CATPRN
Canadian Agricultural Trade Policy And Competitiveness Research Network

THE GRAVITY MODEL AND THE PROBLEM OF ZERO'S IN AGRIFOOD TRADE

CATPRN Working Paper 2011-02
August 2011

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

In the analysis of bilateral trade flows, reported trade of zero or missing observations are quite common and this is a problem when estimating log-linear gravity equations. This has caused many researchers to either ignore the zero trade flows or to replace the 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 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

JEL Classification: F10; F12; F19; Q17; R12

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 geographical distance, 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 "trade resistance" between them. To account for trade resistance analysts have used the geographic distance between trading partners, and a variety of dummy variables including common borders, Commonwealth membership, common language, colonizers, religion, common currency and so on. The theoretical underpinnings of the gravity model were not established until the work of Anderson (1979), Bergstrand (1985), Helpman and Krugman (1985) and Anderson and van Wincoop (2003). Subsequently, 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 the per capita income of trading partners as explanatory variables for bilateral trade flows.

With the theoretical foundations of the gravity equation no longer in doubt 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 (Bucha *et al*., 2004, Egger and Pfaffermayr, 2003). This paper focuses on another important issue in estimating gravity equations, namely selection bias. 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 because of the log-linear functional specification and this non-random selection of data can lead to biased parameter estimates (Heckman, 1979).[i] Recent work by Jayasinghe, et. al. (2010) and Disdier and Marette (2010) note the potential bias caused by zero trade flows but selection bias is not the main focus of their work.

The primary objective of this study is to test for selection bias in gravity modelling, a research area also identified by Anderson and van Wincoop (2004). As a case study, we use agrifood trade data disaggregated to the four digit SITC level from 1990 to 2000, for 52 diverse countries and 46 agrifood products without excluding zero trade flows. Agrifood provides an interesting case study because trade in agrifood products, in contrast to industrial products, is often dominated by zero trade flows; making it an ideal candidate to study the effects of omitting "zero's" on parameter estimation. 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.[ii]

Our analysis makes at least two 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 Disdier and Marette (2010). Second, after accounting for zero trade flows using the Heckman selection model, conditional and unconditional marginal effects are derived making the interpretation of the estimated coefficients straight forward. Both the conditional and unconditional elasticities are important because there are several potential reasons for a bilateral trade flow to be recorded as zero: 1) trade is zero; 2) trade is close to zero and rounded off to zero; or 3) the volume of trade is missing in the UN data set. Unlike, Jayasinghe, et. al. (2010) who argue that a zero in their data is a true no trade situation, with UN data it is not possible to tell whether a zero represents a corner solution (an actual zero) or a potential zero, i.e. all other cases of no-trade. Dow and Norton (2003) argue that in this case, there is a selection problem that can be handled using the Heckman Maximum Likelihood (ML) estimation procedure and it has been commonly used in empirical work (Disdier and Marette 2010, Emlinger et al. 2008, Helpman et al. 2008, Jayasinghe, et. al.2010, Linders and de Groot 2006).

This article is organized into six sections. Section two discusses the derivation of the generalized gravity equation. Section three discusses the empirical 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 ZERO TRADE FLOWS

Recent theoretical developments in the trade literature account for the existence of zero trade flows. Eaton and Kortum (2002) derive the gravity model assuming incomplete specialization to justify the existence of zeros in trade data. Their model includes a limited number of suppliers, each selling to countries having low trade barriers. Although many potential suppliers exist an importer buys from the cheapest supplier, explaining the existence of zeros at the disaggregate level. Haveman and Hummels (2004) use a similar line of reasoning and suggest that zero trade occurs when production/consumption of a product in an exporting/importing country is zero. Anderson and van Wincoop (2003) maintain that the large fixed costs of trading help to explain the existence of zero trade flows. Finally, Helpman et al. (2008) use the profitability of a firm to explain the existence of zero bilateral trade flows at the aggregate level. They suggest that firms have varying levels of productivity and 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 importing country *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*. Using these assumptions, Helpman et al. (2008) derive a gravity model that accounts for firm heterogeneity and fixed trade costs that predicts 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*. The authors 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. The study also decomposes the impact of trade friction on trade flows into the intensive and extensive margins. The intensive margin is the volume of trade per exporter while the extensive margin refers to the number of exporters. The Helpman et al., (2008) model results in a generalized gravity equation that accounts for the self-selection of firms into export markets and their impact on trade volumes.

Helpman et al., (2008) go on to argue that the standard Heckman (1979) correction is valid in a world where firm heterogeneity is not correlated with the export decision. The implication is that all firms are identically affected by trade barriers and country characteristics; assumptions that seem reasonable for our empirical model. Details of the Heckman selection model are presented in section 4. The following section explains the generalized gravity equation used in the analysis.

3.0 GENERALIZED GRAVITY EQUATION

The focus of this paper is largely empirical and we use the generalized gravity equation as derived by Bergstrand (1989, 1990) for several reasons: 1) Bergstrand (1989, 1990) extended the microeconomic foundation of the gravity equation to theoretically justify the inclusion of income as well as the per capita income of exporters and importers in the generalized gravity equation, key variables in explaining agrifood trade; 2) our analysis assumes that agrifood products are differentiated, an assumption explicitly made by Bergstrand (1989); 3) Bergstrand (1990) allows for Linder-type preferences in the gravity model, behavior suitable to explain intra-industry trade; and 4) the representative firm in Bergstrand incurs both fixed and variables costs of production and trade, a behavioural assumption suitable to explain the existence of zero trade flows as discussed in section 2.

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 with variable and fixed costs to produce differentiated products. Firms allocate their products to different markets, in the face of fixed costs of exporting, to maximize profit in monopolistic 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 as a proxy for income inequality in these countries. Linder (1961), Francois and Kaplan (1996) and Dalgin, Mitra and Trindade (2006) argue that the more unequal the income distribution within a country, potentially the greater the share of expenditure on differentiated products. Linder (1961) goes further, suggesting that within country income distribution helps to determine the pattern of trade. Bergstrand's model allows for Linder-type preferences using the Gini coefficient as a measure of income distribution.

4.0 HECKMAN SELECTION MODEL

The problems caused by zero trade flows in gravity models is well known but until very recently has been 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 non-randomly 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 estimated model are always 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 four ways: 1) delete the observations with values equal to zero (McCallum 1995; Frankel 1997); 2) replace the zeros with small positive numbers (McCallum, 1995 and Raballand, 2003); 3) estimate the regression equation as a Tobit model and censor the zero observations at the left tail (Rose 2000); or 4) use the Heckman selection model (Emlinger et al. (2008), Disdier and Marette (2010), Jayasinghe et al. (2010)).[iii] Helpman et al. (2008) developed a variant of the Heckman selection model that has two-stages: an equation for selection into trade partners in the first stage, and a trade flow equation in the second. Linder and de Groot (2006) use a generalized gravity equation for both the selection and outcome equations. 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." Recently, Emlinger et al. (2008) and Disdier and Marette (2010) have used the Heckman selection model to study trade barriers and Jayasinghe et al. (2010) to study United States seed corn trade. Our study adds to the literature using the Heckman selection model to study agrifood trade, albeit in a much broader agrifood product context than the studies noted above and with an explicit focus on the impact of zero trade flows on parameter estimation and economic inference.

The Heckman sample selection model consists of sample selection (equation 1) and outcome equations (equation 2). Consider the following sample selection equation.

$$t_{ijfy}^{*} = \boldsymbol{\eta}'\boldsymbol{Z_i} + u_i \qquad (1)$$

where t_{ijfy}^{*} is a latent variable and it is not observed but we do observe if countries trade or not, such that $t_{ijfy} = 1$ if $t_{ijfy}^{*} > 0$ and $t_{ijfy} = 0$ if $t_{ijfy}^{*} = 0$ and $\boldsymbol{Z_i}$ is a vector of variables that affects t_{ijfy}^{*}. In the outcome equation (equation 2) let, T_{ijfy} be the natural logarithm of the value of country *i*'s per capita imports from country *j* of product *f* in year *y* and $\boldsymbol{X_i}$ is the vector of independent variables determining T_{ijfy}, so

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

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

$$E\left[T_{ijfy} | t_{ijfy} = 1\right] = \boldsymbol{\gamma}'\boldsymbol{X_i} + \rho\sigma_\varepsilon\lambda_i(\alpha_u) \qquad (3)$$

where the function $\lambda_i(\alpha_u) = \frac{\phi\left(\frac{\eta' z}{\sigma_u}\right)}{\Phi\left(\frac{\eta' z}{\sigma_u}\right)}$ is the inverse Mills ratio (IMR), ϕ is the standard normal density function and $\boldsymbol{\Phi}$ is the cumulative standard normal distribution function. Equation (3) estimates the expected values of T_{ijfy} when trade is observed (i.e. greater than zero). Greene (2003) shows that due to the correlation between $\boldsymbol{X_i}$ and the IMR a least squares regression of

T_{ijfy} on $\boldsymbol{X_i}$, omitting $\lambda_i(\alpha_u)$ produces an inconsistent estimator of $\boldsymbol{\gamma}'$. Also, standard regression techniques assume that ρ=0, thus eliminating the IMR in equation (3) and producing biased estimation results if the IMR is statistically significant. A OLS regression yields consistent estimators only if the expected value of the error is known and it is included in the regression — as the Heckman selection model does (Hoffmann and Kassouf, 2005).

Let X_{f_k} denote regressors common to both the selection and outcome equations and consider $\rho\sigma_\varepsilon = \beta_\lambda$, then the marginal effect for the regressor is

$$\frac{\partial E\left[T_{ijfy} | t_{ijfy} = 1\right]}{\partial X_{f_k}} = \boldsymbol{\gamma}_{f_k} - \frac{\eta_{f_k}}{\sigma_u}\beta_\lambda \lambda_i(\alpha_u)[\lambda_i(\alpha_u) - \alpha_u] = \boldsymbol{\gamma}_{f_k} - \frac{\eta_{f_k}}{\sigma_u}\beta_\lambda \delta_i \tag{4}$$

where $\delta_i = \lambda_i(\alpha_u)[\lambda_i(\alpha_u) - \alpha_u]$. 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 values 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. Assume now that Z_{f_k} is a binary variable. Let $\bar{z}_0$ be the vector of explanatory variables in the participation equation with X_{f_k} equal to zero, and all other variables at their mean values and $\bar{z}_1$ be the same vector when X_{f_k} is equal to one. Then the change in the IMR ($\Delta\lambda$) for $\bar{z}$, when it moves from $\bar{z}_1$ to $\bar{z}_0$ is $\frac{\phi\left(\frac{\eta'\bar{z}_1}{\sigma_u}\right)}{\Phi\left(\frac{\eta'\bar{z}_1}{\sigma_u}\right)} - \frac{\phi\left(\frac{\eta'\bar{z}_0}{\sigma_u}\right)}{\Phi\left(\frac{\eta'\bar{z}_0}{\sigma_u}\right)}$. Hence, the conditional marginal effect for the binary variable is

$$\frac{\partial E\left[\Delta T_{ijfy} | t_{ijfy} = 1\right]}{\partial Z_{f_k}} = \boldsymbol{\gamma}_{f_k} + \beta_\lambda \Delta\lambda \tag{5}$$

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 the generalized gravity model, the unconditional marginal effect for a continuous variable that is common to both the selection and outcome equations is

$$\frac{\partial E\left[T_{ijfy}\right]}{\partial X_{f_k}} = \gamma_{f_k} - \frac{\eta_{f_k}}{\sigma_u}\beta_\lambda \delta_i + \left[\Phi\left(\frac{\eta' Z_i}{\sigma_u}\right)\right]^{-1} \phi\left(\frac{\eta' Z_i}{\sigma_u}\right)\frac{\eta_{f_k}}{\sigma_u} \tag{6}$$

The first two terms on the right hand side of equation (6) show the change in trade of agrifood product *f* for the trading partners having observable trade flows (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

$$\frac{\partial E\left[\Delta T_{ijfy}\right]}{\partial Z_{f_k}} = \boldsymbol{\gamma}_{f_k} + \beta_\lambda \Delta\lambda\ + \Delta ln\Phi(-\alpha_u) \tag{7}$$

where $\Delta ln\Phi(-\alpha_u) = ln\Phi\left(\frac{\eta' z_1}{\sigma_u}\right) - ln\Phi\left(\frac{\eta' z_0}{\sigma_u}\right)$. Since the marginal effects vary for each observation we calculate these effects at the mean values.

The existing studies that use the 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. Hillberry (2002) estimates a more restricted variant of the gravity model in which an independent selection equation is estimated. Helpman et al. (2008) estimate 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. Our empirical versions of equations (1) and (2) are

$$t^*_{ijfy} = \begin{matrix} \pi_i + \pi_j + \pi_y + \pi_f + \eta_0 + \eta_1 \ln(dist_{ij}) + \eta_2 DCB_{ij} + \eta_3 DPTA_{ijy} + \\ \eta_4 \ln(Gini_{iy}) + \eta_5 \ln(Gini_{jy}) + \eta_6 \ln(GDP_{iy}) + \eta_7 \ln(GDP_{jy}) + \\ \eta_8 \ln(PCGDP_{iy}) + \eta_9 \ln(PCGDP_{jy}) + \eta_{10} DLandl_i + \eta_{11} DComcol_i + \\ \eta_{12} DColony_{ij} + \eta_{13} DComlang_{ij} + u_i \end{matrix} \quad (1')$$

$$\ln(T_{ijfy}) = \begin{matrix} \psi_i + \psi_j + \psi_y + \gamma_0 + \gamma_1 \ln(dist_{ij}) + \gamma_2 DCB_{ij} + \gamma_3 DPTA_{ijy} + \\ \gamma_4 \ln(Gini_{iy}) + \gamma_5 \ln\left(Gini_{jy}\right) + \gamma_6 \ln(GDP_{iy}) + \gamma_7 \ln(GDP_{jy}) + \\ \gamma_8 \ln(PCGDP_{iy}) + \gamma_9 \ln(PCGDP_{jy}) + \gamma_{10} DLandl_i + \gamma_{11} DComcol_i + \\ \gamma_{12} DColony_{ij} + \gamma_{13} DComlang_{ij} + \gamma_{14} AHT(\rho) + \gamma_{15} \ln(\sigma_u) + \varepsilon_i \end{matrix} \quad (2')$$

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. Importing and exporting country, year and commodity fixed effects are included 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. Fixed effects provide a solution to unobserved heterogeneity (Mátyás, 1997 and Egger, 2002) and account for the 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 (Baier and Bergstrand, 2001, Bergstrand et al., 2007; Baier and Bergstrand, 2009). Two caveats are in order. First, technical and non-technical barriers to trade are unlikely to have changed much over the study period but the estimated models may not fully account for the changes over time since we used year fixed effects rather than time varying fixed effects in the selection and outcome equations. Second, equations $1'$ and $2'$ use macroeconomic variables to explain country level aggregations of agrifood trade flows. Particular firms included in the aggregations might be influenced by firm specific factors that are excluded from our analysis. However, there is no reason to think that these firm level variables are correlated with the error terms of the empirical model in our aggregations over many firms and several commodities. This is similar to the approach taken by Bergstrand (1990), Chow et al. (1999) and Fillat-Castejón and Serrano-sanz (2004) in examining the trade flows of manufactured products. In contrast, Anderson and van Wincoop (2004), Henry de Frahan and Vancauteren (2006) and Emlinger et al. (2008) use sector production and expenditure data but their sample consists of developed countries for which such data is available. Our sample includes 36 developing countries and it is impossible to find capital (or labor) intensities for such a diverse group over a long period of time.

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) 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". This condition is also known as the exclusion restriction.[iv] In our case, the selection equation, in-addition to other variables, includes commodity fixed effects. Exclusion of the commodity fixed effects from the outcome equation provides the exclusion restriction and allows for the calculation of sectoral marginal effects. Finally, the Heckman model is estimated simultaneously using the ML procedure to get homoscedastic standard errors (Greene, 2003).

5.0 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 reconciles United Nations bilateral trade data to develop the WTA and also incorporates zeros into the data. Haveman and Hummels (2004) also use the Statistics Canada data set in their empirical analysis. Statistics Canada stopped updating the World Trade Analyis database in 2004. Given the cost of this database, the fact it is not being updated and the focus of our work we decided to limit our analysis to data from 1990 to 2000. We categorize 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. 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. Observations in each sector are not aggregated. For example, the meat sector includes trade flows of products 0111, 0112, 0113, 0121 and 0149 for 52 countries from 1990 to 2000. The countries included in our sample by income category are: Lower income (Bangladesh, Ethiopia, India, Madagascar, Pakistan and Tanzania); Lower Middle Income (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 (Argentina, Chile, Costa Rica, Hungary, Malaysia, Mexico, Panama, Poland, South Africa, Turkey, Uruguay and Venezuela); and High income (Canada, Denmark, Finland, Germany, France, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and the United States).

Figure 1 shows the proportion of zero trade flows for the selected beverage and agrifood product imports from 1990 to 2000. The outcome equation (which does not use observations with a zero value) contains, on average, 43 percent of the observations. The highest number of censored observations (51%) is in the meat group, while the alcoholic beverage sector has the lowest number (36%) of censored observations. 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 countries are in US dollars, the US GDP deflator is used to obtain 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 on common language, common colonizer and colony are compiled from Glick and Rose (2002).

6.0 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 OLS. The results of the Heckman ML procedure for the outcome equation are reported in table 2 while the results for the selection equation are reported in table 3.

The Wald-test shows that all of the models except cereals are statistically significant at the 99 percent level of significance (Table 2). The Wald-test rejects the hypothesis that all of the coefficients in the outcome equation, except the constant, are zero. 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) equals zero. Rho indicates the correlation between the error terms of the outcome and selection equations; that is, the null hypothesis of cor(u_i,ε_i) = 0 as against the alternative hypothesis of cor(u_i,ε_i) $\neq$ 0. Failure to reject this null hypothesis indicates insignificant sample selection bias, while rejection of the null means OLS produces biased estimates. The null hypothesis that rho = 0 is rejected for all of the product groups and from that we conclude that the use of the Heckman procedure is appropriate. It is important to recognize that ML estimation of the Heckman sample selection model does not provide a value for rho and its standard error. Instead the arc hyperbolic tangent of rho $ln\left(\frac{1+\rho}{1-\rho}\right)$ and the natural logarithm of its standard error are estimated. 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 zero trade flows produce biased estimates. Jayasinghe, et. al. (2010) and Disdier and Marette (2010) obtained similar results. Finally, the ML procedure consistently estimates rho within the range of -1 and 1, suggesting that both the selection and outcome equations have no specification error (StataCorp, 2007).

Collectively these results imply that estimation without considering zero trade flows produces biased estimates, i. e. OLS estimates (based on ignoring zeros) are biased. Helpman et al. (2008) find that the omission of zero trade flows does not yield any selection bias as their estimates both incorporating and ignoring zero trade flows are very similar. However, these authors use different techniques and aggregate data with only a few zero trade flows.

Importing and exporting country, and year fixed effects are included in the models to account for unobserved heterogeneity. Ethiopia is the base importing and exporting country for the fixed effects while 1990 is the base year. The base product in the meat equation is 0149 (other prepared or preserved meat), in dairy 0980 (edible products and preparations), in fresh fish 0350 (fish, dried, salted or in brine), in frozen fish 0372 (crustaceans), in cereals 0488 (preparation of flour), in fresh fruits 0579 (other fresh or dried fruit), in processed fruits 1110 (non alcoholic beverages), in vegetables 0565 (prepared or preserved vegetables), in tea and coffee 0741 (tea) and for alcoholic beverages 1124 (other spirits and liqueurs). These fixed effects are tested with the null hypothesis that their combined effect is 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 would produce biased estimates.

Table 2 contains the estimates of the outcome (regression) equation of the generalized gravity model. The results show that the estimated coefficients all have the expected signs. The distance variable is expected to have a negative sign because an increase in the distance between trading partners is expected to decrease trade. Similarly, variables representing countries with a common border, preferential trade agreements (PTA), common colonizer and trading partners that colonized each other (colony) are expected to have positive signs. All of the estimates for distance are negative and for a common border and a PTA positive and statistically significant, with the exception of the meat sector. Similarly, estimates for a 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 significant. These results show that the model produces the expected results — a sign of suitable specification.

Bergstrand (1990) argues that an exporter’s GDP shows its national output, but an exporter’s per capita GDP is a proxy for its capital-labor ratio; while an importer’s GDP depicts its national output and the importer’s per capita GDP represents its per capita income. Bergstrand (1990) goes on to argue that since the exporter’s per capita income is a proxy for the exporter’s per capita capital, the sign of the exporter’s per capita income variable determines if the product is produced in a capital- or labor-intensive industry. A positive (negative) and statistically significant coefficient for the exporter’s per capita income variable (a proxy for exporter’s capital-labor ratio) suggests that the sector is capital (labor) intensive in production. Like Bergstrand (1990) we also consider the exporter’s per capita income a proxy for the capital-labor ratio because capital and labor data for our sample of countries and commodities, over time, is not available. Bergstrand (1990) shows that if a good is a luxury in consumption and capital intensive in production then the elasticity of substitution for all goods in a sector exceeds unitary. A positive and statistically significant coefficient for exporter income for each sector implies that each sector’s elasticity of substitution exceeds unitary. A positive and statistically significant coefficient for importer income for a sector implies that an increase in importer income increases sector specific imports.

In table 2, the sectoral elasticity of substitution exceeds one for alcoholic beverages and cereals only. Importer's income is negative and statistically significant for cereals, dairy, fresh and frozen fish, fresh and processed fruit and vegetables indicating that as the income of importing countries' increase their imports in these sectors decrease. The estimates show that dairy, fresh fish, fresh fruit, tea and vegetables are capital-intensive sectors while the alcoholic beverage sector is labor intensive. The importer's per capita income shows that cereals, dairy, fresh and frozen fish, fresh fruits, tea and vegetables sectors are luxury goods while processed fruit is a necessity.

Table 3 presents estimates of the selection equation estimated as a Probit model. Commodity fixed effects, omitted from the outcome equation to obtain sectoral marginal effects, are included in the selection equation. All of the gravity type variables have the expected signs with the exception of land locked countries for alcoholic beverages and trading partners ever in a colonial relationship for the meat sector. In the next section, we examine the marginal effects of the explanatory variable.

6.1 Marginal Effects

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 from the selection model. The derived elasticities can either be 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 (conditional) or if they are a potential zero trade flow (unconditional) as a result of missing or misreported data. Since we don't know if the zero trade flows are cases of no trade or instances where trade is not reported, we derive both the conditional and unconditional elasticities (Tables 4 and 5). The conditional marginal effect of the importers' income on alcoholic beverages shows that the value of trade increases by 8.1 percent for a 10 percent increase in the importers' income, among those countries who have positive trade flows, *cetris peribis* (Table 4). The unconditional marginal effect for the same variable is larger, 8.5 percent for a 10 percent increase in income, because it accounts for the larger proportion of countries engaging in trade after the increase in income, *cetris peribis* (Table 5). The conditional elasticities are conceptually similar to the ones estimated using OLS (Table 6). In the case of OLS, a 10 percent increase in importer's income increases trade in alcoholic beverages by 7.2 percent, *cetris peribis* (Table 6).

The summary statistics in table 6 shows that all of the models explained the data well except for meat and alcoholic beverages, but all of the models are statistically significant. The next section contains a comparison of the estimated elasticities across the estimation methods.

6.2 Does It Really Matter?

It is useful to compare the elasticities estimated using OLS which ignores the zero trade flows (Table 6) with the conditional elasticity estimates using Heckman estimation (Table 4) 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 acceptable. We approach the comparison from both a statistical and an economics point of view.

In our model there are 130 parameters related to variables based on the gravity model. Our comparison focuses on these parameters and ignores the parameters capturing the fixed effects. For one key variable, the importers' GDP, using Heckman the elasticities for alcoholic beverages (0.81), fresh fruit (0.76) and processed fruit (0.87) are significant and positive while they are statistically zero using OLS. Switching our focus to the 62 parameters that are statistically significant using both estimation techniques, of these, there are three cases where a parameter changes sign between the two estimation techniques. Two of the sign switches relate to alcoholic beverages where the PTA (0.20; -0.36) and common language (0.12; -0.23) coefficients are positive using Heckman and negative using OLS. The third sign switch is for the importing countries GDP in the tea, coffee and mate equation where it is negative using Heckman (-1.91) and positive using OLS (0.99). So sign switches, while rare are present. 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 concentrate on whether the effect becomes more or less elastic, without worrying about the sign of the coefficient. For 71 percent of the elasticities the Heckman estimate is more elastic than the OLS estimate, 13 percent of the elasticities differ by more than 50 percent and 32 percent of the elasticities differ by more than 20 percent. For the three variables accounting for the effect of being a colonizer, being a colony and having a common language of the 20 parameter estimates that are significant with both estimation techniques 15 are within 10 percent of each other. However, for the variables distance, having a common border and having a PTA of the 26 parameter estimates only eleven are within 10 percent of each other. Some of the differences in parameter estimates are strikingly large, for example using the Heckman estimate for the effect of distance, on the trade of fresh fish, the elasticity is -1.26 while it is only -0.60 using OLS (Tables 4 and 6). As a final check, we look at the number of times an elasticity estimate changed from elastic using Heckman to inelastic using OLS. This happened five times. In some cases the absolute difference between the two coefficients is small (for exporters GDP in vegetables trade: -1.00 (Heckman) and -0.97 (OLS)) but in other cases it is quite large (for being land locked in fresh fish trade: 1.28 (OLS) and -0.80 (Heckman)).

It is also meaningful to compare the elasticities from the conditional (Table 4) and the unconditional (Table 5) estimates to see how much difference the assumption about the nature of the zero trade flows makes. There are 69 common statistically significant elasticities among the conditional and unconditional estimates. There are no sign reversals and most of the unconditional estimates are more elastic than their conditional counterparts. Twenty-three percent of the unconditional elasticities are more than 50 percent more elastic than the conditional estimate and 62 percent are more than 20 percent more elastic. There are eight 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: -1.61(conditional) and 0.99 (unconditional); and for distance in fresh fish trade: -1.77 (conditional) and -0.60 (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 researchers do not know the true nature of the recorded data. However, based on our case study the conditional elasticities will most often be more inelastic than their unconditional counterparts.

7.0 CONCLUSION

The fact that two countries often do not trade individual agrifood products (or product aggregations) with one another can be explained using the recent theories of Helpman et al. (2008) based on the characteristics of individual firms[v]. For the gravity model, which is a common framework used to examine a wide variety of bilateral trade issues, zero trade flows are challenging. This is because the dependent variable in gravity models is generally specified as the logarithm of the value of trade and the logarithm of zero is undefined. As a result, researchers have often ignored zero trade flows in their analysis or replaced the zero with a small positive number.

The omission of zero trade flows from quantitative analysis raises the issue of sample selection bias. Helpman et al. (2008) argue that sample selection bias is small in their sample when parameter estimates calculated by ignoring zero trade flows are compared with parameter estimates calculated using estimation procedures that take into account potential sample selection bias.

Our work examines the question of selection bias using data on bilateral trade flows for 46 agrifood products among 52 diverse countries. In our sample, zero trade is common with 43 percent of the trade flows equalling zero. We find that selection bias is a serious problem and it can lead to incorrect statistical inferences as well as parameter estimates that differ in economically meaningful ways between ordinary least squares estimates, which ignore zero trade flows, and estimates based on the Heckman selection model that corrects for sample selection bias. Since the true nature of a zero trade flow is unknown (is it an actual zero trade flow or missing data) we calculate both conditional and unconditional marginal effects for the variables that are common to the selection and outcome equations. We find that ignoring selection bias rarely effects the signs of variables but often influences the magnitude, statistical significance and economic interpretation of the marginal effects. This conclusion differs sharply from that of Helpman, et al. (2008). Our results suggest that whenever zero trade flows are prominent, in a trade data set, selection bias should be taken into account in estimating the gravity model.

REFERENCES

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

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

Baier, L. and Bergstrand, J. (2001). The growth of world trade: tariffs, transport costs, and income similarity. *Journal of International Economics* 53:1–27.

Baier, L. and Bergstrand, J. (2007). Do free trade agreements actually increase members' international trade. *Journal of International Economics* 71: 72–95.

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

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

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

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

Bergstrand, J., Egger, P. and Larch, M. (2007). Gravity redux: Structural estimation of gravity equations with asymmetric bilateral trade costs. Working Paper. Accessed on January 20, 2008. http://www.nd.edu/~jbergstr/working_papers.html.

Bucha, C., Kleinerta, J. and Toubalb, F. (2004). The distance puzzle: on the interpretation of the distance coefficient in gravity equations. *Economics Letters* 83:293–298.

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

Chow, P., Kellman, M. and Shachmurove, Y. (1999). A test of the Linder hypothesis in pacific NIC trade 1965–1990. *Applied Economics* 31: 175–82.

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

Disdier, A. and Marette, S. (2010). The Combination of Gravity and Welfare Approaches for Evaluating Nontariff Measures. *American Journal of Agricultural Economics* 92:713-726.

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

Eaton, J. and Kortum, S. (2002).Technology, geography and trade. *Econometrica* 70:1741–79.

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.

Emlinger, C., Jacquet, F. and Lozza, E. (2008). Tariffs and other trade costs: assessing obstacles to Mediterranean countries' access to EU-15 fruit and vegetable markets. *European Review of Agricultural Economics* 35: 409–438.

Fillat-Castejón, C. and Serrano-sanz, M. (2004). Linder revisited: trade and development in the Spanish economy. *International Review of Applied Economics* 18:323–48.

Francois, F. and Kaplan, S. (1996). Aggregate demand shifts, income distribution, and the Linder hypothesis. *Review of Economics and Statistics* 78:244–50.

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

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

Haq, Z., K. D. Meilke and J. Cranfield. 2010. *Does the gravity model suffer from selection bias?,* CATPRN Working Paper 2010-01, Canadian Agricultural Trade Policy and Competitiveness Research Network. Guelph.

Haveman, J. and Hummels, D. (2004). Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization. *Canadian Journal of Economics* 37:199–218.

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

Helpman, E. and Krugman, R. (1985). *Market structure and foreign trade. Increasing returns, imperfectcCompetition, and the international economy*, Cambridge, MA: MIT Press.

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

Henry de Frahan, B. and Vancauteren, M. (2006). Harmonisation of food regulations and trade in the Single Market: evidence from disaggregated data. *European Review of Agricultural Economics* 33:337–360.

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

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

Jayasunghe, S., Beghin, C. and Moschini, G. (2010). Determinants of world demand for U.S. corn seeds: the role of trade costs. *American Journal of Agricultural Economics* 92:999-1010.

Linder, B. (1961). *Essay on trade and transformation*. New York: John Wiley.

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

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

Martin, W. and Pham, C. (2008). Estimating the Gravity Model When Zero Trade Flows are Frequent." MPRA Paper No. 9453. Accessed on November 1, 2010. http://mpra.ub.uni-muenchen.de/9453/1/MPRA_paper_9453.pdf

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

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

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

Silva, S. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics* 88: 641–658

StataCorp. (2007). Stata 10 Base Reference Manual. College Station, TX: Stata Press.

Takayama, T. And G. Judge. (1971). Spatial and temporal price and allocation models. Amsterdam: North-Holland.

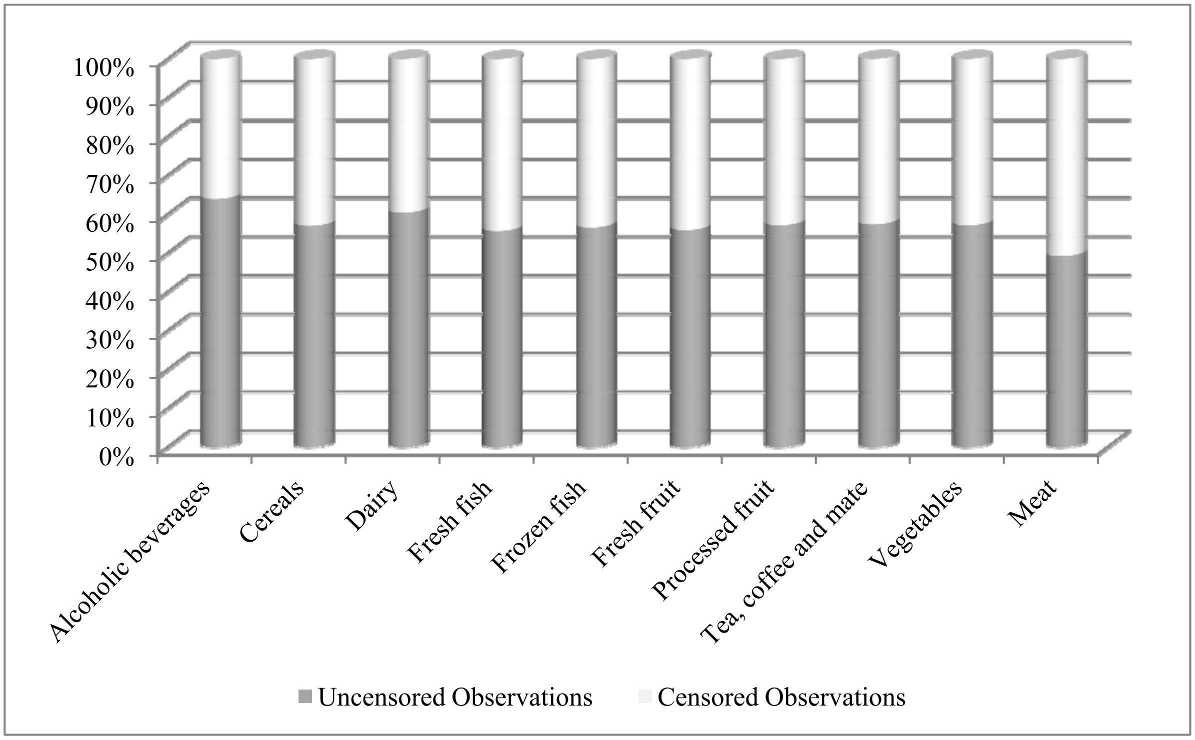
Tinbergen, J. (1962). Shaping the world economy. New York: The Twentieth Century Fund.

Wang, K. and Winters, A. (1991). The Trading Potential of Eastern Europe, *CEPR Discussion Paper*, No. 610, London.

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

Table 1: Description of Variables

Variable/Symbol	**Description**
T_{ijfy}	Real value of country *i*'s imports from 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
$\pi_i, \pi_j, \pi_f, \pi_y$	Respectively importing, exporting, product, and year fixed effects in selection equation
ψ_i, ψ_j and ψ_y	Respectively importing, exporting, and year fixed effects in outcome equation
$dist_{ij}$	Distance between bilateral trade partners
DCB_{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 preferential 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; zero otherwise
$DComlang_{ij}$	Binary variable, which is unity if bilateral trade partners have a common language and zero otherwise
$AHT(\rho)$	Arc hyperbolic tangent of rho. Rho shows correlation between error terms of the selection and outcome equations.
σ_u	Standard errors of the selection equation
η_0 and γ_0	Respectively intercepts of selection and outcome equations
ln	Natural logarithm
u_i and ε_i	Error terms of the selection and outcome equations



Data Source: World Trade Analyzer (WTA), Statistics Canada

Figure 1: Censored and uncensored observations for beverages and agrifood products (1990-2000)

Table 2: Estimates of the regression equation for beverages and agrifood products estimated using Heckman selection model

Variable	Alcoholic beverages	Cereals	Dairy	Fresh fish	Frozen fish	Fresh fruit	Processed fruit	Tea, coffee & mate	Vegetables	Meat
Distance	-0.996***	-1.000***	-0.682***	-0.602***	-0.506***	-0.428***	-0.958***	-0.169***	-0.739***	0.014
	(0.024)	(0.126)	(0.022)	(0.032)	(0.035)	(0.030)	(0.019)	(0.022)	(0.022)	(0.025)
Common Border	0.842***	0.770***	0.414***	0.675***	0.791***	0.374***	0.760***	0.546***	0.595***	0.126
	(0.069)	(0.091)	(0.069)	(0.083)	(0.068)	(0.073)	(0.050)	(0.074)	(0.047)	(0.080)
PTA	0.415***	0.989***	1.448***	1.152***	1.122***	0.571***	0.773***	0.660***	0.917***	-0.368***
	(0.079)	(0.105)	(0.074)	(0.098)	(0.083)	(0.079)	(0.053)	(0.083)	(0.060)	(0.112)
Importer Income	0.892*	-2.418***	-1.245**	-1.542**	-2.835***	-2.652***	2.730***	-3.006***	-0.315	0.852
	(0.517)	(0.617)	(0.465)	(0.773)	(0.632)	(0.678)	(0.409)	(0.531)	(0.395)	(0.548)
Exporter Income	5.374***	1.590*	-2.337***	-2.635***	-0.399	-1.633**	0.273	-1.814**	-1.051**	-0.278
	(0.689)	(0.899)	(0.593)	(0.738)	(0.741)	(0.642)	(0.450)	(0.614)	(0.419)	(0.679)
Importer Per Capita Income	-0.099	3.165***	1.447**	1.627**	3.114***	2.874***	-1.515***	2.932***	0.911**	-0.974*
	(0.548)	(0.623)	(0.487)	(0.806)	(0.662)	(0.706)	(0.429)	(0.555)	(0.411)	(0.588)
Exporter Per Capita Income	-5.387***	-1.241	2.583***	2.689***	0.766	2.066**	-0.367	2.016**	1.157**	0.287
	(0.725)	(0.953)	(0.624)	(0.775)	(0.784)	(0.672)	(0.479)	(0.648)	(0.441)	(0.712)
Importer Gini	0.026	-0.038	0.066	-0.236	-0.022	-0.076	-0.042	-0.085	-0.125	0.092
	(0.115)	(0.136)	(0.105)	(0.146)	(0.126)	(0.136)	(0.089)	(0.124)	(0.084)	(0.134)
Exporter Gini	0.104	-0.021	0.021	-0.09	0.087	0.074	0.15	-0.1	0.11	-0.098
	(0.112)	(0.142)	(0.109)	(0.143)	(0.131)	(0.132)	(0.092)	(0.122)	(0.086)	(0.134)
Land Locked	0.366	-0.18	0.902***	0.644	-2.248***	-1.455***	0.450*	-0.3	-0.132	-0.126
	(0.276)	(0.289)	(0.189)	(0.739)	(0.564)	(0.344)	(0.235)	(0.206)	(0.194)	(0.224)
Common Colonizer	1.049***	0.021	0.143	-0.056	2.001***	0.811***	1.697***	0.665***	1.145***	-0.095
	(0.211)	(0.199)	(0.164)	(0.217)	(0.415)	(0.178)	(0.146)	(0.177)	(0.142)	(0.198)
Colony	0.897***	0.184	0.171**	0.424***	0.604***	0.477***	0.393***	0.249***	0.513***	0.332***
	(0.076)	(0.123)	(0.070)	(0.090)	(0.080)	(0.080)	(0.054)	(0.075)	(0.054)	(0.088)
Common Language	0.270***	0.336***	0.147**	-0.238***	-0.057	-0.048	0.358***	-0.278***	-0.224***	-0.241***
	(0.049)	(0.077)	(0.047)	(0.062)	(0.054)	(0.055)	(0.037)	(0.051)	(0.034)	(0.055)
Selection Bias										
Arc Hyperbolic	1.388***	-0.171	-0.952***	-1.023***	-0.597***	-0.941***	1.138***	-1.335***	-0.239***	-0.296***

Tangent of Rho	(0.065)	(0.311)	(0.023)	(0.034)	(0.068)	(0.032)	(0.047)	(0.025)	(0.040)	(0.019)
ln(sigma)	0.796***	0.521***	0.731***	0.831***	0.672***	0.841***	0.783***	0.930***	0.648***	0.920***
	(0.012)	(0.023)	(0.006)	(0.009)	(0.015)	(0.009)	(0.011)	(0.006)	(0.005)	(0.005)
Fixed Effects										
Importing Countries	2087.0***	684.6***	759.7***	1162.1***	888.4***	600.1***	1362.9***	466.4***	1132.1***	1915.7***
Exporting Countries	7183.4***	1520.4***	1138.6***	1390.8***	1753.5***	2918.3***	6000.0***	1534.5***	3735.6***	319.8***
Year	32.8**	21.0**	21.8**	22.5**	12.2	9.7	50.8***	29.0**	15.6	16.8*
Summary Statistics										
Total Observations	35145	20273	38533	28776	28875	36707	66077	44088	67386	52085
Censored Observations	12673	8707	15248	12753	12535	16184	28293	18782	28873	26451
LR Test	4201.5***	3414.2***	8279.5***	4108.0***	4200.3***	4777.4***	19204.4***	5618.3***	7423.8***	2424.3***
Wald Chi	687.4***	2.02	1936.1***	803.4***	50.99***	840.4***	582.7***	2317.5***	55.3***	26.5***

*, **, *** respectively show significance at 90, 95 and 95 percent levels. Robust standard errors are given in parentheses.

Table 3: Selection 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	Meat
Distance	-0.434***	-0.696***	-0.456***	-0.551***	-0.460***	-0.485***	-0.418***	-0.267***	-0.538***	-0.019**
	(0.013)	(0.021)	(0.013)	(0.015)	(0.015)	(0.014)	(0.009)	(0.010)	(0.010)	(0.009)
Common Border	0.408***	0.314***	0.473***	0.112**	0.076	0.381***	0.364***	0.278***	0.317***	0.001
	(0.044)	(0.074)	(0.043)	(0.047)	(0.047)	(0.035)	(0.031)	(0.034)	(0.030)	(0.028)
PTA	0.228***	0.458**	0.306***	0.188**	0.502***	-0.066	0.384***	0.194***	0.259***	0.031
	(0.059)	(0.175)	(0.072)	(0.077)	(0.097)	(0.044)	(0.044)	(0.040)	(0.058)	(0.040)
Importer Gini	0.082	-0.049	0.936***	1.923***	0.758**	3.575***	1.814***	2.202***	1.878***	-0.328*
	(0.272)	(0.368)	(0.267)	(0.361)	(0.319)	(0.309)	(0.201)	(0.235)	(0.208)	(0.200)
Exporter Gini	3.237***	0.733	2.416***	2.132***	0.251	0.641***	0.332	-0.076	0.164	0.000
	(0.327)	(0.468)	(0.320)	(0.346)	(0.368)	(0.290)	(0.214)	(0.254)	(0.218)	(0.245)
Importer Income	0.132	0.486	-0.510*	-1.218***	-0.169	-3.054***	-1.273***	-1.926***	-1.304***	0.254
	(0.286)	(0.388)	(0.279)	(0.378)	(0.339)	(0.326)	(0.211)	(0.247)	(0.219)	(0.213)
Exporter Income	-3.325***	-0.811*	-2.574***	-2.347***	-0.056	-0.781**	-0.505**	0.037	-0.313	0.066
	(0.341)	(0.488)	(0.335)	(0.361)	(0.387)	(0.302)	(0.225)	(0.268)	(0.227)	(0.256)
Importer Per Capita Income	0.100*	-0.036	0.006	0.200**	0.130*	0.038	-0.001	0.109**	0.075	-0.081*
	(0.060)	(0.084)	(0.059)	(0.069)	(0.068)	(0.062)	(0.044)	(0.054)	(0.046)	(0.047)
Exporter Per Capita Income	0.032	0.046	-0.117*	-0.039	0.074	-0.003	0.080*	0.054	0.185***	0.004
	(0.059)	(0.088)	(0.062)	(0.067)	(0.070)	(0.060)	(0.044)	(0.051)	(0.047)	(0.048)
Land Locked	0.304**	0.072	-0.191	-0.986**	-0.562	0.424**	0.213*	0.034	0.053	0.097
	(0.144)	(0.169)	(0.136)	(0.446)	(0.568)	(0.165)	(0.111)	(0.094)	(0.112)	(0.098)
Common Colonizer	0.238**	0.242**	0.077	0.579***	0.316**	0.536***	0.610***	0.118	0.386***	0.370***
	(0.100)	(0.119)	(0.093)	(0.105)	(0.147)	(0.079)	(0.062)	(0.075)	(0.060)	(0.089)
Colony	0.423***	0.415***	0.211***	0.519***	0.307***	0.361***	0.219***	0.281***	0.374***	-0.116***
	(0.044)	(0.055)	(0.040)	(0.044)	(0.043)	(0.038)	(0.029)	(0.035)	(0.030)	(0.030)
Common Language	0.156***	0.316***	0.369***	0.144***	0.333***	0.025	0.313***	0.116***	0.191***	0.018
	(0.027)	(0.044)	(0.026)	(0.029)	(0.029)	(0.026)	(0.019)	(0.023)	(0.019)	(0.020)
Fixed Effects										
Importing Countries	1360.9***	750.1***	956.6***	1172.1***	693.6***	1288.9***	1598.3***	1045.6***	1519.3***	762.6***
Exporting Countries	3243.4***	1673.5***	2573.6***	1817.4***	2630.9***	3404.8***	5215.7***	2929.9***	4967.2***	304.2***
Year	50.3***	90.3***	13.2	22.0**	25.9**	17.9*	109.5***	49.5***	42.1***	120.9***
Commodity	84.3***	41.8***	5197.0***	2869.7***	852.1****	3201.5***	268.4***	3716.6***	3560.6***	159.0***

*, **, *** respectively show significance at 90, 95 and 95 percent levels. Robust standard errors are given in parentheses.

Table 4: 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**	**Meat**
Distance	-0.556***	-1.121***	-1.177***	-1.263***	-0.757***	-1.014***	-0.530***	-0.302***	-0.895***	0.023
	(0.021)	(0.127)	(0.022)	(0.034)	(0.037)	(0.030)	(0.018)	(0.020)	(0.022)	(0.024)
Common Border	0.469***	0.822***	0.898***	0.808***	0.832***	0.814***	0.414***	0.666***	0.683***	0.126
	(0.069)	(0.091)	(0.069)	(0.083)	(0.068)	(0.073)	(0.050)	(0.074)	(0.047)	(0.080)
PTA	0.196**	1.062***	1.768***	1.372***	1.361***	0.491***	0.411***	0.747***	0.989***	-0.354**
	(0.079)	(0.105)	(0.074)	(0.098)	(0.083)	(0.079)	(0.053)	(0.083)	(0.060)	(0.112)
Importer Income	0.809*	-2.426	-0.230	0.764	-2.421***	1.671**	0.872***	-1.911***	0.232	0.700
	(0.419)	(0.616)	(0.448)	(0.699)	(0.607)	(0.649)	(0.347)	(0.478)	(0.394)	(0.542)
Exporter Income	2.095***	1.718*	0.284	-0.080	-0.262	-0.857	-0.067	-1.852***	-1.003**	-0.277
	(0.569)	(0.889)	(0.575)	(0.683)	(0.717)	(0.586)	(0.392)	(0.556)	(0.417)	(0.670)
Importer Per Capita Income	-0.232	3.249***	0.893	0.167	3.022***	-0.818	-0.211	1.974***	0.532	-0.856
	(0.443)	(0.624)	(0.470)*	(0.725)	(0.637)	(0.680)	(0.361)	(0.500)	(0.410)	(0.581)
Exporter Per Capita Income	-2.020***	-1.382	-0.211	-0.124	0.735	1.122*	0.151	2.034***	1.066**	0.318
	(0.599)	(0.945)	(0.604)	(0.717)	(0.758)	(0.612)	(0.417)	(0.587)	(0.439)	(0.702)
Importer Gini	-0.075	-0.044	0.072	0.004	0.049	-0.030	-0.041	-0.031	-0.103	0.055
	(0.096)	(0.136)	(0.098)	(0.133)	(0.122)	(0.126)	(0.077)	(0.112)	(0.083)	(0.133)
Exporter Gini	0.072	-0.013	-0.106	-0.136	0.127	0.071	0.068	-0.073	0.163*	-0.097
	(0.094)	(0.141)	(0.103)	(0.128)	(0.125)	(0.121)	(0.080)	(0.109)	(0.086)	(0.132)
Land Locked	-0.196	-0.114	0.910***	0.096	-2.420***	-0.795**	0.060	0.068	-0.037	-0.132
	(0.276)	(0.289)	(0.189)	(0.739)	(0.564)	(0.344)	(0.235)	(0.206)	(0.194)	(0.224)
Common Colonizer	0.834***	0.058	0.208	0.475**	2.180***	1.293***	1.166***	0.814***	1.236***	0.065
	(0.211)	(0.199)	(0.164)	(0.217)	(0.415)	(0.178)	(0.146)	(0.177)	(0.142)	(0.198)
Colony	0.532***	0.245**	0.341***	0.913***	0.780***	0.818***	0.181***	0.595***	0.601***	0.277**
	(0.076)	(0.123)	(0.070)	(0.090)	(0.080)	(0.080)	(0.054)	(0.075)	(0.054)	(0.088)
Common Language	0.123**	0.385***	0.442***	-0.089	0.137**	-0.023	0.054	-0.130**	-0.176***	-0.233**
	(0.049)	(0.077)	(0.047)	(0.062)	(0.054)	(0.055)	(0.037)	(0.051)	(0.034)	(0.055)

*, **, *** respectively show significance at 90, 95 and 95 percent levels. Robust standard errors are given in parentheses.

Table 5: 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	Meat
Distance	-0.780***	-1.622***	-1.647***	-1.774***	-0.998***	-1.520***	-0.797***	-0.338***	-1.345***	0.038
	(0.022)	(0.135)	(0.029)	(0.042)	(0.040)	(0.037)	(0.018)	(0.020)	(0.025)	(0.025)
Common Border	0.644***	1.023***	1.319***	0.908***	0.871***	1.170***	0.615***	0.697***	0.920***	0.127
	(0.069)	(0.091)	(0.069)	(0.083)	(0.068)	(0.073)	(0.050)	(0.074)	(0.047)	(0.080)
PTA	0.302***	1.334***	2.053***	1.534***	1.567***	0.421***	0.619***	0.770***	1.186***	-0.329**
	(0.079)	(0.105)	(0.074)	(0.098)	(0.083)	(0.079)	(0.053)	(0.083)	(0.060)	(0.112)
Importer Income	0.851*	-2.462***	0.735	2.548***	-2.024***	5.407***	2.030***	-1.609***	1.801***	0.433
	(0.451)	(0.678)	(0.585)	(0.812)	(0.630)	(0.783)	(0.374)	(0.468)	(0.439)	(0.569)
Exporter Income	3.766***	2.245**	2.776***	1.898**	-0.130	-0.187	0.145	-1.862***	-0.866**	-0.277
	(0.611)	(0.929)	(0.731)	(0.802)	(0.745)	(0.694)	(0.416)	(0.544)	(0.461)	(0.698)
Importer Per Capita Income	-0.164	3.599***	0.367	-0.963	2.933***	-4.008***	-1.024**	1.711***	-0.558	-0.649
	(0.478)	(0.697)	(0.613)	(0.844)	(0.662)	(0.826)	(0.391)	(0.489)	(0.459)	(0.609)
Exporter Per Capita Income	-3.735***	-1.965**	-2.866***	-2.301**	0.706	0.306	-0.172	2.039***	0.805*	0.372
	(0.643)	(0.992)	(0.766)	(0.839)	(0.789)	(0.723)	(0.443)	(0.575)	(0.485)	(0.732)
Importer Gini	-0.023	-0.070	0.078	0.190	0.117	0.009	-0.041	-0.016	-0.040	-0.012
	(0.102)	(0.149)	(0.126)	(0.155)	(0.128)	(0.151)	(0.082)	(0.109)	(0.093)	(0.138)
Exporter Gini	0.088	0.020	-0.227	-0.172	0.166	0.068	0.119	-0.066	0.318***	-0.093
	(0.099)	(0.156)	(0.134)	(0.150)	(0.131)	(0.144)	(0.085)	(0.107)	(0.096)	(0.138)
Land Locked	0.046	0.116	0.916***	-0.288	-2.613***	-0.410	0.281	0.245	0.177	-0.142
	(0.276)	(0.289)	(0.189)	(0.739)	(0.564)	(0.344)	(0.235)	(0.206)	(0.194)	(0.224)
Common Colonizer	0.934***	0.190	0.251	0.779***	2.359***	1.582***	1.453***	0.890***	1.440***	0.324
	(0.211)	(0.199)	(0.164)	(0.217)	(0.415)	(0.178)	(0.146)	(0.177)	(0.142)	(0.198)
Colony	0.696***	0.458***	0.454***	1.199***	0.956***	1.032***	0.306***	0.763***	0.803***	0.179**
	(0.076)	(0.123)	(0.070)	(0.090)	(0.080)	(0.080)	(0.054)	(0.075)	(0.054)	(0.088)
Common Language	0.193***	0.563***	0.636***	0.005	0.336***	-0.007	0.233***	-0.054	-0.062*	-0.218**
	(0.049)	(0.077)	(0.047)	(0.062)	(0.054)	(0.055)	(0.037)	(0.051)	(0.034)	(0.055)

*, **, *** respectively show significance at 90, 95 and 95 percent levels. Robust standard errors are given in parentheses.

Table 6: 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 & mate**	**Vegetables**	**Meat**
Distance	0.022	-1.066***	-0.842***	-0.602***	-0.723***	-0.765***	-0.650***	-0.425***	-0.826***	0.022
	(0.024)	(0.026)	(0.021)	(0.032)	(0.023)	(0.027)	(0.015)	(0.020)	(0.015)	(0.024)
Common Border	0.124	0.799***	0.530***	0.675***	0.787***	0.610***	0.632***	0.763***	0.627***	0.124
	(0.079)	(0.070)	(0.067)	(0.083)	(0.065)	(0.068)	(0.045)	(0.068)	(0.047)	(0.079)
PTA	-0.359**	0.976***	1.560***	1.152***	1.127***	0.550***	0.800***	0.813***	0.874***	-0.359**
	(0.111)	(0.099)	(0.070)	(0.098)	(0.080)	(0.074)	(0.049)	(0.079)	(0.058)	(0.111)
Importer Income	0.718	-2.305***	-0.115	-1.542**	-2.304***	0.538	0.267	0.988**	0.248	0.718
	(0.543)	(0.588)	(0.438)	(0.773)	(0.605)	(0.633)	(0.357)	(0.458)	(0.381)	(0.543)
Exporter Income	-0.288	1.810**	0.303	-2.635***	-0.248	-1.038*	-0.473	-1.517**	-0.967**	-0.288
	(0.671)	(0.806)	(0.556)	(0.738)	(0.724)	(0.606)	(0.410)	(0.548)	(0.417)	(0.671)
Importer Per Capita Income	-0.871	3.121***	0.576	1.627**	2.913***	0.169	0.46	-0.694	0.474	-0.871
	(0.583)	(0.623)	(0.460)	(0.806)	(0.637)	(0.662)	(0.377)	(0.478)	(0.402)	(0.583)
Exporter Per Capita Income	0.323	-1.477*	-0.168	2.689***	0.708	1.338**	0.522	1.617**	1.048**	0.323
	(0.704)	(0.854)	(0.584)	(0.775)	(0.766)	(0.634)	(0.435)	(0.575)	(0.440)	(0.704)
Importer Gini	0.053	-0.043	0.073	-0.236	0.041	-0.079	-0.045	0.03	-0.104	0.053
	(0.133)	(0.136)	(0.101)	(0.146)	(0.122)	(0.128)	(0.080)	(0.111)	(0.083)	(0.133)
Exporter Gini	-0.099	-0.013	0.007	-0.09	0.116	0.049	0.065	-0.035	0.147*	-0.099
	(0.133)	(0.142)	(0.106)	(0.143)	(0.127)	(0.126)	(0.083)	(0.109)	(0.085)	(0.133)
Land Locked	-0.086	-0.167	0.721***	0.644	-2.335***	-1.281***	0.243	-0.354**	-0.138	-0.086
	(0.221)	(0.291)	(0.173)	(0.739)	(0.258)	(0.329)	(0.210)	(0.175)	(0.195)	(0.221)
Common Colonizer	0.038	0.041	0.126	-0.056	2.020***	1.355***	1.325***	0.874***	1.217***	0.038
	(0.198)	(0.195)	(0.154)	(0.217)	(0.399)	(0.165)	(0.131)	(0.146)	(0.141)	(0.198)
Colony	0.278**	0.233**	0.284***	0.424***	0.793***	0.788***	0.173***	0.564***	0.607***	0.278**
	(0.087)	(0.084)	(0.069)	(0.090)	(0.074)	(0.074)	(0.049)	(0.068)	(0.051)	(0.087)
Common Language	-0.229***	0.369***	0.321***	-0.238***	0.104**	-0.004	0.113***	-0.164***	-0.185***	-0.229***
	(0.054)	(0.054)	(0.045)	(0.062)	(0.051)	(0.052)	(0.033)	(0.046)	(0.034)	(0.054)
Fixed Effects										
Importing Countries	41.5***	(19.78)***	(20.57)***	(44.67)***	(36.10)***	(21.75)***	(22.04)***	(15.23)***	(26.60)***	(41.53)***
Exporting Countries	7.2***	(65.95)***	(48.50)***	(70.92)***	(85.57)***	(133.59)***	(106.28)***	(96.61)***	(170.22)***	(7.16)***
Year	1.2	(3.85)***	(2.33)**	(2.09)**	(1.73)*	-1.15	(2.09)**	(4.23)****	(1.98)**	-1.2
Summary Statistics										
R-squared	0.087	0.42	0.455	0.426	0.411	0.37	0.382	0.32	0.376	0.087
#Observations	25634	23285	11566	16023	16340	20523	37784	25306	38513	25634

F-Statistics	21.5***	155.9***	78.8***	271.1***	93.0***	118.6***	224.6***	103.2***	197.1***	21.5***
RMSE	2.452	1.899	1.682	2.019	1.838	2.039	1.821	2.057	1.892	2.452

*, **, *** respectively show significance at 90, 95 and 95 percent levels. Robust standard errors are given in parentheses

[i] Hillberry (2002) illustrates selection bias by using the same data set as McCallum (1995) and shows that McCallum'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.

[ii] This is the same approach as used by Jayasinghe, et. al.(2010) and Disdier and Marette (2010) to account for potential selection bias.

[iii] A fifth option is suggested by Silva and Tenreyro (2006) who propose a Poisson Pseudo-maximum-likelihood (PPML) method to account for zero trade flows. However, Martin and Pham (2008) showed that the PPML estimator can be severely biased when the zero trade flows are frequent as is the case for agrifood products. See Jayasinghe, et. al.(2010) for a brief review of the technique.

[iv] The results in this paper differ from those in Haq, Meilke and Cranfield (2010) by the addition of one product sector (meat) and though the use of a different exclusion restriction. The results of the analysis are robust to these changes.

[v] Of course, simple spatial equilibrium models, long used in agricultural economics research,can also generate zero trade flows (Takayama and Judge 1971).