CATPRN
Canadian Agricultural Trade Policy And Competitiveness Research Network

DOES THE GRAVITY MODEL SUFFER FROM SELECTION BIAS?

CATPRN Working Paper 2010-01
January 2010

Zahoor Haq
Department of Agricultural Economics
NWFP Agricultural University
Peshawar, Pakistan

Karl Meilke and John Cranfield
Department of Food, Agricultural and Resource Economics
University of Guelph

http://www.catrade.org

Funding for this project was provided by the Canadian Agricultural Trade Policy and Competitiveness Research Network (CATPRN) and the Ontario Ministry of Food, Agriculture and Rural Affairs. CATPRN is funded by Agriculture and Agri-Food Canada. The views in paper are those of the authors and should not be attributed to the funding agencies.

Abstract

When analyzing bilateral trade flow data, zero trade flows are quite common and problematic when a gravity equation is estimated with a log-linear functional form. This has caused many researchers to either ignore the zero trade flows or to replace zero with a small positive number. Both of these actions bias the resulting parameter estimates of the gravity equation. In this study we correct for this misspecification by using the Heckman selection model to estimate the bilateral trade flows for 46 agrifood products, for the period 1990 to 2000, for 52 countries. In our sample, selection bias rarely affects the signs of variables but often has a substantial effect on the magnitude, statistical significance and economic interpretation of the marginal effects. Hence, treating zero trade flows properly is important from both a statistical and an economics perspective.

Keywords: Gravity model; selection bias; Agrifood Trade; Heckman Selection Model; marginal effects

1.0 INTRODUCTION

The gravity equation is an important tool in the empirical analysis of international trading relationships and has been used to investigate the effects of trade costs, regional trade agreements, national borders, foreign direct investment and other trade related policies on the volume of trade. Indeed, the gravity model has become "... the workhorse for empirical studies ..." (Eichengreen and Irwin, 1997 p. 33). In its early inception, Tinbergen (1962) argued that bilateral trade flows are proportional to the product of the economic size of trading partners and the measures of "trade resistance" between them. Trade resistance was measured by geographic distance and dummy variables were used to account for common borders and Commonwealth membership. Later the gravity model was extended to include variables to account for common language, colonizers, religion, common currency and so on. Subsequently, Anderson (1979), Bergstrand (1985), Helpman and Krugman (1985) and Anderson and van Wincoop (2003) provided the theoretical foundation for the gravity equation while Bergstrand (1989 and 1990) extended this work and provided the justification for the generalized gravity equation. In comparison to the gravity equation the generalized gravity equation includes per capita income of trading partners as explanatory variables for bilateral trade flows.

With the theoretical foundations of the gravity equation no longer an issue attention has shifted to a number of empirical issues, for example, the specification and estimation of gravity equations using panel data and the interpretation of the effect of distance on patterns of bilateral trade (Buch *et al*., 2004, Egger, 2000, Egger and Pfaffermayr, 2003, and Matyas, 1997). In addition, Anderson and van Wincoop (2003) question the absence of multilateral resistance (MR) terms—that is the non-linear endogenous price terms for both exporting and importing countries—from the traditional cross-section gravity equation. Three approaches have been used to tackle this problem. Bergstrand (1985, 1989) and Baier and Bergstrand (2001) used available price indices to approximate the MR terms; however, the available price indices are poor proxies for the value of traded goods, and this is especially true for agrifood products which is our focus. The second approach to account for the missing price terms is to use country-specific fixed effects. This approach is very popular and yields unbiased bilateral trade estimates (Bergstrand et al., 2007). However Baier and Bergstrand (2009) show that although fixed-effects estimation is easy, it has at least two drawbacks. First, it is not possible to generate region or pair-specific comparative statistics without the underlying system of structural equations. Second, many explanatory variables are subsumed into the fixed effects and hence fixed effects preclude direct estimation of partial effects associated with, say, prices. For example, a gravity equation estimated with exporter and importer fixed effects using one year's data on bilateral trade subsumes variables like population in the fixed effects. However, in a cross-section study using more than one year's data, variables like population are time variant and can be included directly. A third approach to handle MR terms involves estimating a set of non-linear price equations under the assumption of symmetric bilateral trade costs using a non-linear estimation procedure; which then generates multilateral price terms before and after any counterfactual experiment. While this approach provides unbiased estimates and general equilibrium comparative statics, it does so under the symmetric bilateral trade cost assumption (Bergstrand et al., 2007). Consequently, the estimation of gravity equations using region and time specific fixed effects is more popular among trade economists.

Another potential problem with the gravity equation is that GDP includes the value of net exports making GDP endogenous to bilateral trade flows. However, endogeneity of GDP is generally ignored in the empirical literature, because net trade is a small proportion of GDP.

Frankel (1997) reported that coefficient estimates in gravity equations change little after accounting for the endogeneity of GDP using an instrumental variables approach. Recently, Baier and Bergstrand (2007) raise the issue of the endogeneity of binary variables representing free-trade agreements (FTAs). They argue that a policy variable such as an FTA is not an exogenous variable but rather endogenous and needs to be represented by a proxy in econometric analysis. However, this issue is of key importance only when the impacts of FTAs are investigated.

In this paper we focus on another important issue in estimating gravity equations, namely selection bias — a largely ignored but important empirical problem. Selection bias occurs when a subset of the data is systematically excluded due to a particular attribute. The exclusion of the subset can influence the statistical significance of test results and produce biased findings. In estimating the gravity equation it is common to omit zero trade flows.[1] This non-random selection of data can lead to selection bias and biased parameter estimates (Heckman, 1979). For example, Hillberry (2002) uses the same data set as MacCallum (1995) and shows that MacCallum's findings — in particular, that the volume of trade among Canadian provinces exceeds the provinces' trade with the US states by a factor of more than 20.9 — is reduced to a factor of 5.7 when selection bias is taken into account. It is important to precisely measure and estimate international trade flows in order to understand the structure and pattern of world trade. The purpose of this study is to test for selection bias in gravity modelling. As a case study, we use agrifood trade data disaggregated to the four digit SITC level from 1990 to 2000 for 52 countries of diverse development levels and 46 agrifood products without excluding zero trade flows in the analysis. Agrifood provides an interesting case study because Helpman et al. (2008)

[1] Zero trade-flows are generally omitted because it is common to use a log-linear functional form.

and Silva and Tenreyro (2006) find that selection bias in their sample is small and does not affect the estimates. It remains to be seen whether their findings hold for agrifood trade data? The Heckman Maximum Likelihood (ML) procedure is used to test for selection bias and the marginal effects (elasticities) of the explanatory variables are derived to help interpret the estimated parameters.

The analysis makes at least three important contributions to the existing literature. First, the study tests for selection bias in gravity modeling using agrifood trade data and concludes that ignoring zero trade flows can lead to incorrect statistical and economic inferences, a result that contrasts with Helpman et al. (2008) and Silva and Tenreyro (2006). Second, conditional and unconditional marginal effects are derived depending on the nature of the zero trade flows[2], making the interpretation of the estimated coefficients easy. Third, the analysis adds to the literature on the drivers of agrifood product trade.

The article is organized into six sections. Section two discusses the derivation of the generalized gravity equation. Section three discusses the problem of dealing with zero trade flows in gravity trade modeling. Section four presents a description of the data used in the analysis. The empirical results are explained in the fifth section, followed by the conclusions in the sixth and final section.

2.0 GENERALIZED GRAVITY EQUATION

Recent theoretical developments in the trade literature justify the existence of zero trade flows. Helpman et al. (2008) use the profitability of a firm to explain the existence of zero bilateral trade flows at the aggregate level, suggesting that firms have varying levels of productivity and

[2] That is whether zero trade-flows are actual (corner solution) or potential (missing).

only the more productive firms find it profitable to export. In addition, the profitability of exports varies by destination and profitability is higher when firms export to countries with larger markets and where fixed and variable export costs are lower. Hence, for every importer *i*, there is a marginal exporter in country *j* that just breaks even by exporting to country *i*. Firms in country *j* with higher productivity than the marginal exporter receive positive profits from exporting to country *i*. Helpman et al. (2008) also argue that the features of marginal exporters can be identified from the variation in the characteristics of the importers and that aggregate data can be used to predict the volume of exports of heterogeneous firms. Hence, Helpman et al. (2008) derive a gravity model that accounts for firm heterogeneity and fixed trade costs, and predict zero trade flows by allowing all firms in country *j* to choose not to export to country *i* if it is not possible for any firm in country *j* to make a profit shipping to country *i*. Helpman et al. (2008) also decomposes the impact of trade friction on trade flows into the intensive and extensive margins. The intensive margin is the trade volume per exporter while the extensive margin refers to the number of exporters. Their model results in a generalized gravity equation that accounts for the self-selection of firms into export markets and their impact on trade volumes. The model is estimated using a two-stage estimation procedure consisting of selection and outcome equations. The selection equation models the selection of trade partners and the outcome equation models the trade flow.

We assume that firm-level heterogeneity is not correlated with the export decision, implying that all firms are similarly affected by barriers to trade between trade partners. Hence, the selection bias arising due to the unobserved country-pair level random errors can be corrected using the Heckman selection model. The Heckman procedure also consists of two equations: a sample selection equation and an outcome equation. The sample selection equation follows a selection rule where trade between bilateral trade partners are observed when the trade flow is greater than zero, while the outcome equation investigates the relationship of interest when the outcome is observable, i.e., when trade is greater than zero.

The focus of this paper is largely empirical and we use the generalized gravity equation as derived by Bergstrand (1989, 1990)[3]. Bergstrand (1989, 1990) starts with consumers maximizing utility defined over differentiated and homogenous products subject to a budget constraint to derive demand and inverse demand functions. On the supply side, firms use linear technologies to produce products and allocate these to different markets to maximize profit in monopolistically competitive markets. Finally, the generalized gravity equation, in value terms, is obtained by multiplying the inverse demand functions by the profit-maximizing quantity of differentiated products. In our analysis the generalized gravity equation is augmented with Gini variables for the importing and exporting countries to proxy for income distribution in these countries as suggested by Linder (1969), Francois and Kaplan (1996) and Dalgin, Mitra and Trindade (2006) who emphasized the role of income inequality in trade.

3.0 HECKMAN SELECTION MODEL

The issue of ignoring zero trade flows in gravity models is well known but largely ignored. Wooldridge (2006) argues that if the data are randomly missing, then ignoring zeros reduces sample size but does not create any bias. Further, if the data is nonrandomly missing (or zero in the context of a log-linear gravity model) but the sample selection procedure uses an exogenous sampling rule then estimates could still be unbiased. However, if the sample selection is based on the value of the dependent variable (endogenous sample selection) then the parameters of the

[3] See Bergstrand (1989, 1990) for detailed derivation of generalized gravity equation.

estimated model will always be biased if estimated using ordinary least squares (OLS). Obviously, the typical gravity model uses endogenous sample selection since only trade flows greater than zero are considered.

Researchers have dealt with trade values of zero in three ways: 1) delete the observations with values equal to zero (MacCallum 1995; Frankel 1997); 2) replace the zeros with small positive numbers (Linnemann, 1966; Wang and Winters, 1991; MacCallum, 1995 and Raballand, 2003); or 3) estimate the regression equation as a Tobit model and censor the zero observations at the left tail (Rose 2000). Irrespective of the sample selection procedure only a few studies have properly treated zero trade flows in the context of the gravity model (Linders and De Groot, 2006 and Bikker and De Vos, 1992). Linders and De Groot (2006, p. 2) observe that "the sample selection model, which has been widely used in other fields of applied economics, is rather novel to the literature on bilateral trade. Because the sample selection model offers a theoretically sound and econometrically elegant solution to include zero flows in the gravity model of bilateral trade, it deserves more attention in applied work."

The Heckman sample selection model consists of a sample selection (equation 1) and an outcome equation (equation 2). Define a binary variable t^*_{ijfy} such that $t^*_{ijfy} = 1$ if $T_{ijfy} > 0$ and otherwise zero where T_{ijfy} is the value of country i's trade with country j of commodity sector f in year y, $\boldsymbol{z_i}$ and $\boldsymbol{X_i}$ are the matrices of independent variables in the selection and outcome equations and $\boldsymbol{\eta}'$ and $\boldsymbol{\gamma}'$ are the respective vectors of parameters. Error terms of the selection and outcome equations are represented as u_i and ε_i. Hence, the selection equation can be represented as:

$$t^*_{ijfy} = \boldsymbol{\eta}'\boldsymbol{z_i} + u_i \quad (1)$$

and the outcome equation as

$$T_{ijfy} = \boldsymbol{\gamma}'\boldsymbol{X_i} + \varepsilon_i \quad (2)$$

The errors u_i and ε_i, i=1,...,N have a bivariate normal distribution with zero means, standard derivation of σ_u and σ_ε and correlation ρ. Greene (2003) and Hoffmann and Kassouf (2005) show that

$$E\left[T_{ijfy}|t^*_{ijfy} = 1\right] = \boldsymbol{\gamma}'\boldsymbol{X_i} + \rho\sigma_\varepsilon\lambda_i(\alpha_u) \quad (3)$$

Where $\lambda_i(\alpha_u) = \frac{\phi\left(\frac{\eta' z}{\sigma_u}\right)}{\Phi\left(\frac{\eta' z}{\sigma_u}\right)}$ is the inverse Mills ratio (IMR), ϕ is the standard normal density function and Φ is the cumulative standard normal distribution. Equation (3) estimates the expected values of T_{ijfy} when trade is observed. Greene (2003) shows that due to the correlation between $\boldsymbol{X_i}$ and $\lambda_i(\alpha_u)$, a least squares regression of T_{ijfy} on $\boldsymbol{X_i}$, omitting $\lambda_i(\alpha_u)$ will produce an inconsistent estimator of $\boldsymbol{\gamma}'$. Also, the standard regression techniques *a priori* assume that ρ=0, thus eliminating the IMR in equation (3) and producing biased estimation results if the IMR is statistically significant. A least square regression will yield consistent estimators only if the expected value of the error is known and included in the regression — as the Heckman selection model does (Hoffmann and Kassouf, 2005).

Let X_f denote regressors common to both the selection and outcome equations, then the marginal effect for the regressor is

$$\frac{\partial E\left[T_{ijfy}|t^*_{ijfy} > 0\right]}{\partial X_f} = \boldsymbol{\gamma_f} - \frac{\eta_f}{\sigma_u}\beta_\lambda\delta_i \quad (4)$$

The marginal effect given in (4) is composed of a change in the value of trade (T_{ijfy}) due to a change in X_f for the bilateral trade partners participating in trade. Hence, this effect is conditional on the bilateral partners trading non-zero quantities of product *f* and it is called the conditional marginal effect. Greene (2003) and Hoffmann and Kassouf (2005) also derive the conditional marginal effect for a common binary variable. Let z_1 be a binary explanatory variable that is common to both the selection and outcome equations; then the change in the IMR ($\Delta\lambda$) when z_1 moves from one (z_{11}) to zero (z_{10}) is $\Delta\lambda = \frac{\phi\left(\frac{z_{11}\boldsymbol{\eta_i}}{\sigma_u}\right)}{\Phi\left(\frac{z_{11}\boldsymbol{\eta_i}}{\sigma_u}\right)} - \frac{\phi\left(\frac{z_{10}\boldsymbol{\eta_i}}{\sigma_u}\right)}{\Phi\left(\frac{z_{10}\boldsymbol{\eta_i}}{\sigma_u}\right)}$. Hence, the conditional marginal effect for the binary variable is

$$\frac{E\left[T_{ijfy}|t^*_{ijfy} > 0\right]}{\square z_f} = \boldsymbol{\gamma_f} - \beta_\lambda \Delta\lambda \tag{5}$$

Hoffmann and Kassouf (2005) also derive the unconditional marginal effects for the continuous and binary variables that are common to both the selection and outcome equations. For a logarithmic specification of gravity model, the unconditional marginal effect for a continuous variable that is common to both the selection and outcome equations is[4]

$$\frac{E\left[T_{ijfy}|t^*_{ijfy} > 0\right]}{\square X_f} = \boldsymbol{\gamma_f} - \frac{\eta_f}{\sigma_u}\beta_\lambda\delta_i + \left[\Phi\left(\frac{z_i\boldsymbol{\eta_i}}{\sigma_u}\right)\right]^{-1}\phi\left(\frac{z_i\boldsymbol{\eta_i}}{\sigma_u}\right)\frac{\boldsymbol{\eta_f}}{\sigma_u} \tag{6}$$

Using the analogy of Hoffmann and Kassouf (2005), the first two terms on the right hand side show the change in trade of agrifood product *f* for the trading partners having observable trade (i.e. more than zero) while the last term shows the effect due to a change in the probability of the trading partners being involved in trade. Similarly, the unconditional marginal effect for the binary variable that is common to both the selection and outcome equations is

$$\frac{E\left[T_{ijfy}|t^*_{ijfy} > 0\right]}{\square z_f} = \boldsymbol{\gamma_f} + \beta_\lambda \Delta\lambda + \Delta ln\Phi(-\alpha_u) \tag{7}$$

[4] See Hoffmann and Kassouf (2005) for details.

Where $\Delta ln\Phi(-\alpha_u) = ln\Phi\left(\frac{z_{11}\eta_i}{\sigma_u}\right) - ln\phi\left(\frac{z_{10}\eta_i}{\sigma_u}\right)$. Since marginal effects vary for each observation we calculate these effects at the mean values.

The existing studies that use Heckman selection model specify the selection and outcome equations as either a gravity equation or generalized gravity equation. Linder and de Groot (2006) use a generalized gravity equation for both the selection and outcome equations. Rose (2000) estimates a variant of the gravity model without explaining it in detail. Hillberry (2002) estimates a more restricted variant of the gravity model in which an independent selection equation is estimated. Helpman et al. (2008) estimates selection and outcome equations that include only the variables that affect trade costs. Hence, the exact specification of the selection and outcome equations differ across studies but a generalized gravity equation incorporating the variables determining trade costs are generally incorporated in the selection equation. This study also specifies the selection and outcome equations as a generalized gravity model that is similar to the one used by Rose (2000) and it includes most of the variables indentified by Helpman et al. (2008). Hence, the extended equations (1) and (2) are

$$t^*_{ijfy} = \begin{array}{l} \pi_i + \pi_i + \pi_y + \pi_f + \eta_1 lndist_{ij} + \eta_2 DCB_{ij} + \eta_3 DPTA_{ijy} + \\ \eta_4 lnGini_{iy} + \eta_5 lnGini_{jy} + \eta_6 lnGDP_{iy} + \eta_7 lnGDP_{jy} + \\ \eta_8 lnPCGDP_{iy} + \eta_9 lnPCGDP_{jy} + \eta_{10} DLandl_i + \eta_{11} DComcol_i + \\ \eta_{12} DColony_{ij} + \eta_{13} DComlang_{ij} + u_{ijfy} \end{array} \tag{1'}$$

$$\mathrm{T}_{ijfy} = \begin{array}{l} \psi_i + \psi_i + \psi_y + \psi_f + \gamma_1 lndist_{ij} + \gamma_2 DCB_{ij} + \gamma_3 DPTA_{ijy} + \\ \gamma_4 lnGini_{iy} + \gamma_5 lnGini_{jy} + \gamma_6 lnGDP_{iy} + \gamma_7 lnGDP_{jy} + \\ \gamma_8 lnPCGDP_{iy} + \gamma_9 lnPCGDP_{jy} + \gamma_{10} DLandl_i + \gamma_{11} DComcol_i + \\ \gamma_{12} DColony_{ij} + \gamma_{13} DComlang_{ij} + \gamma_{14} AHT(\rho) + \gamma_{15} ln(\sigma_u) + \epsilon_{ijfy} \end{array} \tag{2'}$$

The description of the variables included in equations (1′) and (2′) is given in table 1. Note that equation (2′) includes the arc hyperbolic tangent of rho ($AHT(\rho)$) and the logarithm of standard errors of the selection equation ($ln(\sigma_u)$) — the variables which determine the IMR.

Table 1: Description of Variables

Variable/Symbol	**Description**
T_{ijfy}	Real value of country i's trade with country j in product f in year y (US$)
t^*_{ijfy}	A binary variable such that $t^*_{ijfy} = 1$ if $T_{ijfy} > 0$ and zero otherwise
$dist_{ij}$	Distance between bilateral trade partners
CBD_{ij}	Binary variable, which is unity if bilateral trade partners have a common border and zero otherwise
$DPTA_{ijy}$	Binary variable, which is unity if bilateral trade partners have or belong to the same regional trade agreement and zero otherwise
$Gini_{iy}$	Gini coefficient for country i in year y
$Gini_{jy}$	Gini coefficient for country j in year y
GDP_{iy}	Real gross domestic product of country i in year y (US$)
GDP_{jy}	Real gross domestic product of country j in year y (US$)
$PCGDP_{iy}$	Real per capita gross domestic product of country i in year y (US$)
$PCGDP_{jy}$	Real per capita gross domestic product of country j in year y (US$)
$Landl_i$	Binary variable, which is unity if trading partners are land locked and zero otherwise
$DComcol_i$	Binary variable, which is unity if trading partners were ever colonized by the same colonizer, after 1945, and zero otherwise
$DColony_{ij}$	Binary variable, which is unity if trading partners ever colonized each other; that is, one of the partners is a colony and the other its colonizer
$DComlang_{ij}$	Binary variable, which is unity if bilateral trade partners have a common language and zero otherwise
$AHT(\rho)$	Arc hyperbolic tangent of rho
σ_u	Standard errors of the selection equation
π_i, π_j, π_f and π_y	Respectively importing, exporting, product, and year fixed effects
ψ_i, ψ_j, ψ_f and ψ_y	Respectively importing, exporting, product, and year fixed effects
Ln	Natural logarithm
u_{ijfy} and ϵ_{ijfy}	Error terms

Sufficient variation is required to identify the parameters of the selection and outcome equations, requiring identification of separate variables that affect the IMR from those that determine the outcome equation. However, in practice this is seldom possible. Maddala (1983, p.

233-34) in discussing the Heckman selection model suggests that the "condition for identification for the simultaneous-equations model are well known; namely, Cov (u_1, u_2) = 0 [i.e. u_i, ε_i in this study] or there is at least one variable in X_i not included in Z_i. These are the conditions for identification in Heckman's model". The condition is also known as the exclusion restriction. In our case, the selection and outcome equations contain a similar set of variables. Hence, the validity of the estimates depends on the normality of the residuals. Therefore, a restricted model imposing the exclusion restriction is estimated to check whether imposing the restriction affects the estimates. The issue is explored further in section 5.3. Finally, the Heckman model can be estimated either simultaneously or as two separate equations. However, Greene (2003) shows that the estimates generated with the Heckman model estimated simultaneously using the maximum likelihood (ML) procedure is homoskadestic, hence the selection and outcome equations are estimated using this procedure.

4.0 DATA

The trade data come from the World Trade Analyzer (WTA) of Statistics Canada, covering trade flows from 1990 to 2000 for most countries of the world. The data is organized by the Standard International Trade Classification (SITC), revision 3, at the four-digit level. Statistics Canada uses United Nations bilateral trade data to develop the WTA and also incorporates zeros into the data. We categorized SITC codes into ten differentiated agrifood product sectors: meat; dairy products; fresh fish; frozen fish; cereals; fresh fruit; processed fruit; vegetables; tea, coffee and mate; and alcoholic beverages.[5] However, individual products included in each sector are not

[5] Meat includes products having SITC codes 0111, 0112, 0113, 0121 and 0149; dairy 0223, 0230, 0240 and 0980; fish fresh 0341 and 0350; frozen fish 0342, 0343, 0360, 0371 and 0372; cereals 0481, 0483, 0484 and 0488; fresh fruits 0571, 0572, 0574, 0575 and 0579; processed fruits 0577, 0583, 0585, 0586, 0589 and 1110; vegetables 0541, 0542, 0544, 0545, 0546, 0561 and 0565; tea and coffee 0711, 0712, 0730, 0741 and 0742 and alcoholic beverages - 1121, 1123 and 1124.

aggregated. For example meat sector includes trade flows of products 0111, 0112, 0113, 0121 and 0149 for 52 countries from 1990 to 2000.

Countries are categorized as lower income (LI), lower middle income (LMI), upper middle income (UMI) and high income (HI) using the World Bank per capita gross national product thresholds.[6] Gross domestic product (GDP) and per capita GDP data come from the World Bank's World Development Indicators. Also, since GDP data for all of the selected countries are in US dollars, the US GDP deflator is used to get GDP estimates in real terms. Income inequality data come from the UN-WIDER data set. Estimates of the distance between capitals and border sharing are obtained from the World Bank's website (World Bank 2005). The dummy variable representing multilateral trade agreements is developed from the Tuck Trade Agreement database (CIB 2007). The data required for the other gravity variables in the trade model are compiled from Glick and Rose (2002).

5.0 RESULTS AND DISCUSSION

The empirical analysis involves ten agrifood product sectors and two estimation procedures. Each product sector consists of individual products with 46 products in total. For comparison across estimators, the model is estimated using the Heckman maximum likelihood (ML) and ordinary least squares (OLS) procedures. The results of the Heckman ML procedure for the outcome equation are reported in table 2 while the results of the selection equation are not reported (these are available upon request)

[6] Lower income countries include (Bangladesh, Ethiopia, India, Madagascar, Pakistan and Tanzania), Lower Middle Income countries include (Bolivia, Brazil, China, Colombia, Dominican Republic, Ecuador, Egypt, El Salvador, Guatemala, Indonesia, Jamaica, Jordan, Paraguay, Peru, Philippines, Romania, Sri Lanka and Thailand), Upper Middle income countries include (Argentina, Chile, Costa Rica, Hungary, Malaysia, Mexico, Panama, Poland, South Africa, Turkey, Uruguay and Venezuela) and High income countries include (Canada, Denmark, Finland, Germany, France, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and the United States).

The outcome equation used 43 percent of the total observations (the non-zero observations), with the highest proportion of censored observations in the fresh fruit group (Table 2). The Wald-test shows that all of the models are statistically significant at the 99 percent level of significance.[7] A likelihood ratio test is used to test for the independence of the selection and outcome equations. Specifically, it tests the null hypothesis that rho (ρ)[8] is equal to zero. Failure to reject this null hypothesis indicates insignificant sample selection bias, while rejection of this null means OLS will produce biased estimated. Indeed, the null that $\rho = 0$ is rejected for all of product groups and we conclude use of the Heckman procedure is appropriate. It is important to recognize, however, the ML estimation of the Heckman sample selection model does not directly report ρ and σ, but the arc hyperbolic tangent of rho as $ln\left(\frac{1+\rho}{1-\rho}\right)$ and it calculates the natural logarithm of σ. Table 2 shows that for all of the models both of the estimates of selection hazard (ρ and σ) are statistically significant indicating that ignoring zeros will produce biased estimates. Finally, the ML procedure consistently estimates ρ within the range of -1 and 1, suggesting that both the selection and outcome equations have no specification error. Collectively these results imply that estimation without considering zero trade flows would produce biased estimates. Hence, in this case, OLS estimates (based on ignoring zeros) would be biased. Helpman et al. (2008) and Silva and Tenreyro (2006) find that omission of zero trade-flows does not yield any selection bias as their estimates using both zero and positive trade flows are very similar. However, these authors used different techniques. Helpman et al. (2008) used a variant of the Heckman selection model while Silva and Tenreyro (2006) used Poisson estimation.

[7] The Wald-test tests the hypothesis that all of the coefficients in the outcome equation (i.e., regression model), except the constant, are zero.
[8] Rho indicates the correlation between the error terms of the outcome and selection equations; that is, the null hypothesis of cor(u_i,ξ_i) = 0 as against the alternative hypothesis of cor(ui,ξ_i) ≠ 0.

Importing and exporting country, product, and year fixed effects are kept in the models to account for unobserved factors including prices, commodity specific characteristics, domestic and trade related policies, industry specific border related hindrances, unmeasurable product quality characteristics, and technical and non-technical barriers to trade. [9] Fixed effects provide a solution to the unobserved heterogeneity (Mátyás, 1997 and Egger, 2002). These fixed effects are tested with the null hypothesis that their combined effects are zero. The results indicate that the fixed effects are statistically significant for all of the products, except for the year specific fixed effect for fresh fruit. The implication is that ignoring these effects in the empirical analysis will produce biased estimates.

5.1 Marginal Effects

Before discussing the marginal effects, it is important to mention that the estimated parameters of the generalized gravity equation using the Heckman selection model (table 2) generally have the expected sign. The distance variable is theoretically expected to have a negative sign because an increase in the distance between trading partners is expected to decrease trade. Similarly, countries with a common border, preferential trade agreements (PTA), common colonizer and trading partners ever colonized each other (colony) are expected to have positive signs. However, these variables may take signs other than what is expected in empirical analyses (Hallak, 2006). In our case, all of the estimates for distance are negative and for a common border and a PTA positive and statistically significant. Similarly estimates for the common colonizer and trading partners ever in a colonial relationship are positive and mostly statistically significant while the parameters for a common language are positive when statistically

[9] Each bilateral trade flow for each individual product is included as an observation during estimation, so product specific fixed effects are added to the model to account for product specific omitted heterogeneity.

significant. These results imply that the model produced the expected results — a sign of suitable specification.

One of the problems with the Heckman selection model is that the estimated parameters of the variables that are common to both the selection and outcome equations cannot be interpreted as the usual elasticities in the case of a log-linear gravity model. This is due to the inclusion of the IMR in the outcome equation. However, marginal effects (or elasticities) can be derived using the estimates of the Heckman selection model. But the derived elasticities can be either conditional or unconditional depending on the assumption made about the nature of zero trade flows; that is whether the zero's represent an actual trade flow of zero or if they are a potential zero trade flow as a result of missing or misreported data . Since we don't know if the zero trade flows represent cases with no trade flow or instances where trade is not reported we derive both the conditional and unconditional elasticities.[10] Marginal effects are compiled in tables 3 and 4. Hence the conditional marginal effect of the importing country income for alcoholic beverages suggests that keeping other variables constant, as income of the importing country increases by 10 percent, trade of alcoholic beverages increases by 8.3 percent, given that there is an actual zero trade flow. The unconditional marginal effect for the same variable implies that for countries participating in agrifood trade, as the income of the importing country increases by 10 percent, the trade of alcoholic beverages increases by 8.8 percent, keeping all other variables constant. Both the interpretations are different than the estimates using OLS (table 5) where the data on zero trade flows is simply discarded. The next section contains a comparison of the estimated elasticities across the estimation methods.

[10] The derivation of the marginal effects after estimation of the Heckman selection model also depends on the specification of the estimated model (see Hofmann and Kassouf, 2005 for details).

5.2 Does It Really Matter?

It is useful to compare the elasticities estimated using OLS which ignores the zero trade flows (Table 5) with the conditional elasticity estimates using Heckman estimation (table 3) where the zero trade flow is explicitly treated as a no trade situation. Although the OLS estimates are biased, the bias might be small enough to make little or no difference when making economic inferences. If so, OLS provides a straightforward means of obtaining statistically biased but potentially economically meaningful results and makes parameter estimates obtained in studies that ignored zero trade flows more meaningful. We approach the comparison from both a statistical and an economics point of view.

In our model there are 117 parameters related to variables based on the gravity model. Our comparison focuses on these parameters and ignores the parameters capturing the fixed effects. Using the Heckman approach, 62 percent of the parameters are statistically significant at the 90 percent level or above, 56 percent are statistically significant when estimated with OLS. The major difference between the two estimation techniques relates to the 63 parameters associated with continuous variables where 43 percent are statistically significant using Heckman and only 33 percent using OLS. The most important contributor to the difference is the effect of the importing countries income on trade where using Heckman six of the eight parameters are significant while with OLS only two are significant. The four elasticities found significant using Heckman and insignificant using OLS are all large enough to be economically important ranging from 0.77 for tea, coffee and mate to 1.52 for fresh fish (table 3).

Turning to an economic comparison of the OLS and Heckman results we focus on the 64 parameters that are statistically significant using both estimation techniques. Of these, there is only one case where a parameter switches sign between the two estimation techniques and this is for the effect of being a land locked country, trading dairy products, where the elasticity using OLS is 0.65 and using Heckman is -0.47 which is the expected sign (tables 3 and 5). So, from a direction of effect perspective the two methods are almost identical. Still economists are interested not only in the direction of the effect but the size of the effect – this is where we would expect bias to come into play. In this comparison we will concentrate on if the effect becomes more or less elastic without worrying about the sign of the coefficient. For 83 percent of the elasticities the OLS estimate is more elastic than the Heckman estimate; 11 percent of the elasticities differ by more than 50 percent and 28 percent of the elasticities differ by more than 20 percent. The largest difference in parameter estimates is for the effect of preferential trade agreements where all nine parameter estimates are more than 20 percent more elastic using OLS than using Heckman. For example, using the OLS estimate for the effect of a preferential trade agreement, on the trade of fresh fish, the elasticity is 1.04 while it is only 0.51 using Heckman (tables 3 and 5). Estimates of the effect of distance and a common border also tend to be quite different across the two estimation techniques. As a final check, from an economics perspective, we looked at the number of times an elasticity estimate changed from elastic using OLS to inelastic using Heckman. This happened seven times. In some cases the absolute difference between the two coefficients was small (for distance in cereal trade: -0.99 (OLS) and -1.12 (Heckman)) but in other cases it was quite large (for a PTA in fresh fish trade: 1.04 (OLS) and 0.51 (Heckman)).

It is also meaningful to compare the elasticities from the conditional (table 3) and the unconditional (table 4) elasticities calculated using Heckman to see how much difference the assumption about the nature of the zero trade flows makes. There are 66 common statistically significant elasticities using the conditional and unconditional Heckman estimates. There are no sign reversals and all but two unconditional estimates are more elastic than their conditional counterparts. Twenty-one percent of the unconditional elasticities are more than 50 percent more elastic than the conditional estimate and 80 percent are more than 20 percent more elastic. There are eleven cases where a conditional estimate is inelastic and the unconditional estimate is elastic and in some cases the difference is quite large (for importers income in tea, coffee and mate trade: 0.77 (conditional) and 2.28(unconditional); and for importers income in processed fruit trade: 0.73 (conditional) and 2.08 (unconditional)).

To conclude, ignoring zero-trade flows and using OLS results in biased parameter estimates and in many situations this bias can be large and economically meaningful. In addition, the assumption about the nature of the zero trade flows is also important – this is unfortunate because the researcher will almost never know the true nature of the recorded data. However, based on our case study the conditional elasticities will almost always be more inelastic than there unconditional counterparts.

5.3 THE RESTRICTED MODELS

The Heckman selection model estimated using the ML procedure showed that ignoring zeros in the analyses would produce biased estimates. However, these estimates are based on the assumption of normality of errors. Hence, it is relevant and important to verify the robustness of the results given in table 2. Restricted models, implementing an exclusion restriction, are estimated assuming that the parameter of the Gini coefficient of either the importing or the exporting country is zero (i.e. $\gamma_4 = 0$ or $\gamma_5 = 0$). The Gini coefficient is selected for implementing the exclusion restriction because in most of the estimated models (table 2), the variable is statistically insignificant. However, this exclusion restriction may result in specification error. Hence, before implementing the exclusion restriction, the Hausman and Likelihood Ratio (LR) specification tests are calculated to test for any resulting misspecification. The Hausman's test is based on estimating the variance of the difference of the estimators while the LR tests the null hypothesis that the parameter vector of a statistical model satisfies some constraint. The results of the tests are compiled in table 6 while the results of the restricted models are given in table 7.

The Hausman specification test is implemented in three ways: i) the unrestricted outcome equation of the always-consistent estimator is tested against the restricted estimator; ii) the unrestricted selection equation is tested against the restricted selection equation; and iii) the unrestricted selection and outcome equations are tested against the restricted equations. The Hausman specification test rejects the specification error for all three cases, while the Likelihood Ratio specification test indicates specification error only for processed fruits and vegetables. Hence, for these products, the parameter on the Gini coefficient of the exporting country is restricted to zero and the specification tests are calculated and they indicate no specification error. Now that the restricted models are not having any specification error, the results are compiled in table 7. Comparing these results with those of the unrestricted model (i.e. Table 2) shows that the magnitude and statistical significance of the estimated parameters of the restricted and unrestricted models are similar. Hence, applying the exclusion restriction does not change the estimates of Heckman selection model.

6.0 CONCLUSION

Corner solutions, or zero trade flows, are commonly observed when using disaggregated trade data. This is especially true for agrifood trade because in many cases the international market is quite thin. Trade studies based on log-linear gravity models, generally omit zero trade-flows to facilitate estimation. This study investigates whether the omission of zero trade flows in estimating the generalized gravity model leads to selection bias and incorrect statistical and economic inferences.

The study employs the Heckman selection model to test and correct for selection bias. The analysis provides strong evidence of selection bias in the generalized gravity model using disaggregated agrifood trade data as a case study. The implication is that parameters estimated using ordinary least squares, ignoring zero trade flows, are biased. These results are important because the gravity equation is the "work horse" of applied trade analysis and its correct specification and estimation is vital in deriving the correct policy implications from quantitative analysis. We derived the conditional and unconditional marginal effects for the variables common to both the selection and outcome equations using the Heckman specification; the only study we are aware of deriving these effects in the context of the gravity equation. The derived marginal effects are compared with the parameters estimated using OLS.

We find that ignoring selection bias rarely affects the signs of variables but often influences the magnitude, statistical significance and economic interpretation of the marginal effects. Some economic effects that are elastic when estimated using OLS become inelastic when selection bias is accounted for using the Heckman selection model. The analysis also shows that ignoring selection bias could also lead to wrong policy implications. For example, the effects of preferential trade agreements on trade are larger based on OLS estimates in comparison to Heckman estimates. Hence, treating zero trade flows properly is important from both an economic and a statistical perspective and should be included in quantitative analysis, especially when gravity trade models are employed.

Table 2: Generalized gravity equation estimated for beverages and agrifood products using Heckman selection model

Variable	**Alcoholic beverages**	**Cereals**	**Dairy**	**Fresh fish**	**Frozen fish**	**Fresh fruit**	**Processed fruit**	**Tea, coffee and mate**	**Vegetables**
Distance	-1.028***	-1.335***	-1.228***	-1.574***	-1.082***	-1.143***	-0.953***	-0.665***	-1.181***
	(0.023)	(0.032)	(0.020)	(0.030)	(0.027)	(0.029)	(0.017)	(0.020)	(0.020)
Common Border	0.899***	0.840***	0.872***	0.776***	0.750***	0.888***	0.743***	1.005***	0.816***
	(0.070)	(0.072)	(0.061)	(0.075)	(0.067)	-0.064	(0.049)	(0.064)	(0.048)
PTA	0.411***	0.939***	1.159***	0.972***	1.151***	0.403***	0.799***	0.956***	0.769***
	(0.080)	(0.094)	(0.064)	(0.085)	(0.079)	(0.066)	(0.051)	(0.067)	(0.056)
Importer Gini	0.028	-0.033	0.102	0.151	0.124	-0.053	-0.039	0.095	-0.041
	(0.113)	(0.135)	(0.095)	(0.137)	(0.128)	(0.127)	(0.086)	(0.111)	(0.086)
Exporter Gini	0.111	0.061	-0.064	-0.111	0.187	0.089	0.144	0.04	0.271**
	(0.114)	(0.143)	(0.099)	(0.132)	(0.128)	(0.123)	(0.088)	(0.107)	(0.089)
Importer Income	0.935*	-2.116***	0.652	3.331***	-1.257*	3.223***	2.690***	2.689***	2.009***
	(0.513)	(0.586)	(0.428)	(0.750)	(0.628)	(0.666)	(0.398)	(0.492)	(0.414)
Exporter Income	5.527***	2.387**	1.945***	1.453*	-0.168	-0.526	0.353	-1.524**	-0.644
	(0.681)	(0.810)	(0.567)	(0.700)	(0.749)	(0.591)	(0.439)	(0.543)	(0.434)
Importer Per Capita Income	-0.14	3.150***	0.08	-1.938*	2.364***	-2.189**	-1.471***	-2.123***	-0.847
	(0.544)	(0.619)	(0.449)	(0.780)	(0.661)	(0.699)	(0.419)	(0.513)	(0.434)
Exporter Per Capita Income	-5.521***	-2.098*	-1.963**	-1.745*	0.802	0.68	-0.454	1.581**	0.617
	(0.717)	(0.860)	(0.598)	(0.736)	(0.795)	(0.622)	(0.465)	(0.576)	(0.457)
Land Locked	0.505	-0.233	0.523**	-1.681*	-2.185**	-0.903**	0.506*	-0.218	-0.158
	(0.287)	(0.309)	(0.168)	(0.709)	(0.782)	(0.300)	(0.242)	(0.187)	(0.193)
Common Colonizer	1.135***	0.199	0.073	1.176***	2.280***	1.780***	1.752***	0.969***	1.451***
	(0.224)	(0.184)	(0.162)	(0.211)	(0.447)	(0.163)	(0.142)	(0.172)	(0.144)
Colony	0.896***	0.408***	0.457***	1.307***	1.084***	1.022***	0.403***	0.760***	0.868***
	(0.074)	(0.082)	(0.063)	(0.083)	(0.075)	(0.073)	(0.052)	(0.072)	(0.053)
Common Language	0.289***	0.504***	0.540***	0.004	0.349***	0.002	0.359***	-0.056	-0.063
	(0.050)	(0.055)	(0.043)	(0.058)	(0.055)	(0.052)	(0.036)	(0.046)	(0.036)
Selection Bias									
Arc Hyperbolic Tangent of	1.481***	0.567***	0.630***	1.078***	0.960***	0.619***	1.092***	0.789***	0.812***
Rho	(0.055)	(0.054)	(0.025)	(0.052)	(0.050)	(0.032)	(0.035)	(0.032)	(0.042)
ln(sigma)	0.790***	0.545***	0.599***	0.768***	0.720***	0.726***	0.745***	0.765***	0.723***
	(0.010)	(0.013)	(0.007)	(0.012)	(0.013)	(0.009)	(0.009)	(0.010)	(0.010)
Fixed Effects									
Importing Countries	2,155.4***	129.1***	214.7***	3364.4***	1478.8***	1877.1***	160.1***	130.1***	161.6***
Exporting Countries	7,539.7***	267.4***	500.5***	3963.1***	4654.5***	9673.1***	668.3***	651.0***	775.4***
Year	38.2***	67.9***	42.2***	18.8**	25.5***	14.0	52.4***	72.8***	31.1**
Commodity	1,136.8***	322.6***	5088.1***	86.29***	1233.8***	2987.2***	2720.6***	4342.8***	2225.2***
Summary Statistcis									
Total Number of Observations	35,145	20,273	38,533	28,776	28,875	36,707	66,077	44,088	67,386
Censored Observations	12,673	8,707	15,248	12,753	12,535	16,184	28,293	18,782	28,873
LR Test	1,178.7***	109.8***	467.2***	516.5***	404.3***	257.3***	1089.5***	410.8***	626.4***
Wald Chi	22021.2***	6283.1***	26456.0***	14843.8***	12156.2***	15408.4***	24446.6***	17693.8***	17280.9***

[A] *, **, *** respectively show significance at 90, 95 and 95 percent levels.

[B] Robust standard errors.

Table 3: Conditional marginal effects of the generalized gravity equation using Heckman selection model

Variable	Alcoholic beverages	Cereals	Dairy	Fresh fish	Frozen fish	Fresh fruit	Processed fruit	Tea, coffee and mate	Vegetables
Distance	-0.593***[A]	-0.987***	-0.941***	-0.910***	-0.627***	-0.788***	-0.546***	-0.413***	-0.741***
	(0.019)[B]	(0.032)	(0.020)	(0.028)	(0.025)	(0.028)	(0.016)	(0.019)	(0.020)
Common Border	0.525***	0.671***	0.630***	0.617***	0.657***	0.621***	0.418***	0.800***	0.552***
	(0.069)	(0.072)	(0.061)	(0.075)	(0.067)	(0.064)	(0.049)	(0.064)	(0.048)
PTA	0.204**	0.660***	0.891***	0.510***	0.745***	0.317***	0.461***	0.682***	0.505***
	(0.080)	(0.094)	(0.064)	(0.085)	(0.079)	(0.066)	(0.051)	(0.067)	(0.056)
Importer Gini	-0.067	-0.015	0.100	-0.019	0.006	-0.067	-0.038	0.015	-0.100
	(0.092)	(0.132)	(0.092)	(0.124)	(0.116)	(0.122)	(0.076)	(0.103)	(0.080)
Exporter Gini	0.080	0.046	-0.010	-0.121	0.122	0.095	0.066	-0.009	0.129
	(0.093)	(0.138)	(0.095)	(0.119)	(0.119)	(0.118)	(0.078)	(0.100)	(0.082)
Importer Income	0.827**	-2.098***	0.093	1.522**	-1.746**	0.885	0.935**	0.774*	0.520
	(0.408)	(0.558)	(0.408)	(0.643)	(0.565)	(0.616)	(0.338)	(0.445)	(0.381)
Exporter Income	2.300***	2.048**	0.624	-0.391	-0.170	-0.939*	0.037	-1.325**	-0.843**
	(0.550)	(0.776)	(0.526)	(0.632)	(0.682)	(0.564)	(0.386)	(0.505)	(0.400)
Importer Per Capita Income	-0.252	2.907***	0.391	-0.856	2.332***	-0.211	-0.236	-0.445	0.185
	(0.432)	(0.590)	(0.427)	(0.666)	(0.590)	(0.643)	(0.354)	(0.463)	(0.397)
Exporter Per Capita Income	-2.214***	-1.717**	-0.537	0.380	0.638	1.210**	0.027	1.455**	0.929**
	(0.579)	(0.823)	(0.554)	(0.662)	(0.721)	(0.593)	(0.410)	(0.534)	(0.421)
Land Locked	-0.074	-0.428	-0.474**	-1.083	-1.702**	-1.271***	0.134	-0.429**	-0.402**
	(0.286)	(0.309)	(0.168)	(0.709)	(0.782)	(0.300)	(0.242)	(0.187)	(0.193)
Common Colonizer	0.894***	0.073	0.025	0.696***	2.003***	1.436***	1.238***	0.855***	1.136***
	(0.223)	(0.184)	(0.162)	(0.211)	(0.447)	(0.163)	(0.142)	(0.172)	(0.144)
Colony	0.523***	0.234**	0.328***	0.809***	0.794***	0.796***	0.197***	0.554***	0.574***
	(0.074)	(0.082)	(0.063)	(0.083)	(0.075)	(0.073)	(0.052)	(0.072)	(0.053)
Common Language	0.135**	0.336***	0.350***	-0.131**	0.075	-0.029	0.062*	-0.159**	-0.191***
	(0.049)	(0.055)	(0.043)	(0.709)	(0.055)	(0.052)	(0.036)	(0.046)	(0.036)

[A] *, **, *** respectively show significance at 90, 95 and 95 percent levels.
[B] Robust standard errors.

Table 4: Unconditional marginal effects of the generalized gravity equation estimated using Heckman selection model

Variable	Alcoholic beverages	Cereals	Dairy	Fresh fish	Frozen fish	Fresh fruit	Processed fruit	Tea, coffee and mate	Vegetables
Distance	-0.802***[A]	-1.410***	-1.226***	-1.339***	-0.955***	-1.140***	-0.812***	-0.611***	-1.088***
	(0.020)[B]	(0.033)	(0.020)	(0.028)	(0.026)	(0.029)	(0.016)	(0.020)	(0.020)
Common Border	0.690***	0.862***	0.848***	0.716***	0.722***	0.865***	0.615***	0.953***	0.746***
	(0.069)	(0.072)	(0.061)	(0.075)	(0.067)	(0.064)	(0.049)	(0.064)	(0.048)
PTA	0.299***	0.955***	1.127***	0.776***	1.007***	0.400***	0.664***	0.882***	0.698***
	(0.080)	(0.094)	(0.064)	(0.085)	(0.079)	(0.066)	(0.051)	(0.067)	(0.056)
Importer Gini	-0.021	-0.037	0.102	0.091	0.091	-0.053	-0.039	0.078	-0.053
	(0.098)	(0.138)	(0.095)	(0.128)	(0.122)	(0.127)	(0.080)	(0.108)	(0.083)
Exporter Gini	0.095	0.064	-0.063	-0.114	0.169	0.089	0.117	0.030	0.241**
	(0.099)	(0.145)	(0.099)	(0.123)	(0.123)	(0.123)	(0.082)	(0.104)	(0.086)
Importer Income	0.879**	-2.120***	0.649	2.691***	-1.393**	3.201***	2.079***	2.283***	1.696***
	(0.442)	(0.599)	(0.428)	(0.694)	(0.598)	(0.665)	(0.367)	(0.475)	(0.402)
Exporter Income	3.853***	2.460**	1.939***	0.800	-0.168	-0.530	0.243	-1.482**	-0.686
	(0.595)	(0.825)	(0.567)	(0.658)	(0.717)	(0.590)	(0.410)	(0.528)	(0.421)
Importer Per Capita Income	-0.198	3.202***	0.081	-1.555**	2.355***	-2.170**	-1.041**	-1.768***	-0.630
	(0.469)	(0.633)	(0.449)	(0.720)	(0.628)	(0.698)	(0.385)	(0.496)	(0.421)
Exporter Per Capita Income	-3.805***	-2.180**	-1.958***	-0.993	0.756	0.685	-0.286	1.554**	0.683
	(0.627)	(0.876)	(0.598)	(0.690)	(0.760)	(0.621)	(0.435)	(0.560)	(0.444)
Land Locked	0.169	-0.213	-0.521**	-1.506**	-2.083**	-0.950**	0.357	-0.272	-0.224
	(0.286)	0.309	(0.168)	(0.709)	(0.782)	(0.300)	(0.242)	(0.187)	(0.193)
Common Colonizer	1.003***	0.217	0.071	0.970***	2.188***	1.736***	1.531***	0.942***	1.361***
	(0.223)	(0.184)	(0.162)	(0.211)	(0.447)	(0.163)	(0.142)	(0.172)	(0.144)
Colony	0.687***	0.430***	0.450***	1.097***	0.988***	1.004***	0.325***	0.708***	0.787***
	(0.074)	(0.082)	(0.063)	(0.083)	(0.075)	(0.073)	(0.052)	(0.072)	(0.053)
Common Language	0.207***	0.532***	0.530***	-0.045	0.264***	0.001	0.248***	-0.080*	-0.093**
	(0.049)	(0.055)	(0.043)	(0.058)	(0.055)	(0.052)	(0.036)	(0.046)	(0.036)

[A] *, **, *** respectively show significance at 90, 95 and 95 percent levels.
[B] Robust standard errors.

Table 5: Generalized gravity equation estimated for beverages and agrifood products using Least Squares

Variable	Alcoholic beverages	Cereals	Dairy	Fresh fish	Frozen fish	Fresh fruit	Processed fruit	Tea, coffee and mate	Vegetables
Distance	-0.721***	-1.119***	-1.061***	-1.156***	-0.768***	-0.888***	-0.655***	-0.489***	-0.882***
	(0.020)	(0.026)	(0.020)	(0.025)	(0.023)	(0.026)	(0.014)	(0.019)	(0.015)
Common Border	0.689***	0.776***	0.741***	0.762***	0.776***	0.716***	0.623***	0.848***	0.694***
	(0.061)	(0.071)	(0.060)	(0.070)	(0.062)	(0.063)	(0.045)	(0.060)	(0.045)
PTA	0.513***	0.970***	1.175***	1.037***	1.109***	0.455***	0.828***	0.870***	0.886***
	(0.073)	(0.095)	(0.064)	(0.082)	(0.076)	(0.066)	(0.047)	(0.065)	(0.054)
Importer Gini	-0.094	-0.028	0.091	-0.042	0.039	-0.058	-0.043	0.021	-0.103
	(0.100)	(0.133)	(0.092)	(0.127)	(0.119)	(0.123)	(0.078)	(0.104)	(0.081)
Exporter Gini	0.071	0.020	-0.037	-0.075	0.132	0.094	0.065	-0.009	0.159
	(0.101)	(0.139)	(0.096)	(0.124)	(0.122)	(0.118)	(0.081)	(0.101)	(0.083)
Importer Income	0.333	-2.432***	0.015	0.875	-2.132***	1.016	0.393	0.546	0.254
	(0.452)	(0.568)	(0.412)	(0.682)	(0.589)	(0.623)	(0.353)	(0.444)	(0.379)
Exporter Income	1.687**	1.776*	0.398	-0.339	-0.454	-0.995	-0.313	-1.665**	-0.908*
	(0.606)	(0.796)	(0.542)	(0.644)	(0.700)	(0.571)	(0.402)	(0.513)	(0.408)
Importer Per Capita	0.283	3.257***	0.537	-0.074	2.760***	-0.298	0.366	-0.207	0.506
Income	(0.479)	(0.600)	(0.432)	(0.711)	(0.620)	(0.654)	(0.372)	(0.464)	(-0.400)
Exporter Per Capita	-1.586*	-1.436	-0.318	0.220	0.931	1.246*	0.347	1.765**	0.977*
Income	(0.638)	(0.845)	(0.572)	(0.676)	(0.742)	(0.600)	(0.426)	(0.543)	(0.429)
Land Locked	-0.043	-0.253	0.651***	-0.910*	-2.214***	-1.103***	0.334	-0.241	-0.162
	(0.231)	(0.294)	(0.156)	(0.362)	(0.263)	(0.280)	(0.221)	(0.166)	(0.178)
Common Colonizer	1.592***	0.129	0.077	0.601**	2.194***	1.418***	1.377***	0.860***	1.244***
	(0.182)	(0.180)	(0.154)	(0.189)	(0.427)	(0.156)	(0.128)	(0.158)	(0.138)
Colony	0.476***	0.251**	0.369***	0.910***	0.837***	0.814***	0.196***	0.563***	0.600***
	(0.068)	(0.079)	(0.061)	(0.077)	(0.071)	(0.071)	(0.047)	(0.066)	(0.050)
Common Language	0.166***	0.407***	0.402***	-0.097	0.113*	-0.009	0.120***	-0.138**	-0.187***
	(0.044)	(0.053)	(0.041)	(0.054)	(0.050)	(0.050)	(0.032)	(0.043)	(0.033)
Fixed Effects									
Importing Countries	33.2***	21.6***	35.9***	56.4***	35.4***	30.1***	25.1***	20.3***	28***
Exporting Countries	149.8***	69.2***	79.5***	64.5***	88.3***	179.8***	106.3***	112.3***	180.3***
Year	1.8*	4.6***	3.8***	2.2**	1.65*	1.26	2.25**	5.6***	1.6**
Commodity	353.4***	623.5***	1572.1***	1415.8***	455.1***	649.2***	446.6***	828.4***	318.1***
Summary Statistcis									
R-squared	0.49	0.48	0.53	0.52	0.44	0.44	0.42	0.41	0.41
Number of Observation	22,472	11,566	23,285	16,023	16,340	20,523	37,784	25,306	38,513
F-Statistics	172.0***	87.6***	247.0***	214.0***	104.2***	157.0***	242.0***	149.0***	225.0***
RMSE	1.8	1.6	1.7	1.9	1.8	1.9	1.8	1.9	1.8

[A] *, **, *** respectively show significance at 90, 95 and 95 percent levels.

[B] Robust standard errors

Table 6: Specification tests of the restricted model against the unrestricted model

Commodity	Hausman Specification Test			LR Specification Test
	Outcome Equation	Selection Equation	All Equations	
Alcohol	0.23	0.27	0.27	0.27
Cereals	0.19	0.11	0.19	0.19
Dairy	2.71	2.38	2.37	2.38
Frozen Fish	0.86	1.11	1.12	1.11
Fresh Fish	0.74	1.05	1.05	1.04
Fresh Fruits	0.02	0.02	0.02	0.02
Processed Fruits	1.98	3.44	3.45	3.45*
Processed Fruits[A]	0.00	0.00	0.00	0.00
Tea	1.05	1.29	1.29	1.29
Vegetables	2.46	15.55	15.41	15.44***
Vegetables [A]	0.55	2.73	2.72	2.71

[A] Specification tests when Gini of importing country dropped.

Table 7: Restricted generalized gravity equation estimated for beverages and agrifood products using Heckman selection model

Variable	**Alcoholic beverages**	**Cereals**	**Dairy**	**Frozen fish**	**Fresh fish**	**Fresh fruit**	**Processed fruit**[A]	**Tea, coffee and mate**	**Vegetables**[A]
Log of Distance	-1.028***	-1.335***	-1.227***	-1.081***	-1.574***	-1.143***	-0.953***	-0.665***	-1.181***
	(0.022)[B]	(0.030)	(0.020)	(0.026)	(0.028)	(0.028)	(0.017)	(0.021)	(0.019)
Common Border	0.899***	0.840***	0.872***	0.750***	0.776***	0.888***	0.743***	1.005***	0.816***
	(0.070)	(0.073)	(0.057)	(0.074)	(0.079)	(0.066)	(0.050)	(0.063)	(0.048)
PTA	0.411***	0.939***	1.159***	1.151***	0.972***	0.404***	0.799***	0.956***	0.769***
	(0.077)	(0.096)	(0.067)	(0.098)	(0.106)	(0.073)	(0.056)	(0.069)	(0.062)
Log of Gini of Importing Country	0.026	-0.034	0.103	0.118	0.153	-0.054	0.145[A]	0.094	0.271**[A]
	(0.112)	(0.133)	(0.097)	(0.128)	(0.140)	(0.126)	(0.085)	(0.113)	(0.089)
Log of GDP of Importing Country	0.93	-2.117***	0.656	-1.244	3.326***	3.224***	2.691***	2.685***	2.009***
	(0.516)	(0.584)	(0.426)	(0.641)	(0.759)	(0.642)	(0.403)	(0.508)	(0.405)
Log of GDP of Exporting Country	5.540***	2.402**	1.934***	-0.157	1.445*	-0.518	0.353	-1.519**	-0.644
	(0.64)	(0.822)	(0.568)	(0.723)	(0.682)	(0.604)	(0.433)	(0.544)	(0.433)
Log of PC GDP of Importing Country	-0.135	3.150***	0.075	2.352***	-1.932*	-2.192**	-1.474***	-2.120***	-0.849*
	(0.546)	(0.617)	(0.450)	(0.677)	(0.795)	(0.669)	(0.424)	(0.531)	(0.427)
Log of PC GDP of Exporting Country	-5.528***	-2.113*	-1.954**	0.79	-1.743*	0.679	-0.454	1.579**	0.618
	(0.67)	(0.866)	(0.598)	(0.760)	(0.712)	(0.631)	(0.454)	(0.576)	(0.455)
Land Locked	0.503	-0.234	0.523*	-2.189	-1.688	-0.906*	0.507*	-0.218	-0.158
	(0.315)	(0.292)	(0.211)	(1.669)	(1.283)	(0.453)	(0.228)	(0.215)	(0.219)
Common Colonizer	1.137***	0.199	0.072	2.280***	1.176***	1.781***	1.752***	0.969***	1.451***
	(0.245)	(0.198)	(0.173)	(0.298)	(0.216)	(0.177)	(0.125)	(0.152)	(0.127)
Colony	0.896***	0.408***	0.457***	1.084***	1.307***	1.022***	0.403***	0.760***	0.868***
	(0.074)	(0.080)	(0.063)	(0.078)	(0.084)	(0.075)	(0.052)	(0.069)	(0.053)
Common Language	0.289***	0.504***	0.540***	0.348***	0.004	0.001	0.359***	-0.056	-0.063
	(0.048)	(0.054)	(0.041)	(0.053)	(0.058)	(0.051)	(0.035)	(0.046)	(0.035)

[A] Restricted model when Gini of exporting country is included in the analysis.
[B] Robust standard errors.

References:

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review, 69*(1), 106–16.

Anderson, J., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review, 93*(1), 170–92.

Baier, S. L., & Bergstrand, J. H. 2001. The growth of world trade: Tariffs, transport costs, and income similarity. *Journal of International Economics, 53*, 1–27.

Baier, S. L., & Bergstrand, J. H. 2009. Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics*, 77(1): 77-85

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics, 67*(3), 474–81.

Bergstrand, J. H. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *Review of Economics and Statistics, 71*(1), 143–53.

Bergstrand, J. H. (1990). The Heckscher-Ohlin-Samuelson model, the Linder hypothesis, and the determinants of bilateral intra-industry trade. *Economic Journal, 100*(4), 1216–29.

Bergstrand, J. H., Egger, P., & Larch, M. (2007). Gravity redux: Structural estimation of gravity equations with asymmetric bilateral trade costs. Working Paper. http://www.nd.edu/~jbergstr/working_papers.html.

Bikker, J. A. and A. F. de Vos. 1992. An international trade flow model with zero observations: An extension of the Tobit model. *Brussels Economic Review* 135: 379–404.

Center for International Business(CIB). 2007. Trade Agreement Database. Tuck School of Business, Dartmouth College, Hanover, NH (http://cibresearch.tuck.dartmouth.edu/).

Dalgin, M., Mitra, D., & Trindade, V. (2006). *Inequality, nonhomothetic preferences and trade: A gravity approach*. National Bureau of Economic Research.

Dow, W. H. and E. C. Norton. 2003. Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Services & Outcome Research Methodology* 4: 5–18.

Egger, P. 2000. A note on the proper specification of the gravity equation. *Economics Letters*, 66(1):25−31

Egger, P. 2002. An econometric view on the estimation of gravity models and calculation of trade potential. *The World Economy*, 297–312.

Egger, P. and Pfaffermayr, M. 2003. The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects. *Empirical Economics*, 28:571-80

Eichengreen B. and Irwin D. 1997. The role of history in bilateral trade flows, in: *The Regionalization of the World Economy*, Frankel J., ed., University of Chicago Press, 33–57

Frankel, J. A. (1997). *Regional trading blocs in the world economic system*. Washington DC: Institute for International Economics.

Greene, W. H. 2003. *Econometric analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Glick, R. and A. K. Rose. 2002. Does a currency union affect trade? Time series evidence. *European Economic Review* 46: 1125–1151.

Hallak, J. C. 2006. Product quality and the direction of trade. *Journal of International Economics* 68 (1): 238–265.

Hillberry, R. H. 2002. Aggregation bias, compositional change and the border effect. *The Canadian Journal of Economics* 35 (3): 517–30.

Heckman, J. 1979. Sample selection bias as a specification error. *Economterica* 47 (1): 153–61.

Helpman, E., M. Melitz and Y. Rubinstein. 2008. Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2): 441-87

Hoffmann, R. and A. L. Kassouf. 2005. Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure. *Applied Economics* 37: 1303–1311.

Hummels, David (1999) 'Toward a geography of trade costs,' mimeo, University of Chicago

Linnemann, H. (1966). *An econometric study of international trade flows*. Amsterdam: North Holland.

Linders, G. M. and H. L. F. de Groot. 2006. Estimation of the gravity equation in the presence of zero trade flow. Tinbergen Institute Discussion Paper No. TI 2006-072/3.

Maccallum, J. T. 1995. National borders matter: Canada-U.S. regional trade partners. *American Economic Review* 85: 615–23.

Maddala, G. S. 1983. Limited-dependent and qualitative variables in econometrics. Cambridge, UK: Cambridge University Press.

Mátyás, L. 1997. Proper econometric specification of the gravity model. *The World Economy* 20, 363–68.

Persson, Torsten and Guido Tabellini (2003), The Economic Effects of Constitutions (Cambridge, MA: The MIT Press).

Raballand, G. (2003). Determinants of the negative impact of being landlocked on trade: An empirical investigation through the central Asian case. *Comparative Economic Studies, 45*, 520–36.

Rose, A. K. (2000). One money, one market: The effect of common currencies on trade. *Economic Policy, 15*, 8–45.

Silva, J. M. C. Santos, and Silvana T. 2006. The log of gravity. *Review of Economics and Statistics*, 88: 641–658

Tinbergen, Jan. 1962. Shaping the world economy. New York: The Twentieth Century Fund

Wang, Z. K. and Winters, L.A. 1991. The Trading Potential of Eastern Europe, *CEPR Discussion Paper*, no. 610, London

Wooldridge, J. M. 2006. Introductory econometrics: A modern Approach. Third Edition Thomson South-Western