	0.075936
	Heckman 2SLS	-0.013369	0.036735	-0.016955	0.060524	0.084405	0.044126	0.081808	0.073303
	GHeckman 2SLS	-0.056072	0.022896	-0.061612	0.041653	0.087218	0.027448	0.081922	0.052258
	heckman2sls_Log	0.028376	0.031208	0.043105	0.065995	0.272549	0.080166	0.276058	0.103259
	Gheckman2sls_Log	0.028994	0.030300	0.043765	0.065766	0.272700	0.079514	0.275979	0.101415
	NNLS	-0.032683	0.031909	-0.038758	0.055679	0.081452	0.039031	0.078523	0.069632
	Gnnls	-0.100966	0.023702	-0.106076	0.040388	0.075017	0.022999	0.068921	0.049679
	nnls_C	-0.032683	0.031909	-0.038758	0.055679	0.142418	0.037905	0.145455	0.073001
	Gnnls_C	-0.100966	0.023702	-0.106076	0.040388	0.229963	0.034358	0.235574	0.063682
	PPML	-0.092457	0.017172	-0.098181	0.037970	0.073418	0.022434	0.067722	0.048877
	PPML_C	-0.092457	0.017172	-0.098181	0.037970	0.254988	0.023664	0.263316	0.056814
	Optimal_PPML	-0.144901	0.015578	-0.149469	0.032430	0.004084	0.021415	-0.001508	0.042718
	Optimal_PPML_C	-0.116357	0.015724	-0.118643	0.033319	0.108022	0.020685	0.109615	0.045046
case 3	OLS	0.002244	0.030700	0.006043	0.071415	0.260502	0.054979	0.263740	0.096668
	Heckman 2SLS	0.010259	0.127420	0.009494	0.171189	0.052020	0.152695	0.056064	0.202497
	GHeckman 2SLS	0.000273	0.040058	0.000611	0.052400	0.080191	0.056622	0.079576	0.077325
	heckman2sls_Log	0.004381	0.033469	0.007444	0.083293	-0.056831	0.087225	-0.044759	0.123284
	Gheckman2sls_Log	0.003383	0.033329	0.007197	0.083549	-0.061149	0.094021	-0.049458	0.127348
	NNLS	0.008428	0.098625	0.007028	0.146092	0.073983	0.115998	0.079550	0.173924
	Gnnls	0.000168	0.025736	0.000803	0.047652	0.150864	0.052426	0.151866	0.074261
	nnls_C	0.008428	0.098625	0.007028	0.146092	0.144665	0.141628	0.152544	0.216254
	Gnnls_C	0.000168	0.025736	0.000803	0.047652	0.307932	0.114812	0.313139	0.134037
	PPML	0.001328	0.038143	-0.000645	0.065411	0.111266	0.051526	0.113930	0.086712
	PPML_C	0.001328	0.038143	-0.000645	0.065411	0.264010	0.055522	0.273620	0.109295
	Optimal_PPML	-0.001226	0.035463	-0.003864	0.057976	0.092258	0.048755	0.093923	0.077408
	Optimal_PPML_C	-0.001417	0.037847	-0.003700	0.062998	0.179382	0.045103	0.185134	0.078402
case 4	OLS	-0.109181	0.034464	0.113723	0.072513	0.092147	0.054887	0.387417	0.097672
	Heckman 2SLS	0.002217	0.167117	0.284995	0.196706	0.030968	0.190879	0.343026	0.233789
	GHeckman 2SLS	-0.029991	0.053761	0.220240	0.069170	0.030790	0.058947	0.339595	0.083248
	heckman2sls_Log	-0.099222	0.060203	0.123223	0.105190	-0.013960	0.202183	0.301103	0.186908
	Gheckman2sls_Log	-0.097456	0.059583	0.125572	0.106720	-0.016705	0.206160	0.298041	0.189819
	NNLS	-0.011025	0.134052	0.269991	0.164096	0.034585	0.150482	0.350621	0.188960
	Gnnls	-0.103031	0.040032	0.153179	0.059376	0.042947	0.042948	0.353521	0.070287
	nnls_C	-0.011025	0.134052	0.269991	0.164096	0.106065	0.187123	0.308853	0.239775
	Gnnls_C	-0.103031	0.040032	0.153179	0.059376	0.194854	0.078650	0.421702	0.114398
	PPML	-0.054276	0.046248	0.208100	0.069238	0.033451	0.057857	0.346193	0.089176
	PPML_C	-0.054276	0.046248	0.208100	0.069238	0.166316	0.065501	0.392535	0.114948

	Optimal_PPML	-0.061662	0.042108	0.217946	0.061460	0.018410	0.053389	0.345916	0.078611
	Optimal_PPML_C	-0.050993	0.045592	0.170142	0.068416	0.087105	0.052068	0.373561	0.082400
case 5	**OLS**	-0.234880	0.035789	-0.203404	0.076002	-0.044927	0.047332	-0.022552	0.095804
	Heckman 2SLS	0.014609	0.121959	0.013174	0.167363	0.061534	0.151760	0.065790	0.205260
	GHeckman 2SLS	-0.009806	0.116451	-0.001760	0.106332	0.099058	0.226382	0.083459	0.489654
	heckman2sls_Log	**-0.136704**	0.040879	-0.111648	0.088104	0.210685	0.052987	0.233838	0.109937
	Gheckman2sls_Log	**-0.136536**	0.040881	-0.111516	0.088080	0.215916	0.082847	0.236830	0.124394
	NNLS	-0.019203	0.100565	-0.023210	0.144284	0.051168	0.116053	0.053809	0.172384
	Gnnls	-0.310293	0.216619	-0.272210	0.207657	-0.083536	0.201532	-0.062999	0.174836
	nnls_C	-0.019203	0.100565	-0.023210	0.144284	0.127802	0.141660	0.133166	0.213375
	Gnnls_C	-0.310293	0.216619	-0.272210	0.207657	0.173574	0.110114	0.182916	0.122883
	PPML	-0.104393	0.040160	-0.104476	0.064300	0.024750	0.049408	0.022625	0.082720
	PPML_C	-0.104393	0.040160	-0.104476	0.064300	0.172506	0.054858	0.177775	0.104043
	Optimal_PPML	-0.137068	0.035720	-0.134232	0.057395	-0.011418	0.044544	-0.012332	0.072263
	Optimal_PPML_C	-0.119493	0.036864	-0.115216	0.061516	0.060233	0.043205	0.064680	0.074411

Table A1-2

Errors Lognormally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG Single equations									
case 1	OLS	0.390174	0.034754	0.354656	0.050923	0.195054	0.059256	0.225888	0.088609
	Heckman 2SLS	0.000092	0.009777	-0.001151	0.020299	0.139065	0.013628	0.142010	0.025236
	GHeckman 2SLS	0.000346	0.009862	-0.000904	0.020582	0.146573	0.016835	0.148969	0.027213
	heckman2sls_Log	0.390174	0.034754	0.354656	0.050923	0.566820	0.051194	0.603837	0.082126
	Gheckman2sls_Log	0.174967	0.073055	0.174138	0.071020	0.561676	0.061538	0.597218	0.088986
	NNLS	0.000092	0.009777	-0.001151	0.020299	0.125074	0.011964	0.125549	0.024640
	Gnnls	0.000346	0.009862	-0.000904	0.020582	0.129395	0.015062	0.129379	0.026382
	nnls_C	0.000092	0.009777	-0.001151	0.020299	0.156634	0.011474	0.162301	0.024918
	Gnnls_C	0.000346	0.009862	-0.000904	0.020582	0.178746	0.026331	0.183997	0.035827
	PPML	-0.001140	0.019069	-0.000966	0.030545	0.098231	0.033798	0.099699	0.048577
	PPML_C	-0.001140	0.019069	-0.000966	0.030545	0.273547	0.028834	0.290752	0.046500
	Optimal_PPML	-0.181495	0.100943	-0.179886	0.116196	-0.159823	0.143600	-0.165088	0.160869
	Optimal_PPML_C	-0.181495	0.100943	-0.179886	0.116196	-0.063794	0.143962	-0.059267	0.168756
case 2	OLS	0.211405	0.027346	0.200551	0.049285	0.185778	0.058600	0.202092	0.093877
	Heckman 2SLS	0.002652	0.031416	-0.000507	0.058176	0.125617	0.043994	0.123536	0.077602
	GHeckman 2SLS	0.000712	0.019848	-0.002459	0.044425	0.160714	0.037427	0.157816	0.066022
	heckman2sls_Log	0.211405	0.027346	0.200551	0.049285	0.558242	0.060763	0.585536	0.094290
	Gheckman2sls_Log	0.176842	0.025757	0.173866	0.047125	0.552399	0.068402	0.578154	0.101733
	NNLS	0.002652	0.031416	-0.000507	0.058176	0.113988	0.040893	0.110003	0.074869
	Gnnls	0.000707	0.019850	-0.002491	0.044471	0.118574	0.031737	0.110773	0.061754
	nnls_C	0.002652	0.031416	-0.000507	0.058176	0.158032	0.038225	0.161758	0.075359
	Gnnls_C	0.000707	0.019850	-0.002491	0.044471	0.230651	0.034965	0.236076	0.067296
	PPML	0.000708	0.019592	-0.001869	0.043993	0.087160	0.030707	0.078740	0.062515
	PPML_C	0.000708	0.019592	-0.001869	0.043993	0.286378	0.028359	0.298563	0.065830
	Optimal_PPML	-0.127158	0.046611	-0.129649	0.070018	-0.104810	0.071308	-0.115978	0.100112
	Optimal_PPML_C	-0.127158	0.046611	-0.129649	0.070018	0.011936	0.074030	0.012193	0.107203
case 3	OLS	0.001929	0.024262	0.002181	0.057526	0.064549	0.067521	0.058874	0.120852
	Heckman 2SLS	0.047786	0.323882	0.024478	0.276366	0.086508	0.207032	0.085054	0.283903
	GHeckman 2SLS	-0.004541	0.044072	-0.011450	0.082331	0.055103	0.100984	0.037045	0.141603
	heckman2sls_Log	0.001929	0.024262	0.002181	0.057526	0.302370	0.138351	0.301375	0.174700
	Gheckman2sls_Log	0.001356	0.024405	0.001726	0.057813	0.301762	0.139825	0.300961	0.174110
	NNLS	0.019830	0.145841	0.007972	0.211330	0.084620	0.195316	0.076484	0.269281
	Gnnls	-0.004681	0.043149	-0.010916	0.081170	-0.000349	0.081262	-0.017527	0.129727
	nnls_C	0.019830	0.145841	0.007972	0.211330	0.175338	0.183032	0.174121	0.269021
	Gnnls_C	-0.004681	0.043149	-0.010916	0.081170	0.301510	0.101093	0.301984	0.147965
	PPML	0.002928	0.056386	-0.006466	0.098592	0.037693	0.085354	0.021038	0.139457
	PPML_C	0.002928	0.056386	-0.006466	0.098592	0.292494	0.072739	0.294846	0.143516
	Optimal_PPML	-0.015046	0.080756	-0.017528	0.146491	-0.032298	0.115177	-0.037163	0.194416
	Optimal_PPML_C	-0.015046	0.080756	-0.017528	0.146491	0.118802	0.097435	0.123157	0.186344
case 4	OLS	0.132504	0.035675	-0.121275	0.079575	-0.191509	0.074097	-0.153559	0.139552
	Heckman 2SLS	0.012510	0.192370	-0.007420	0.260508	0.005874	0.339230	0.101992	0.371598
	GHeckman 2SLS	-0.013542	0.068080	-0.020179	0.122767	-0.079348	0.188153	0.084279	0.211043
	heckman2sls_Log	0.132504	0.035675	-0.121275	0.079575	0.217489	0.137406	0.146956	0.167269
	Gheckman2sls_Log	0.116172	0.035501	-0.135192	0.080579	0.278756	0.247168	0.159068	0.233357
	NNLS	0.013927	0.194697	-0.002161	0.261645	-0.018859	0.308919	0.081082	0.352854

	Gnnls	-0.013274	0.065926	-0.020706	0.121002	-0.234909	0.230711	-0.055840	0.210848
	nnls_C	0.013927	0.194697	-0.002161	0.261645	0.177659	0.346839	0.160177	0.349581
	Gnnls_C	-0.013274	0.065926	-0.020706	0.121002	0.210839	0.113896	0.245534	0.183971
	PPML	0.000236	0.083652	-0.013409	0.135289	-0.132300	0.125993	0.015620	0.177909
	PPML_C	0.000236	0.083652	-0.013409	0.135289	0.227392	0.105890	0.257798	0.179698
	Optimal_PPML	-0.095336	0.136626	0.131402	0.236740	-0.223184	0.180296	0.115381	0.280065
	Optimal_PPML_C	-0.095336	0.136626	0.131402	0.236740	-0.027261	0.154907	0.257223	0.272356
case 5	**OLS**	0.301283	0.038704	0.263431	0.069434	-0.090553	0.070607	-0.087257	0.129669
	Heckman 2SLS	0.019800	0.144135	0.003068	0.217906	0.102449	0.216071	0.098410	0.294665
	GHeckman 2SLS	-0.008053	0.100406	-0.022835	0.125567	0.065668	0.175797	0.056253	0.204103
	heckman2sls_Log	0.301283	0.038704	0.263431	0.069434	0.350662	0.077612	0.366259	0.138218
	Gheckman2sls_Log	0.203544	0.042572	0.187062	0.072246	0.366104	0.120786	0.372423	0.184545
	NNLS	0.020262	0.144409	0.002267	0.217364	0.064407	0.196657	0.057107	0.271371
	Gnnls	-0.003510	0.059653	-0.015613	0.095815	-0.178339	0.308172	-0.191440	0.309387
	nnls_C	0.020262	0.144409	0.002267	0.217364	0.165920	0.177973	0.164562	0.268095
	Gnnls_C	-0.003510	0.059653	-0.015613	0.095815	0.239703	0.084540	0.234614	0.146029
	PPML	0.002144	0.058430	-0.011990	0.103886	-0.049592	0.089490	-0.070367	0.142749
	PPML_C	0.002144	0.058430	-0.011990	0.103886	0.239796	0.074936	0.239873	0.145160
	Optimal_PPML	-0.126306	0.118093	-0.125434	0.177859	-0.196775	0.148040	-0.203934	0.217489
	Optimal_PPML_C	-0.126306	0.118093	-0.125434	0.177859	-0.045974	0.142454	-0.040781	0.217909

2) FDG ETtobit equations

3) Table A2-1

Errors Normally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG ETtobit equations									
case 1	Etobit	-0.004700	0.012300	-0.005500	0.021700	0.000128	0.015900	-0.000326	0.024500
	GETtobit	-0.005100	0.012200	-0.006000	0.021900	0.000700	0.016000	0.000200	0.024700
	n_ETtobit	-0.530400	0.024800	-0.489300	0.028800	-0.494100	0.040300	-0.458400	0.041200
	Gn_ETtobit	-0.006000	0.012100	-0.006900	0.021900	0.000800	0.016000	0.000300	0.024700
	Ettobit_L	-0.125600	0.011800	-0.128900	0.020600	-0.003300	0.017300	-0.004100	0.025300
	GETtobit_L	-0.167000	0.030200	-0.169000	0.034400	-0.004400	0.017800	-0.005200	0.025900
	Ettobit_LOG	-0.522300	0.013600	-0.499400	0.019800	-0.400600	0.020800	-0.379600	0.028600
	GEttobit_LOG	-0.167000	0.030200	-0.169000	0.034400	-0.004500	0.017800	-0.005400	0.025900
case 2	Etobit	-0.030200	0.049700	-0.032900	0.068900	-0.159100	0.055200	-0.165300	0.067900
	GETtobit	-0.006500	0.020600	-0.008500	0.043100	-0.007800	0.035700	-0.009700	0.053000
	n_ETtobit	-0.331300	0.019600	-0.311800	0.034200	-0.318200	0.026000	-0.312100	0.038200
	Gn_ETtobit	-0.007000	0.020400	-0.009000	0.043100	-0.007700	0.035700	-0.009700	0.053000
	Ettobit_L	-0.369300	0.035700	-0.368300	0.050800	-0.111100	0.035000	-0.122000	0.060000
	GETtobit_L	-0.439600	0.019500	-0.420700	0.032300	-0.299200	0.035400	-0.300900	0.046800
	Ettobit_LOG	-0.450200	0.019000	-0.427700	0.031800	-0.346000	0.026800	-0.343300	0.037000
	GEttobit_LOG	-0.439500	0.019600	-0.420600	0.032300	-0.298900	0.035400	-0.300600	0.046900
case 3	Etobit	0.014700	0.192300	0.007900	0.236900	-0.549200	0.169200	-0.558900	0.178100
	GETtobit	0.000700	0.022600	-0.000500	0.040200	0.009100	0.038700	0.007100	0.056600
	n_ETtobit	-0.003700	0.021900	-0.003800	0.040300	-0.000700	0.034700	-0.001900	0.051600
	Gn_ETtobit	0.002000	0.022600	0.000600	0.040400	0.007400	0.035500	0.005300	0.053400
	Ettobit_L	-0.775700	0.031600	-0.777000	0.066400	-0.436600	0.049700	-0.445700	0.092800
	GETtobit_L	-0.443100	0.028600	-0.427600	0.065000	-0.254400	0.042100	-0.261200	0.065400
	Ettobit_LOG	-0.442900	0.028600	-0.427400	0.065000	-0.253900	0.042100	-0.260600	0.065400
	GEttobit_LOG	-0.442900	0.028600	-0.427400	0.065000	-0.253900	0.042100	-0.260600	0.065400
case 4	Etobit	-0.007700	0.267400	0.071300	0.315400	-0.461000	0.213500	-0.431200	0.236500
	GETtobit	-0.035700	0.036800	0.241600	0.057000	-0.004800	0.049800	0.266300	0.081200
	n_ETtobit	-0.142300	0.027500	0.132000	0.046300	-0.078800	0.035600	0.183000	0.061800
	Gn_ETtobit	-0.036500	0.038000	0.240600	0.058100	-0.004700	0.049900	0.266400	0.081400
	Ettobit_L	-0.801100	0.032100	-0.913900	0.064800	-0.547800	0.044100	-0.618600	0.089900
	GETtobit_L	-0.525900	0.032500	-0.544700	0.069000	-0.344000	0.041500	-0.311200	0.072600
	Ettobit_LOG	-0.525700	0.032500	-0.544600	0.069000	-0.343500	0.041500	-0.310700	0.072700
	GEttobit_LOG	-0.525700	0.032500	-0.544600	0.069000	-0.343500	0.041500	-0.310700	0.072700
case 5	Etobit	-0.053700	0.198500	-0.057400	0.242300	-0.271500	0.168100	-0.279600	0.191900
	GETtobit	0.003000	0.037400	0.002300	0.057500	0.029700	0.058400	0.027600	0.076100
	n_ETtobit	-0.340300	0.036800	-0.287600	0.051500	-0.286400	0.054400	-0.240900	0.064500
	Gn_ETtobit	0.003500	0.037300	0.002800	0.057300	0.029800	0.058300	0.027700	0.076000
	Ettobit_L	-0.771000	0.033300	-0.774000	0.056100	-0.563100	0.044400	-0.561400	0.078900
	GETtobit_L	-0.562900	0.027600	-0.551300	0.049000	-0.415900	0.037600	-0.404300	0.061100
	Ettobit_LOG	-0.562700	0.027600	-0.551100	0.049000	-0.418300	0.036900	-0.405700	0.060800
	GEttobit_LOG	-0.562700	0.027600	-0.551100	0.049000	-0.416100	0.037600	-0.404300	0.061100

Table A2-2

Errors Lognormally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG ETtobit equations									
case 1	Etobit	-0.003500	0.011700	-0.006100	0.020700	-0.095300	0.032000	-0.097700	0.036200
	GETtobit	-0.012200	0.025100	-0.014300	0.032000	-0.093800	0.025800	-0.095800	0.030700
	n_ETtobit	-0.464800	0.129000	-0.444700	0.148900	-0.685300	0.092700	-0.683600	0.094700
	Gn_ETtobit	-0.014900	0.058500	-0.013900	0.038300	-0.099000	0.040700	-0.101400	0.043900
	Ettobit_L	0.008487	0.009000	-0.007482	0.017900	-0.000900	0.017800	-0.000800	0.022800
	GETtobit_L	0.001900	0.009200	0.000300	0.018100	-0.005800	0.017700	-0.009800	0.024100
	Ettobit_LOG	0.031900	0.031500	0.047500	0.038400	-0.336300	0.030800	-0.331000	0.036600
	GEttobit_LOG	**0.002000**	**0.009200**	**0.000300**	**0.018100**	**-0.005800**	**0.017700**	**-0.009700**	**0.024200**
case 2	Etobit	-0.004200	0.039900	-0.010300	0.062900	-0.245300	0.054200	-0.255900	0.063000
	GETtobit	0.004400	0.035400	-0.008400	0.079900	-0.219100	0.045800	-0.229200	0.059000
	n_ETtobit	-0.317100	0.094700	-0.311100	0.117100	-0.582000	0.074200	-0.581700	0.083000
	Gn_ETtobit	-0.000900	0.054100	-0.013400	0.084700	-0.218900	0.045800	-0.229000	0.059000
	Ettobit_L	-0.175300	0.018300	-0.183700	0.031300	0.017000	0.032200	0.015900	0.041800
	GETtobit_L	0.001500	0.016600	-0.000600	0.035200	0.005500	0.041500	0.004800	0.047100
	Ettobit_LOG	0.056000	0.028600	0.066100	0.043100	-0.245200	0.045400	-0.240500	0.047300
	GEttobit_LOG	**0.001500**	**0.016600**	**-0.000600**	**0.035100**	**0.005700**	**0.041500**	**0.004900**	**0.047100**
case 3	Etobit	-0.009300	0.171900	-0.021000	0.223200	-0.710600	0.121500	-0.719500	0.118100
	GETtobit	0.019800	0.071400	-0.005700	0.144200	-0.085200	0.125800	-0.111100	0.159700
	n_ETtobit	-0.004700	0.073800	-0.023600	0.145400	-0.250800	0.095600	-0.264400	0.132100
	Gn_ETtobit	0.021200	0.071000	-0.005200	0.144000	-0.084900	0.125700	-0.110800	0.159800
	Ettobit_L	-0.507500	0.019600	-0.506500	0.038000	-0.166900	0.046900	-0.163900	0.064000
	GETtobit_L	-0.004900	0.026800	-0.003200	0.057600	0.016800	0.070700	0.016200	0.078000
	Ettobit_LOG	-0.001500	0.026200	-0.000300	0.057600	0.009000	0.069900	0.009200	0.079500
	GEttobit_LOG	**-0.004200**	**0.026600**	**-0.002500**	**0.057800**	**0.017400**	**0.070900**	**0.016800**	**0.078100**
case 4	Etobit	-0.019200	0.256100	-0.033100	0.309300	-0.801500	0.117100	-0.826900	0.101000
	GETtobit	-0.014300	0.142900	0.187000	0.271500	-0.135100	0.197900	0.042000	0.283400
	n_ETtobit	-0.129100	0.146400	0.085600	0.287400	-0.393900	0.123700	-0.219600	0.243500
	Gn_ETtobit	-0.019000	0.144800	0.185700	0.272100	-0.133900	0.199200	0.044000	0.287600
	Ettobit_L	-0.469100	0.022500	-0.671500	0.045200	-0.204500	0.054300	-0.444400	0.089400
	GETtobit_L	0.036700	0.045300	-0.208300	0.079800	0.035000	0.082800	-0.138100	0.109900
	Ettobit_LOG	0.104300	0.038800	-0.140900	0.077200	0.014900	0.085800	-0.138700	0.109400
	GEttobit_LOG	**0.036600**	**0.**045300	**-0.208500**	**0.079800**	**0.035700**	**0.082900**	**-0.137500**	**0.110100**
case 5	Etobit	-0.021100	0.176600	-0.032800	0.221400	-0.657200	0.119200	-0.662400	0.122800
	GETtobit	-0.032100	0.141900	-0.055100	0.235500	-0.166800	0.132700	-0.186600	0.163400
	n_ETtobit	-0.285100	0.161500	-0.299700	0.253100	-0.523800	0.140100	-0.521000	0.155900
	Gn_ETtobit	-0.041800	0.150800	-0.059600	0.237300	-0.166500	0.132700	-0.186300	0.163400
	Ettobit_L	-0.342300	0.023500	-0.350900	0.042300	-0.189400	0.046600	-0.179500	0.077800
	GETtobit_L	0.006000	0.038600	0.007000	0.059000	0.002800	0.071000	0.003900	0.095300
	Ettobit_LOG	0.267100	0.042300	0.256200	0.072300	-0.065300	0.076500	-0.058300	0.095900
	GEttobit_LOG	**0.006000**	**0.038600**	**0.007000**	**0.059000**	**0.003200**	**0.071000**	**0.004300**	**0.095300**

3) FDG Heckman equations

Table A3-1

Errors Normally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG Heckman equations									
case 1	heckmanml (selectioneq: beta=alpha)	0.001800	0.010300	0.001900	0.020700	0.135400	0.012800	0.135900	0.024400
	Gheckmanml	0.000300	0.010500	0.000500	0.021000	0.165700	0.014600	0.167600	0.027300
	n_heckmanml	-0.419000	0.029200	-0.376700	0.032000	-0.298600	0.051500	-0.271200	0.052300
	Gn_heckmanml	0.000600	0.010400	0.000700	0.020900	0.162600	0.020200	0.164400	0.030700
	heckmanml_L	-0.071300	0.017000	-0.071300	0.029800	-0.079200	0.190200	-0.079300	0.195100
	Gheckmanml_L	-0.071200	0.017100	-0.071100	0.029700	-0.073100	0.189600	-0.073900	0.192300
	heckmanml_LOG	-0.132900	0.028900	-0.102100	0.045600	0.194900	0.046200	0.230800	0.070600
	Gheckmanml_LOG	-0.089600	0.018800	-0.086000	0.030300	0.086600	0.041700	0.089000	0.057100
case 2	heckmanml (selection equation: beta=alpha)	-0.010900	0.036400	-0.013200	0.063300	0.105539	0.075740	0.101197	0.096443
	Gheckmanml	-0.023000	0.021200	-0.025700	0.040500	0.130886	0.063104	0.127684	0.079637
	n_heckmanml	-0.255100	0.018100	-0.250700	0.032800	-0.109774	0.023562	-0.118168	0.047814
	Gn_heckmanml	-0.023500	0.021600	-0.026100	0.040400	0.130491	0.062421	0.127293	0.079129
	heckmanml_L	-0.373800	0.026800	-0.386600	0.049000	-0.324056	0.028699	-0.336859	0.047450
	Gheckmanml_L	-0.190900	0.029500	-0.193200	0.063300	-0.318899	0.026661	-0.329375	0.048579
	heckmanml_LOG	-0.086500	0.048100	-0.079400	0.072700	0.392213	0.043459	0.391481	0.078914
	Gheckmanml_LOG	-0.294200	0.044700	-0.288600	0.073700	-0.240836	0.041557	-0.251309	0.059724
case 3	heckmanml (selection eq: beta=alpha)	0.006200	0.144000	0.005600	0.193800	0.037704	0.263307	0.038376	0.313153
	Gheckmanml	0.006000	0.020100	0.001800	0.037500	0.235309	0.251351	0.229815	0.253978
	n_heckmanml	0.001200	0.020000	-0.001900	0.037200	0.119290	0.039316	0.114846	0.059025
	Gn_heckmanml	0.004000	0.020500	0.000000	0.037000	0.122427	0.038204	0.117813	0.058154
	heckmanml_L	-0.438000	0.039900	-0.449600	0.060900	-0.396221	0.043750	-0.396453	0.066449
	Gheckmanml_L	-0.012100	0.033900	-0.011200	0.062100	-0.202190	0.046071	-0.197240	0.070723
	heckmanml_LOG	0.003200	0.029900	0.001600	0.066400	0.556873	0.052978	0.541624	0.097438
	Gheckmanml_LOG	-0.005500	0.031300	-0.003700	0.066000	-0.079487	0.051430	-0.069641	0.079654
case 4	heckmanml (selectioneq: beta=alpha)	-0.016500	0.303100	0.171200	0.301000	0.050612	0.281423	0.350950	0.345067
	Gheckmanml	-0.033100	0.041200	0.234000	0.061300	0.633924	0.223993	1.110065	0.308847
	n_heckmanml	-0.132600	0.026200	0.131400	0.045400	0.063285	0.053960	0.370274	0.069821
	Gn_heckmanml	-0.073900	0.030700	0.184000	0.049000	0.226453	0.252662	0.579291	0.334638
	heckmanml_L	-0.474300	0.032300	-0.348400	0.053000	-0.445753	0.041052	-0.287810	0.063871
	Gheckmanml_L	-0.146800	0.045400	0.109200	0.076200	-0.205735	0.048293	0.101729	0.085133
	heckmanml_LOG	-0.092100	0.038700	0.098200	0.072000	0.350288	0.054172	0.575937	0.098907
	Gheckmanml_LOG	-0.115900	0.040900	0.095800	0.073600	-0.137478	0.044966	0.192682	0.077184
case 5	heckmanml (selectioneq: beta=alpha)	-0.045900	0.169200	-0.040400	0.222300	0.127225	0.219928	0.139055	0.301881
	Gheckmanml	0.020600	0.043700	0.018000	0.066200	0.369827	0.061450	0.373936	0.096250
	n_heckmanml	-0.382100	0.035900	-0.335200	0.052400	-0.276299	0.052770	-0.238572	0.066767
	Gn_heckmanml	-0.012500	0.040800	-0.016700	0.061200	0.214047	0.087810	0.210933	0.114008
	heckmanml_L	-0.571300	0.037000	-0.569400	0.053600	-0.519983	0.032932	-0.520559	0.055781
	Gheckmanml_L	-0.366700	0.042900	-0.351100	0.075000	-0.410348	0.050146	-0.390076	0.076840
	heckmanml_LOG	-0.269900	0.039100	-0.242100	0.069500	0.006636	0.051609	0.031156	0.090850
	Gheckmanml_LOG	-0.383900	0.077700	-0.363300	0.101600	-0.344804	0.073616	-0.327595	0.102404

Table A3-2

Errors Lognormally Distributed		k=0				k=1			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG Heckman equations									
case 1	heckmanml (selectioneq: beta=alpha)	0.0004431	0.0099000	-0.0006477	0.0207000	0.140300	0.015100	0.143800	0.028900
	Gheckmanml	-0.0015000	0.0183000	-0.0015000	0.0301000	0.139800	0.073600	0.150600	0.037800
	n_heckmanml	-0.3453000	0.1329000	-0.3548000	0.2028000	-0.426500	0.147300	-0.459300	0.231800
	Gn_heckmanml	-0.0072000	0.0173000	-0.0102000	0.0330000	0.114500	0.063400	0.117900	0.079700
	heckmanml_L	0.0006810	0.0092000	-0.0002488	0.0180000	0.153400	0.014100	0.171400	0.032400
	Gheckmanml_L	0.0018000	0.0093000	0.0009000	0.0182000	0.156800	0.015100	0.174300	0.032700
	heckmanml_LOG	0.3892000	0.0358000	0.3516000	0.0510000	0.384100	0.059400	0.418000	0.088600
	Gheckmanml_LOG	0.0388000	0.1114000	0.0342000	0.1006000	0.146400	0.020900	0.163500	0.036200
case 2	heckmanml (selectioneq: beta=alpha)	0.0030000	0.0314000	0.0008000	0.0592000	0.147100	0.052100	0.148000	0.087600
	Gheckmanml	0.0132000	0.0291000	0.0031000	0.0773000	0.197100	0.049000	0.195300	0.098900
	n_heckmanml	-0.2351000	0.0953000	-0.2423000	0.1398000	-0.330800	0.106600	-0.357100	0.170000
	Gn_heckmanml	0.0133000	0.0289000	0.0034000	0.0770000	0.151000	0.072400	0.144100	0.119200
	heckmanml_L	-0.1755000	0.0183000	-0.1829000	0.0311000	-0.054100	0.081200	-0.059300	0.096200
	Gheckmanml_L	0.0018000	0.0164000	0.0003000	0.0350000	0.052700	0.146600	0.051200	0.163500
	heckmanml_LOG	0.2110000	0.0283000	0.1979000	0.0485000	0.441400	0.067900	0.460200	0.102300
	Gheckmanml_LOG	0.0216000	0.0633000	0.0194000	0.0671000	0.033400	0.108800	0.031100	0.121800
case 3	heckmanml (selectioneq: beta=alpha)	0.0197000	0.1412000	0.0156000	0.2138000	0.223200	0.299700	0.255900	0.409800
	Gheckmanml	0.0242000	0.0660000	0.0052000	0.1364000	0.493800	0.328200	0.479500	0.422200
	n_heckmanml	0.0013000	0.0682000	-0.0114000	0.1397000	-0.066500	0.114000	-0.085300	0.201000
	Gn_heckmanml	0.0244000	0.0661000	0.0057000	0.1371000	0.319800	0.370900	0.297500	0.424400
	heckmanml_L	-0.5069000	0.0205000	-0.5079000	0.0372000	-0.517300	0.048900	-0.526700	0.072900
	Gheckmanml_L	-0.0020000	0.0264000	-0.0035000	0.0561000	-0.377500	0.110900	-0.386200	0.151900
	heckmanml_LOG	0.0025000	0.0251000	0.0003000	0.0559000	0.492800	0.077100	0.487600	0.133500
	Gheckmanml_LOG	-0.0024000	0.0262000	-0.0037000	0.0561000	-0.276200	0.084000	-0.289700	0.125500
case 4	heckmanml (selectioneq: beta=alpha)	0.0254000	0.2125000	0.0027000	0.2838000	0.303000	0.528000	0.489500	0.987300
	Gheckmanml	-0.0045000	0.1315000	0.2012000	0.2554000	0.384900	0.417400	0.872000	0.586300
	n_heckmanml	-0.1174000	0.1366000	0.0988000	0.2755000	-0.238700	0.160600	0.090600	0.292100
	Gn_heckmanml	-0.0100000	0.1315000	0.1925000	0.2630000	0.388500	0.449000	1.158400	4.416600
	heckmanml_L	-0.4683000	0.0224000	-0.6728000	0.0450000	-0.660800	0.071500	-0.577500	0.094300
	Gheckmanml_L	0.0358000	0.0441000	-0.2106000	0.0783000	-0.513800	0.140600	-0.221000	0.245300
	heckmanml_LOG	0.1334000	0.0357000	-0.1250000	0.0763000	0.388000	0.102400	0.266200	0.162200
	Gheckmanml_LOG	0.0374000	0.0456000	-0.2043000	0.0790000	-0.457700	0.127000	-0.247600	0.187000
case 5	heckmanml (selectioneq: beta=alpha)	0.0241000	0.1451000	0.0063000	0.2198000	0.286900	0.245300	0.335700	0.371200
	Gheckmanml	0.0213000	0.1073000	-0.0260000	0.2292000	0.555700	0.178600	0.550100	0.350200
	n_heckmanml	-0.2577000	0.1727000	-0.2981000	0.2978000	-0.335600	0.161800	-0.351300	0.263400
	Gn_heckmanml	0.0044000	0.1248000	-0.0410000	0.2301000	0.500500	0.239900	0.512700	0.402000
	heckmanml_L	-0.3399000	0.0240000	-0.3520000	0.0437000	-0.612500	0.068500	-0.614700	0.089100
	Gheckmanml_L	0.0093000	0.0375000	0.0063000	0.0593000	-0.524500	0.156200	-0.533000	0.185600
	heckmanml_LOG	0.3927000	0.0423000	0.3523000	0.0744000	0.346400	0.132400	0.349300	0.173100
	Gheckmanml_LOG	0.0052000	0.0370000	0.0030000	0.0589000	-0.397000	0.156600	-0.401000	0.174200

4) FDG Heckman_New equations (k=1)

Table 4A

		Errors Lognormally Distributed				Errors Normally Distributed			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
FDG Heckman_New equations									
case 1	heckmanml_New	0.1051645	0.0177979	0.1020574	0.0271954	0.1474483	0.0242358	0.1515898	0.0352009
	Gheckmanml_New	0.1134649	0.0237337	0.1113514	0.0335982	0.1618608	0.0268892	0.1641362	0.0347838
	n_heckmanml_New	-0.4407339	0.1206466	-0.4445395	0.1439046	-0.3097255	0.0477075	-0.2894354	0.0459534
	Gn_heckmanml_New	0.1093478	0.0234268	0.1065589	0.0319956	0.1537632	0.0189221	0.1551650	0.0291044
	heckmanml_L_New	0.1142010	0.0129079	0.1233716	0.0301763	0.1081790	0.0135750	0.1098425	0.0353737
	Gheckmanml_L_New	0.1141924	0.0129093	0.1233606	0.0301743	0.1098295	0.0170151	0.1119816	0.0387504
	heckmanml_LOG_New	0.1683703	0.0349560	0.1725191	0.0518956	0.0504444	0.0346720	0.0731051	0.0526355
	Gheckmanml_LOG_New	0.1142637	0.0129131	0.1234324	0.0301797	0.1116753	0.0180350	0.1139662	0.0389828
case 2	heckmanml_New	0.0653139	0.0467078	0.0565494	0.0680626	0.1906376	0.1037579	0.1947218	0.1330444
	Gheckmanml_New	0.0879373	0.0582558	0.0776510	0.0865373	0.0291372	0.0871875	0.0193026	0.0949631
	n_heckmanml_New	-0.3425345	0.0911376	-0.3499148	0.1090748	-0.1392668	0.0244590	-0.1494869	0.0423050
	Gn_heckmanml_New	0.0859773	0.0544027	0.0754181	0.0806206	0.0283588	0.0866443	0.0184600	0.0941573
	heckmanml_L_New	-0.1131180	0.0391016	-0.1188260	0.0514405	-0.4932877	0.1300610	-0.4944335	0.1269576
	Gheckmanml_L_New	-0.1021326	0.0510591	-0.1079961	0.0614657	-0.4182524	0.1212188	-0.4208555	0.1206756
	heckmanml_LOG_New	0.1510390	0.0366482	0.1508526	0.0597620	-0.0187624	0.1851020	-0.0362179	0.1899963
	Gheckmanml_LOG_New	-0.0791443	0.0787824	-0.0843768	0.0903990	-0.4146893	0.1224339	-0.4174794	0.1212445
case 3	heckmanml_New	0.3043070	0.5660993	0.2519512	0.5654774	2.5993300	4.5170941	2.1381827	7.2300229
	Gheckmanml_New	0.3212235	0.3881788	0.2864554	0.4286031	0.1970178	0.0365249	0.1938200	0.0617270
	n_heckmanml_New	-0.0680456	0.1033382	-0.0929285	0.1657961	0.1969591	0.0367707	0.1937348	0.0617141
	Gn_heckmanml_New	-0.0639607	0.1758323	-0.0905667	0.1873459	0.1965825	0.0363258	0.1934147	0.0616182
	heckmanml_L_New	-0.7283151	0.1103671	-0.7309193	0.1114052	-0.9765264	0.0060786	-0.9768663	0.0063984
	Gheckmanml_L_New	-0.7324449	0.1276009	-0.7348477	0.1286752	-0.9770393	0.0059290	-0.9774126	0.0062200
	heckmanml_LOG_New	-0.7588123	0.2084236	-0.7608336	0.2056740	-0.9567682	0.0076415	-0.9572758	0.0079068
	Gheckmanml_LOG_New	-0.7388126	0.1348711	-0.7413294	0.1357117	-0.9772858	0.0056891	-0.9776608	0.0059692
case 4	heckmanml_New	0.2932249	1.1650845	-0.2420139	0.7853144	0.8877846	2.8549708	-4.9919401	7.7923822
	Gheckmanml_New	0.0377612	0.2978128	0.4100827	0.4754297	0.0832135	0.1734528	0.4119990	0.2496062
	n_heckmanml_New	-0.2620667	0.1445128	0.0427514	0.2890251	0.0473959	0.0358388	0.3652111	0.0616529
	Gn_heckmanml_New	0.0429849	0.4374145	0.3829275	0.6583539	0.2140011	0.4028977	0.5797304	0.5194255
	heckmanml_L_New	-0.9697667	0.0720616	-0.9759762	0.0546349	-0.9937753	0.0015051	-0.9935302	0.0018389
	Gheckmanml_L_New	-0.9704137	0.0723883	-0.9765614	0.0544459	-0.9938554	0.0014965	-0.9936519	0.0018510
	heckmanml_LOG_New	-0.9388774	0.0207314	-0.9479375	0.0163830	-0.9839469	0.0034058	-0.9802728	0.0043075
	Gheckmanml_LOG_New	-0.9742202	0.0662823	-0.9792925	0.0503757	-0.9938651	0.0014663	-0.9936580	0.0018009
case 5	heckmanml_New	0.4513706	0.4848255	0.3865275	0.4748339	4.7538940	4.8890449	4.9480261	8.7972943
	Gheckmanml_New	0.3266703	0.3979702	0.2968351	0.4446286	0.1623431	0.2946069	0.1601291	0.3073835
	n_heckmanml_New	-0.3282809	0.1394866	-0.3439494	0.2099681	-0.1606979	0.0553970	-0.1288501	0.0673285
	Gn_heckmanml_New	0.0151054	0.2672389	-0.0083687	0.3018288	0.2668820	0.4356768	0.2682926	0.4279863
	heckmanml_L_New	-0.8303721	0.1294735	-0.8302651	0.1310152	-0.9923375	0.0019679	-0.9917323	0.0023473
	Gheckmanml_L_New	-0.8649739	0.1373338	-0.8647736	0.1385397	-0.9924622	0.0019668	-0.9918700	0.0022959
	heckmanml_LOG_New	-0.6219982	0.2389182	-0.6239701	0.2393787	-0.9708853	0.0060742	-0.9700206	0.0066772
	Gheckmanml_LOG_New	-0.8347032	0.1828181	-0.8348146	0.1821333	-0.9924790	0.0019399	-0.9919138	0.0023767

5) SDG Single equations

Table A5-1

Errors Normally Distributed		k=0.5				k=1.25			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
SDG Single equations									
case 1	OLS	-0.209610	0.032099	-0.178106	0.051180	-0.201849	0.035994	-0.185044	0.051391
	Heckman 2SLS	0.002772	0.010608	0.001240	0.020374	0.003650	0.011066	0.002995	0.021289
	GHeckman 2SLS	0.002730	0.010779	0.001056	0.020605	0.003732	0.011108	0.002972	0.021514
	heckman2sls_Log	0.065578	0.041816	0.088171	0.058593	0.078555	0.052743	0.098025	0.063843
	Gheckman2sls_Log	0.065438	0.041486	0.088096	0.059287	0.077401	0.052361	0.096312	0.063350
	NNLS	-0.027348	0.010023	-0.031616	0.018485	-0.028538	0.010643	-0.033613	0.019195
	Gnnls	-0.031092	0.010951	-0.035549	0.019206	-0.031470	0.011635	-0.036773	0.019820
	nnls_C	0.030872	0.010055	0.033348	0.020026	0.071138	0.010807	0.078251	0.021493
	Gnnls_C	0.033241	0.010743	0.035576	0.020654	0.085454	0.018611	0.092499	0.027420
	PPML	-0.121526	0.013696	-0.123897	0.023785	-0.110729	0.014617	-0.118718	0.025064
	PPML_C	0.085155	0.017084	0.087718	0.030952	0.203966	0.017909	0.213294	0.035608
	Optimal_PPML	-0.208050	0.017534	-0.207966	0.024201	-0.192381	0.019460	-0.198311	0.026110
	Optimal_PPML_C	-0.152718	0.018773	-0.147563	0.025707	-0.086087	0.022856	-0.082829	0.031018
case 2	OLS	-0.162594	0.029406	-0.147640	0.061768	-0.178807	0.035901	-0.174478	0.064978
	Heckman 2SLS	-0.014937	0.037414	-0.018387	0.060832	-0.014304	0.039257	-0.017551	0.063040
	GHeckman 2SLS	-0.061925	0.023557	-0.068884	0.045100	-0.061945	0.027049	-0.069213	0.049959
	heckman2sls_Log	0.047124	0.053011	0.055083	0.073160	0.046366	0.057192	0.052031	0.076028
	Gheckman2sls_Log	0.047302	0.052699	0.055151	0.072925	0.047086	0.056828	0.052668	0.075313
	NNLS	-0.041590	0.034412	-0.048531	0.054339	-0.044916	0.036944	-0.052623	0.056839
	Gnnls	-0.130313	0.038230	-0.134604	0.047686	-0.138288	0.044837	-0.147483	0.055383
	nnls_C	0.029970	0.032122	0.030569	0.056424	0.067211	0.032935	0.072146	0.059206
	Gnnls_C	0.079662	0.022852	0.080365	0.047996	0.150488	0.033763	0.158170	0.057888
	PPML	-0.112361	0.020225	-0.119062	0.038861	-0.116658	0.023549	-0.127706	0.042572
	PPML_C	0.096657	0.021289	0.096662	0.046829	0.205958	0.023374	0.212657	0.052432
	Optimal_PPML	-0.154815	0.017135	-0.160723	0.033794	-0.159041	0.018728	-0.167425	0.036080
	Optimal_PPML_C	-0.085830	0.016852	-0.088313	0.034726	-0.036497	0.017991	-0.037351	0.038024
case 3	OLS	-0.097458	0.036871	-0.090018	0.073155	-0.140861	0.045369	-0.142495	0.081829
	Heckman 2SLS	0.013748	0.134028	0.013752	0.176147	0.013223	0.139739	0.013294	0.183565
	GHeckman 2SLS	-0.024973	0.038981	-0.025182	0.055888	-0.048362	0.047680	-0.050697	0.062914
	heckman2sls_Log	0.012717	0.067603	0.013913	0.089374	0.014688	0.090144	0.016657	0.105653
	Gheckman2sls_Log	0.011607	0.067733	0.012760	0.089340	0.014627	0.090407	0.016357	0.105805
	NNLS	-0.007755	0.105299	-0.014652	0.147805	-0.016302	0.112728	-0.024424	0.155316
	Gnnls	-0.061050	0.031977	-0.058642	0.052342	-0.091052	0.039365	-0.094136	0.060652
	nnls_C	0.044822	0.156767	0.040020	0.217665	0.077982	0.158468	0.077695	0.225973
	Gnnls_C	0.135869	0.067746	0.129963	0.100637	0.238594	0.095877	0.238475	0.125065
	PPML	-0.035508	0.046900	-0.039719	0.071364	-0.054047	0.054877	-0.062498	0.080243
	PPML_C	0.098594	0.063267	0.095479	0.107228	0.198622	0.066905	0.201850	0.116742
	Optimal_PPML	-0.020471	0.039472	-0.023553	0.061404	-0.035161	0.044003	-0.040301	0.066621
	Optimal_PPML_C	0.028134	0.036531	0.026637	0.061204	0.066647	0.037178	0.067755	0.063823
case 4	OLS	-0.178179	0.042678	0.048960	0.081330	-0.207733	0.052701	0.015174	0.089899
	Heckman 2SLS	0.009758	0.176397	0.295281	0.203195	0.010183	0.181614	0.297824	0.211219
	GHeckman 2SLS	-0.057416	0.058018	0.199657	0.074210	-0.077329	0.064525	0.184099	0.080755
	heckman2sls_Log	-0.047261	0.079050	0.119165	0.088047	-0.021788	0.106212	0.154017	0.110677
	Gheckman2sls_Log	-0.047575	0.079521	0.119296	0.088073	-0.022104	0.106581	0.154242	0.110398
	NNLS	-0.021391	0.139217	0.256561	0.166969	-0.027309	0.147513	0.250556	0.176249
	Gnnls	-0.145847	0.047504	0.115650	0.067014	-0.164749	0.053141	0.096438	0.073251
	nnls_C	0.054049	0.272773	0.052558	0.305227	0.081270	0.266670	0.090288	0.308170
	Gnnls_C	0.083692	0.072925	0.123127	0.128709	0.184153	0.095521	0.231189	0.147509
	PPML	-0.077133	0.055177	0.185574	0.076686	-0.088686	0.064138	0.173362	0.086501
	PPML_C	0.052432	0.097872	0.077404	0.148852	0.146683	0.102357	0.188374	0.157744
	Optimal_PPML	-0.072761	0.046456	0.207396	0.065271	-0.080931	0.051668	0.199926	0.071275
	Optimal_PPML_C	-0.020050	0.040686	0.244062	0.062300	0.022113	0.041694	0.291606	0.065628
case 5	OLS	-0.264000	0.042728	-0.240116	0.080010	-0.267044	0.052456	-0.259886	0.088884

	Heckman 2SLS	0.018273	0.128118	0.018539	0.172400	0.017263	0.134953	0.017858	0.180072
	GHeckman 2SLS	-0.021186	0.267953	-0.010039	0.165875	-0.027282	0.090899	-0.027415	0.102749
	heckman2sls_Log	-0.062141	0.069716	-0.050477	0.102582	-0.011085	0.096050	-0.003253	0.117000
	Gheckman2sls_Log	-0.061742	0.069681	-0.050162	0.102255	-0.010487	0.096287	-0.002884	0.116757
	NNLS	-0.029495	0.106378	-0.038350	0.146537	-0.033916	0.113335	-0.043698	0.154299
	Gnnls	-0.302273	0.187291	-0.271865	0.161800	-0.275647	0.150376	-0.260418	0.136717
	nnls_C	0.037883	0.157358	0.032268	0.217146	0.071950	0.158742	0.070582	0.225352
	Gnnls_C	0.069584	0.061659	0.070627	0.100871	0.190227	0.079252	0.191826	0.115273
	PPML	-0.111641	0.048026	-0.116650	0.070722	-0.111262	0.055487	-0.122140	0.079701
	PPML_C	0.059634	0.063798	0.057350	0.107328	0.167221	0.067024	0.168760	0.116745
	Optimal_PPML	-0.134642	0.039456	-0.135708	0.060426	-0.128251	0.044130	-0.134438	0.065582
	Optimal_PPML_C	-0.084267	0.036034	-0.081840	0.059355	-0.026391	0.038289	-0.024614	0.062931

Table A5-2

Errors Lognormally Distributed

		k=0.5				k=1.25			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
	SDG Single equations								
case 1	OLS	0.166734	0.051249	0.156000	0.058886	0.011376	0.059907	0.009033	0.069182
	Heckman 2SLS	0.000290	0.011386	-0.002056	0.020976	0.002703	0.013013	0.001817	0.022714
	GHeckman 2SLS	0.000244	0.011426	-0.002132	0.021139	0.002730	0.013041	0.001818	0.022797
	heckman2sls_Log	0.165859	0.059334	0.154695	0.071817	0.091313	0.064938	0.083099	0.080910
	Gheckman2sls_Log	0.060749	0.052588	0.060206	0.055900	0.051675	0.051787	0.050984	0.057500
	NNLS	-0.015237	0.010448	-0.018389	0.020811	-0.024494	0.011175	-0.028781	0.020500
	Gnnls	-0.015393	0.010424	-0.018590	0.020866	-0.024576	0.011137	-0.028875	0.020515
	nnls_C	0.036494	0.010310	0.039191	0.021521	0.073587	0.011055	0.081308	0.022060
	Gnnls_C	0.037744	0.011455	0.040291	0.022114	0.077921	0.016253	0.085510	0.025426
	PPML	-0.064338	0.024769	-0.065535	0.035166	-0.098656	0.030377	-0.104149	0.040662
	PPML_C	0.113716	0.022776	0.115949	0.037657	0.211753	0.026047	0.223780	0.042452
	Optimal_PPML	-0.216316	0.105348	-0.215830	0.117829	-0.243394	0.111016	-0.247537	0.123203
	Optimal_PPML_C	-0.162030	0.106751	-0.159961	0.122588	-0.134955	0.115637	-0.130700	0.132490
case 2	OLS	0.049644	0.040628	0.054181	0.057552	-0.056241	0.048577	-0.051303	0.069378
	Heckman 2SLS	0.001255	0.035536	-0.001179	0.062385	0.001764	0.038285	-0.000618	0.065560
	GHeckman 2SLS	-0.009183	0.023649	-0.013558	0.049297	-0.011494	0.028564	-0.016053	0.054378
	heckman2sls_Log	0.137940	0.047792	0.134746	0.064862	0.117344	0.055013	0.119646	0.074087
	Gheckman2sls_Log	0.128581	0.045369	0.127954	0.061521	0.114476	0.050405	0.117522	0.067722
	NNLS	-0.013660	0.034092	-0.018900	0.060016	-0.027233	0.037103	-0.034234	0.059593
	Gnnls	-0.048279	0.026306	-0.054107	0.047738	-0.075525	0.033552	-0.083941	0.054863
	nnls_C	0.037829	0.031624	0.038455	0.060755	0.071812	0.032793	0.076780	0.060324
	Gnnls_C	0.081719	0.026509	0.081174	0.051939	0.134811	0.034382	0.143076	0.058483
	PPML	-0.059724	0.024228	-0.064380	0.048209	-0.096925	0.028928	-0.105011	0.051788
	PPML_C	0.117201	0.022692	0.115939	0.051133	0.212527	0.026112	0.221895	0.055246
	Optimal_PPML	-0.164694	0.049252	-0.167903	0.071765	-0.195985	0.052713	-0.203228	0.075405
	Optimal_PPML_C	-0.107552	0.050141	-0.109446	0.073858	-0.080750	0.055389	-0.080841	0.080396
case 3	OLS	-0.115802	0.035446	-0.104963	0.070129	-0.185978	0.045457	-0.182788	0.085320
	Heckman 2SLS	0.047896	0.349383	0.030236	0.298963	0.045924	0.351654	0.028510	0.306683
	GHeckman 2SLS	-0.040969	0.056242	-0.050995	0.092073	-0.070919	0.070382	-0.084294	0.105459
	heckman2sls_Log	0.007545	0.046710	0.007803	0.073868	0.013564	0.060844	0.015397	0.089640
	Gheckman2sls_Log	0.007887	0.046595	0.008415	0.073861	0.014245	0.061179	0.016223	0.089694
	NNLS	-0.003547	0.159314	-0.018679	0.214639	-0.018906	0.170412	-0.037168	0.225042
	Gnnls	-0.099345	0.055289	-0.098126	0.091178	-0.170685	0.074673	-0.174198	0.109170
	nnls_C	0.048663	0.149161	0.038251	0.212786	0.091337	0.198495	0.081692	0.239368
	Gnnls_C	0.143960	0.078773	0.134700	0.108930	0.238694	0.104029	0.237772	0.131575
	PPML	-0.057709	0.071303	-0.065064	0.107390	-0.096144	0.083156	-0.110063	0.121044
	PPML_C	0.118087	0.061540	0.112859	0.106497	0.212765	0.065189	0.215622	0.116961
	Optimal_PPML	-0.052693	0.089334	-0.055602	0.153526	-0.087840	0.098593	-0.093573	0.162722
	Optimal_PPML_C	0.003751	0.082186	0.001852	0.149198	0.028363	0.084157	0.028332	0.154009
case 4	OLS	-0.029225	0.052632	-0.281717	0.094019	-0.140715	0.067418	-0.401432	0.114319
	Heckman 2SLS	0.093097	0.544207	0.056481	0.427599	0.089549	0.545221	0.054255	0.434528
	GHeckman 2SLS	-0.052001	0.088895	-0.058287	0.136940	-0.082514	0.112665	-0.085961	0.152480
	heckman2sls_Log	0.072186	0.067796	-0.189140	0.097870	0.061012	0.084193	-0.203198	0.116662
	Gheckman2sls_Log	0.073252	0.066549	-0.187466	0.097105	0.065199	0.084396	-0.198034	0.116346
	NNLS	-0.009338	0.229345	-0.022135	0.286642	-0.029598	0.247347	-0.045379	0.297235
	Gnnls	-0.139782	0.107057	-0.126574	0.146295	-0.242296	0.159999	-0.221778	0.182288
	nnls_C	0.041063	0.207323	0.031265	0.267677	0.076727	0.223887	0.073655	0.281307
	Gnnls_C	0.108100	0.099063	0.107178	0.139522	0.185153	0.121691	0.197202	0.162899
	PPML	-0.078905	0.108589	-0.082686	0.148948	-0.134554	0.127810	-0.139274	0.166984
	PPML_C	0.100822	0.092026	0.099473	0.143466	0.187077	0.097313	0.198648	0.154309
	Optimal_PPML	-0.140938	0.149251	0.085708	0.242881	-0.186393	0.162918	0.038689	0.251688
	Optimal_PPML_C	-0.083788	0.138764	0.146904	0.239437	-0.067804	0.141590	0.169290	0.244438
case 5	OLS	0.082313	0.055698	0.074478	0.084824	-0.067865	0.068883	-0.072013	0.104536
	Heckman 2SLS	0.050171	0.358768	0.032276	0.303604	0.050166	0.358977	0.033185	0.310259
	GHeckman 2SLS	-0.034175	0.073007	-0.044087	0.115777	-0.064013	0.105415	-0.072968	0.137927
	heckman2sls_Log	0.108803	0.065855	0.090467	0.093642	0.056001	0.082668	0.043894	0.111704
	Gheckman2sls_Log	0.082633	0.064159	0.072847	0.092484	0.060160	0.078473	0.055076	0.106677

	NNLS	0.000996	0.204957	-0.016955	0.236390	-0.023926	0.171275	-0.041510	0.225496
	Gnnls	-0.126998	0.092059	-0.124622	0.127504	-0.250682	0.242606	-0.235877	0.255517
	nnls_C	0.053556	0.195005	0.040447	0.233175	0.079593	0.153475	0.072149	0.220221
	Gnnls_C	0.120459	0.079400	0.113463	0.114221	0.207143	0.108367	0.144585	1.386992
	PPML	-0.075535	0.074073	-0.081826	0.111602	-0.126590	0.087714	-0.139934	0.126485
	PPML_C	0.104198	0.063369	0.099324	0.110006	0.193281	0.067887	0.196142	0.120722
	Optimal_PPML	-0.167999	0.125211	-0.167608	0.182948	-0.206992	0.132899	-0.211042	0.190834
	Optimal_PPML_C	-0.112202	0.121621	-0.110935	0.181977	-0.092810	0.126851	-0.089931	0.188767

6) SDG ETtobit equations (k=1)

Table A6

		Errors Lognormally Distributed				Errors Normally Distributed			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
SDG ETtobit equations									
case 1	Etobit	-0.142799	0.021151	-0.147914	0.026957	-0.124229	0.016296	-0.128593	0.022156
	GETtobit	-0.142942	0.021150	-0.148093	0.026918	-0.124344	0.016297	-0.128715	0.022165
	n_ETtobit	-0.662000	0.092323	-0.652870	0.096850	-0.660081	0.026641	-0.648705	0.027303
	Gn_ETtobit	-0.143028	0.021160	-0.148156	0.026973	-0.124444	0.016296	-0.128830	0.022168
	Ettobit_L	-0.157308	0.015995	-0.162616	0.022572	-0.222390	0.014985	-0.229586	0.021633
	GETtobit_L	-0.159062	0.016650	-0.164404	0.023054	-0.234203	0.018353	-0.241667	0.024865
	Ettobit_LOG	-0.513547	0.023576	-0.509309	0.028648	-0.634041	0.012584	-0.631030	0.017323
	GEttobit_LOG	-0.159176	0.016619	-0.164520	0.023025	-0.234090	0.018283	-0.241551	0.024807
case 2	Etobit								
	GETtobit								
	n_ETtobit								
	Gn_ETtobit								
	Ettobit_L								
	GETtobit_L								
	Ettobit_LOG								
	GEttobit_LOG								
case 3	Etobit	-0.451147	0.138165	-0.467634	0.147206	-0.371174	0.155275	-0.383105	0.169576
	GETtobit	-0.027887	0.106662	-0.048050	0.146760	0.003943	0.067287	-0.004942	0.078872
	n_ETtobit	-0.280461	0.077496	-0.263577	0.121335	-0.222890	0.032275	-0.200595	0.044702
	Gn_ETtobit	-0.027787	0.106628	-0.047955	0.146725	0.003967	0.067278	-0.004919	0.078863
	Ettobit_L	-0.345912	0.032472	-0.355190	0.050224	-0.586325	0.044753	-0.599536	0.071320
	GETtobit_L	-0.208809	0.041042	-0.220668	0.057502	-0.477611	0.032884	-0.480958	0.049083
	Ettobit_LOG	-0.255341	0.040575	-0.258500	0.056207	-0.478990	0.032133	-0.481836	0.048864
	GEttobit_LOG	-0.208638	0.041054	-0.220504	0.057510	-0.477793	0.032853	-0.480991	0.049016
case 4	Etobit								
	GETtobit								
	n_ETtobit								
	Gn_ETtobit								
	Ettobit_L								
	GETtobit_L								
	Ettobit_LOG								
	GEttobit_LOG								
case 5	Etobit	-0.465142	0.136856	-0.482124	0.145368	-0.400363	0.153059	-0.412453	0.165504
	GETtobit	-0.085941	0.120118	-0.110345	0.159055	-0.106674	0.048147	-0.121088	0.060446
	n_ETtobit	-0.450239	0.130089	-0.433638	0.165090	-0.445622	0.043269	-0.416878	0.049018
	Gn_ETtobit	-0.085647	0.120160	-0.110054	0.159105	-0.106306	0.048153	-0.120720	0.060460
	Ettobit_L	-0.297116	0.034445	-0.305223	0.053575	-0.613527	0.044098	-0.624711	0.067935
	GETtobit_L	-0.096317	0.042395	-0.111280	0.062967	-0.553224	0.034903	-0.557041	0.047499
	Ettobit_LOG	-0.162150	0.046450	-0.165541	0.064500	-0.560453	0.032268	-0.562735	0.045716
	GEttobit_LOG	-0.095820	0.042478	-0.110854	0.063078	-0.553945	0.034796	-0.557549	0.047376

7) SDG Standard Heckman equations (k=1.25)

Table A7

		Errors Lognormally Distributed				Errors Normally Distributed			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
SDG Old Heckman equations									
case 1	heckmanml (heckman when alpha =4 and beta=3 (one more variable in the selection % equation))	0.006245	0.018937	0.005095	0.026809	0.005339	0.011551	0.003825	0.021386
	Gheckmanml	0.006258	0.019000	0.005167	0.026879	0.004903	0.011655	0.003131	0.021740
	n_heckmanml	-0.494049	0.121006	-0.511605	0.173246	-0.368773	0.047064	-0.356524	0.041617
	Gn_heckmanml	0.006250	0.019019	0.005159	0.026889	0.004889	0.011649	0.003120	0.021745
	heckmanml_L	-0.004871	0.010694	-0.007213	0.020187	0.006064	0.012744	0.004745	0.027650
	Gheckmanml_L	-0.004175	0.010749	-0.006534	0.020445	0.006951	0.012626	0.005578	0.027808
	heckmanml_LOG	0.022920	0.046527	0.022594	0.062253	-0.142628	0.029705	-0.125863	0.048424
	Gheckmanml_LOG	-0.005354	0.011208	-0.007456	0.020449	-0.019421	0.012151	-0.022190	0.027569
case 2	heckmanml (heckman when alpha =4 and beta=3 (one more variable in the selection % equation))	0.029330	0.045921	0.028872	0.070881	0.015117	0.043452	0.014471	0.067993
	Gheckmanml	0.016797	0.041118	0.011947	0.069545	-0.010108	0.033553	-0.014698	0.051361
	n_heckmanml	-0.412435	0.079344	-0.420038	0.120937	-0.257712	0.023704	-0.259976	0.038525
	Gn_heckmanml	0.016836	0.041141	0.011990	0.069550	-0.009920	0.033526	-0.014503	0.051343
	heckmanml_L	-0.211181	0.059877	-0.224911	0.065973	-0.494229	0.043579	-0.510311	0.062137
	Gheckmanml_L	-0.011399	0.027023	-0.015942	0.048682	-0.454833	0.051931	-0.464402	0.069261
	heckmanml_LOG	-0.009996	0.036359	-0.003244	0.060623	-0.105459	0.033689	-0.104708	0.061407
	Gheckmanml_LOG	-0.035450	0.027083	-0.038962	0.047494	-0.129682	0.033918	-0.139519	0.071394
case 3	heckmanml (heckman when alpha =4 and beta=3 (one more variable in the selection % equation))	0.115246	0.258101	0.115309	0.313127	0.076581	0.210539	0.085089	0.266613
	Gheckmanml	0.190556	0.225794	0.176457	0.279832	0.020573	0.072555	0.014505	0.077618
	n_heckmanml	-0.157807	0.090277	-0.168064	0.168160	-0.017554	0.032560	-0.021325	0.052386
	Gn_heckmanml	0.146219	0.226361	0.130954	0.291963	0.014985	0.062278	0.009348	0.070326
	heckmanml_L	-0.608493	0.028624	-0.613757	0.050500	-0.581961	0.043072	-0.597188	0.064591
	Gheckmanml_L	-0.051081	0.148603	-0.059717	0.169075	-0.341580	0.039565	-0.346587	0.070554
	heckmanml_LOG	-0.035318	0.061359	-0.035338	0.099460	-0.263816	0.308468	-0.259230	0.297677
	Gheckmanml_LOG	-0.076905	0.058689	-0.076972	0.097657	-0.346435	0.307930	-0.342195	0.298437
case 4	heckmanml (heckman when alpha =4 and beta=3 (one more variable in the selection % equation))	0.150189	0.341933	0.171518	0.430867	0.086475	0.264161	0.393921	0.320980
	Gheckmanml	0.118448	0.315258	0.425088	0.395321	0.235792	0.209888	0.572379	0.261034
	n_heckmanml	-0.322752	0.139636	-0.058975	0.271910	-0.118103	0.041027	0.154573	0.059056
	Gn_heckmanml	0.109317	0.313154	0.520381	2.184166	0.121416	0.216483	0.433220	0.264022
	heckmanml_L	-0.593876	0.037194	-0.779883	0.059318	-0.614486	0.044602	-0.488507	0.062915
	Gheckmanml_L	-0.134504	0.237685	-0.392119	0.224702	-0.422918	0.042947	-0.112731	0.076223
	heckmanml_LOG	-0.004964	0.081512	-0.266626	0.123634	-0.160600	0.170289	0.054467	0.158280
	Gheckmanml_LOG	-0.136228	0.080913	-0.404532	0.127969	-0.220190	0.196787	0.005186	0.172363
case 5	heckmanml (heckman when alpha =4 and beta=3 (one more variable in the selection % equation))	0.141394	0.245994	0.145994	0.305010	0.128960	0.181104	0.140961	0.245920
	Gheckmanml	0.124095	0.178250	0.120030	0.285527	0.180142	0.088376	0.177448	0.111439
	n_heckmanml	-0.392356	0.140935	-0.400901	0.231391	-0.244287	0.062713	-0.226180	0.064933
	Gn_heckmanml	0.123705	0.181577	0.113859	0.275992	0.176072	0.092374	0.173571	0.115284
	heckmanml_L	-0.542954	0.036503	-0.548133	0.057704	-0.658978	0.039847	-0.666152	0.059669
	Gheckmanml_L	-0.058748	0.098408	-0.065990	0.121886	-0.518893	0.044960	-0.505962	0.075825
	heckmanml_LOG	-0.019102	0.057872	-0.025034	0.105154	-0.194292	0.167131	-0.186034	0.180825
	Gheckmanml_LOG	-0.129573	0.054958	-0.134338	0.090636	-0.251072	0.212484	-0.241226	0.216729

8) SDG Heckman_New equations

Table A8-1

Errors Normally Distributed

		k=0.5				k=1.25			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
	SDG Heckman_New equations								
case 1	heckmanml_New	0.003658	0.010565	0.002980	0.020298	0.002255	0.011133	0.001485	0.020309
	Gheckmanml_New	0.003259	0.010658	0.002399	0.020534	0.001944	0.011194	0.000954	0.020550
	n_heckmanml_New	-0.401001	0.035977	-0.380622	0.033310	-0.372314	0.045408	-0.362259	0.042329
	Gn_heckmanml_New	0.003245	0.010658	0.002380	0.020537	0.001886	0.011202	0.000880	0.020529
	heckmanml_L_New	-0.013549	0.011056	-0.013125	0.027175	-0.001245	0.012173	-0.002319	0.025535
	Gheckmanml_L_New	-0.013594	0.011129	-0.013116	0.027311	-0.000717	0.012000	-0.001840	0.025691
	heckmanml_LOG_New	-0.173036	0.027196	-0.151052	0.043740	-0.158315	0.030228	-0.146115	0.041295
	Gheckmanml_LOG_New	-0.013619	0.011229	-0.013121	0.027259	-0.003980	0.050528	-0.004579	0.052224
case 2	heckmanml_New	0.062254	0.040021	0.068179	0.065588	0.031200	0.042271	0.032776	0.062115
	Gheckmanml_New	0.000391	0.031929	-0.003009	0.048539	-0.003820	0.033522	-0.006851	0.048460
	n_heckmanml_New	-0.266988	0.020523	-0.262651	0.033637	-0.260036	0.023001	-0.258915	0.034546
	Gn_heckmanml_New	0.000533	0.031891	-0.002868	0.048518	-0.003628	0.033494	-0.006647	0.048465
	heckmanml_L_New	-0.648895	0.108652	-0.635025	0.110319	-0.505764	0.173958	-0.504002	0.171417
	Gheckmanml_L_New	-0.113583	0.172619	-0.112028	0.177034	-0.004681	0.105640	-0.011904	0.114685
	heckmanml_LOG_New	-0.149526	0.036250	-0.141203	0.059025	-0.128374	0.040651	-0.125686	0.056649
	Gheckmanml_LOG_New	-0.074222	0.132171	-0.076909	0.138217	-0.003669	0.096381	-0.011292	0.107138
case 3	heckmanml_New	1.206839	1.739992	0.502445	1.530139	0.337733	1.282930	0.064340	1.167908
	Gheckmanml_New	-0.066414	0.025797	-0.066652	0.047337	-0.089081	0.030344	-0.083207	0.055135
	n_heckmanml_New	-0.067914	0.025129	-0.067820	0.046904	-0.090023	0.029667	-0.083951	0.054479
	Gn_heckmanml_New	0.313422	0.443931	0.313011	0.447075	-0.085641	0.067240	-0.079634	0.084587
	heckmanml_L_New	-0.985941	0.005204	-0.984707	0.005830	-0.965523	0.014279	-0.959967	0.074811
	Gheckmanml_L_New	-0.986319	0.004987	-0.985125	0.005476	-0.965351	0.014812	-0.964072	0.015474
	heckmanml_LOG_New	-0.932193	0.009356	-0.932440	0.010007	-0.947086	0.008487	-0.947079	0.009026
	Gheckmanml_LOG_New	-0.986491	0.004939	-0.985310	0.005501	-0.963811	0.062892	-0.971452	0.093170
case 4	heckmanml_New	-0.216786	0.662983	-1.230724	0.845225	-0.268464	0.803449	-0.822075	0.954532
	Gheckmanml_New	-0.078502	0.075406	0.198127	0.107271	-0.136926	0.042357	0.129158	0.071033
	n_heckmanml_New	-0.160666	0.029349	0.105158	0.053211	-0.170004	0.033316	0.093335	0.062402
	Gn_heckmanml_New	0.251132	0.451074	0.629080	0.581399	-0.134285	0.076245	0.132658	0.111113
	heckmanml_L_New	-0.994215	0.001475	-0.993798	0.001816	-0.991203	0.002732	-0.990613	0.002934
	Gheckmanml_L_New	-0.994228	0.001474	-0.993966	0.002034	-0.991227	0.002713	-0.990649	0.002913
	heckmanml_LOG_New	-0.976319	0.003994	-0.971074	0.005096	-0.980042	0.004042	-0.975751	0.004925
	Gheckmanml_LOG_New	-0.994243	0.001460	-0.993845	0.001791	-0.991434	0.002785	-0.990877	0.002981
case 5	heckmanml_New	1.040751	1.622837	0.342244	1.392660	0.046772	1.063652	-0.160319	0.979606
	Gheckmanml_New	-0.095608	0.047491	-0.096153	0.066187	-0.164716	0.049458	-0.158580	0.071118
	n_heckmanml_New	-0.295515	0.041884	-0.271668	0.053931	-0.277585	0.047118	-0.259642	0.060090
	Gn_heckmanml_New	0.029389	0.347163	0.025400	0.335824	-0.162268	0.085215	-0.156105	0.101214
	heckmanml_L_New	-0.986221	0.004458	-0.985082	0.004904	-0.967488	0.036821	-0.967609	0.017732
	Gheckmanml_L_New	-0.985931	0.004712	-0.984708	0.005798	-0.968654	0.011873	-0.967558	0.012219
	heckmanml_LOG_New	-0.957116	0.007062	-0.955884	0.007712	-0.958134	0.008137	-0.957105	0.008817
	Gheckmanml_LOG_New	-0.986487	0.004372	-0.985426	0.004821	-0.969509	0.012480	-0.968442	0.012851

Table A8-2

Errors Lognormally Distributed		k=0.5				k=1.25			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
	SDG Heckman_New equations								
case 1	heckmanml_New	0.003659	0.019902	0.003793	0.028687	0.000420	0.013802	0.000371	0.022971
	Gheckmanml_New	0.003614	0.019899	0.003699	0.028728	0.000391	0.013789	0.000310	0.022983
	n_heckmanml_New	-0.455881	0.109233	-0.444804	0.130779	-0.491683	0.114607	-0.487355	0.128146
	Gn_heckmanml_New	0.003590	0.019860	0.003675	0.028688	0.000387	0.013797	0.000305	0.022993
	heckmanml_L_New	-0.007038	0.010062	-0.008119	0.018715	-0.006502	0.010613	-0.008110	0.019339
	Gheckmanml_L_New	-0.006273	0.010156	-0.007302	0.018813	-0.005858	0.010643	-0.007460	0.019541
	heckmanml_LOG_New	0.167911	0.049449	0.164459	0.057907	0.030637	0.051297	0.034364	0.059814
	Gheckmanml_LOG_New	-0.006229	0.010154	-0.007249	0.018815	-0.005824	0.010641	-0.007418	0.019542
case 2	heckmanml_New	0.058060	0.044364	0.063883	0.067215	0.021028	0.040999	0.021966	0.058932
	Gheckmanml_New	0.034560	0.049461	0.033066	0.067075	0.013414	0.036046	0.011874	0.054840
	n_heckmanml_New	-0.361354	0.075991	-0.355352	0.094490	-0.408845	0.078291	-0.404606	0.085653
	Gn_heckmanml_New	0.034452	0.049401	0.032954	0.067058	0.013451	0.036067	0.011912	0.054850
	heckmanml_L_New	-0.206770	0.034229	-0.210566	0.040845	-0.165651	0.039972	-0.172496	0.047073
	Gheckmanml_L_New	-0.020803	0.021519	-0.020373	0.038728	-0.017508	0.023506	-0.019200	0.040856
	heckmanml_LOG_New	0.070871	0.037839	0.076617	0.051730	-0.007898	0.039152	-0.001992	0.053614
	Gheckmanml_LOG_New	-0.020832	0.021515	-0.020402	0.038720	-0.017513	0.023510	-0.019207	0.040853
case 3	heckmanml_New	0.446555	0.222556	0.484405	0.313042	0.027261	0.131575	0.028120	0.139783
	Gheckmanml_New	0.370786	0.139650	0.359418	0.203238	0.131020	0.136771	0.120970	0.159920
	n_heckmanml_New	-0.091448	0.078477	-0.094251	0.124818	-0.156948	0.077345	-0.158332	0.108180
	Gn_heckmanml_New	0.332572	0.193901	0.321278	0.245776	0.115497	0.145800	0.105546	0.174453
	heckmanml_L_New	-0.587034	0.024406	-0.577654	0.032705	-0.621261	0.028221	-0.614477	0.034967
	Gheckmanml_L_New	-0.051191	0.060661	-0.045521	0.076294	-0.030254	0.045434	-0.028170	0.065594
	heckmanml_LOG_New	-0.010827	0.041902	-0.004559	0.061926	-0.020565	0.049572	-0.015619	0.066638
	Gheckmanml_LOG_New	-0.047624	0.050393	-0.041741	0.068498	-0.030129	0.045395	-0.028022	0.065559
case 4	heckmanml_New	0.565311	0.226829	0.632866	0.332912	-0.006916	0.157018	-0.005237	0.160805
	Gheckmanml_New	0.428123	0.268322	0.733318	0.381578	0.090465	0.192267	0.237206	0.248128
	n_heckmanml_New	-0.225244	0.124744	-0.009147	0.240377	-0.280270	0.127112	-0.153866	0.198069
	Gn_heckmanml_New	0.432421	0.299161	0.769316	0.534979	0.084898	0.196183	0.224868	0.288065
	heckmanml_L_New	-0.579681	0.031873	-0.616430	0.034591	-0.618097	0.034921	-0.639161	0.036244
	Gheckmanml_L_New	-0.072432	0.120270	-0.167715	0.114755	-0.044711	0.086893	-0.104748	0.091961
	heckmanml_LOG_New	0.017799	0.059648	-0.067783	0.075135	-0.017747	0.067886	-0.059343	0.079329
	Gheckmanml_LOG_New	-0.062805	0.102660	-0.158983	0.103029	-0.039351	0.067982	-0.099284	0.078622
case 5	heckmanml_New	0.436780	0.223880	0.469458	0.303406	0.014134	0.128018	0.014706	0.134897
	Gheckmanml_New	0.353395	0.196440	0.349202	0.266796	0.093860	0.126534	0.086710	0.148201
	n_heckmanml_New	-0.315333	0.134341	-0.314356	0.195225	-0.383851	0.132411	-0.378951	0.144677
	Gn_heckmanml_New	0.311011	0.233593	0.302767	0.292257	0.094510	0.126730	0.087471	0.147941
	heckmanml_L_New	-0.496321	0.032253	-0.488606	0.039658	-0.543351	0.038665	-0.536921	0.045205
	Gheckmanml_L_New	-0.038430	0.047232	-0.034832	0.064317	-0.036654	0.050731	-0.035474	0.066676
	heckmanml_LOG_New	0.111666	0.056409	0.111546	0.071987	0.014908	0.062148	0.017191	0.075613
	Gheckmanml_LOG_New	-0.039152	0.049477	-0.035571	0.065640	-0.036615	0.050724	-0.035436	0.066671

9) Third DG Heckman equations (k=1)

Table 9

		Errors Lognormally Distributed				Errors Normally Distributed			
		b1		b2		b1		b2	
		bias	std	bias	std	bias	std	bias	std
Third DG Heckman equations									
case 1	heckmanml_New	0.004017	0.022358	0.004195	0.030869	0.003538	0.010496	0.003178	0.020426
	Gheckmanml_New	0.004016	0.022471	0.004174	0.031082	0.003151	0.010519	0.002621	0.020748
	n_heckmanml_New	-0.451445	0.117874	-0.445835	0.133210	-0.398256	0.036427	-0.379662	0.036943
	Gn_heckmanml_New	0.003217	0.023274	0.003457	0.031833	0.003134	0.010520	0.002600	0.020746
	heckmanml_L_New	-0.007006	0.010260	-0.008534	0.018712	-0.013142	0.011284	-0.013230	0.024951
	Gheckmanml_L_New	-0.006167	0.010307	-0.007730	0.018920	-0.013280	0.011480	-0.013187	0.025032
	heckmanml_LOG_New	0.166197	0.048134	0.163099	0.056918	-0.170060	0.027899	-0.157160	0.039260
	Gheckmanml_LOG_New	-0.006120	0.010301	-0.007674	0.018917	-0.013184	0.011671	-0.013207	0.025126
case 2	heckmanml_New	0.057609	0.045755	0.065432	0.069339	0.061916	0.040290	0.068639	0.065867
	Gheckmanml_New	0.031113	0.047918	0.029330	0.072212	0.001188	0.031239	-0.001056	0.046854
	n_heckmanml_New	-0.359891	0.080321	-0.357282	0.095277	-0.265703	0.019823	-0.262063	0.032820
	Gn_heckmanml_New	0.031036	0.047924	0.029250	0.072218	0.001334	0.031192	-0.000908	0.046859
	heckmanml_L_New	-0.206414	0.034365	-0.210950	0.040712	-0.650036	0.110492	-0.640020	0.113736
	Gheckmanml_L_New	-0.019160	0.020186	-0.020586	0.038150	-0.169056	0.224907	-0.172692	0.220246
	heckmanml_LOG_New	0.069518	0.037316	0.072547	0.051703	-0.147794	0.036877	-0.144697	0.057208
	Gheckmanml_LOG_New	-0.019195	0.020187	-0.020621	0.038149	-0.112725	0.201151	-0.117658	0.201398
case 3	heckmanml_New	0.443617	0.188988	0.497312	0.294359	3.640689	4.380717	7.186315	23.707644
	Gheckmanml_New	0.400216	0.149246	0.385686	0.192757	-0.061971	0.026769	-0.060535	0.045345
	n_heckmanml_New	-0.089806	0.078477	-0.096886	0.125240	-0.063925	0.025816	-0.061966	0.045232
	Gn_heckmanml_New	0.307191	0.233155	0.291715	0.256731	-0.057667	0.065552	-0.056094	0.078344
	heckmanml_L_New	-0.587564	0.026323	-0.579200	0.032754	-0.986598	0.005108	-0.985031	0.005821
	Gheckmanml_L_New	-0.056790	0.073952	-0.054020	0.087252	-0.986475	0.012517	-0.984862	0.014257
	heckmanml_LOG_New	-0.013059	0.041030	-0.009455	0.061850	-0.931511	0.009942	-0.931562	0.010731
	Gheckmanml_LOG_New	-0.056856	0.074103	-0.054055	0.087408	-0.986588	0.012429	-0.984966	0.014170
case 4	heckmanml_New	0.554571	0.222700	0.622611	0.328915	0.172674	1.413135	-2.835240	4.409085
	Gheckmanml_New	0.442377	0.267335	0.756016	0.431469	-0.067930	0.089128	0.212507	0.113621
	n_heckmanml_New	-0.219220	0.134523	-0.004184	0.250679	-0.157699	0.030900	0.111143	0.052037
	Gn_heckmanml_New	0.347679	0.322370	0.642027	0.497115	0.092817	0.387264	0.429613	0.501591
	heckmanml_L_New	-0.578919	0.033040	-0.618900	0.032561	-0.994483	0.001382	-0.994112	0.001738
	Gheckmanml_L_New	-0.069058	0.120339	-0.172195	0.114166	-0.994546	0.001408	-0.994412	0.002039
	heckmanml_LOG_New	0.018459	0.056954	-0.075575	0.071321	-0.976205	0.004249	-0.970774	0.005587
	Gheckmanml_LOG_New	-0.066944	0.117164	-0.170323	0.111690	-0.994514	0.001367	-0.994164	0.001712
case 5	heckmanml_New	0.434083	0.192660	0.481991	0.292560	2.794524	4.174175	2.276350	7.631794
	Gheckmanml_New	0.397521	0.186292	0.393441	0.278725	-0.090488	0.048818	-0.087273	0.065709
	n_heckmanml_New	-0.306056	0.147258	-0.309007	0.206701	-0.292004	0.044276	-0.264456	0.056083
	Gn_heckmanml_New	0.317301	0.246939	0.308937	0.310758	-0.090490	0.048825	-0.087085	0.065631
	heckmanml_L_New	-0.496789	0.033530	-0.490046	0.039544	-0.987106	0.003762	-0.985749	0.004575
	Gheckmanml_L_New	-0.039594	0.051521	-0.038265	0.068246	-0.986865	0.004081	-0.985613	0.004822
	heckmanml_LOG_New	0.108825	0.055086	0.107846	0.072352	-0.957380	0.006601	-0.955701	0.007893
	Gheckmanml_LOG_New	-0.038329	0.046834	-0.036903	0.066195	-0.987300	0.003670	-0.985968	0.004495