Advanced issues of Gravity Model

Dr. Selim Raihan
Executive Director, SANEM
Professor,
Department of Economics,
University of Dhaka

Outline

- Issue of zero trade flows
- Solution to the zero trade flows
- Endogeneity Issues
- Solution to the endogeneity problem
- Summary

Issue of zero trade flows

- There are evidences that show that, a large portion of international trade matrix consists zero trade.
- Haveman and Hummels (2004) find that about one-third of the bilateral trade matrix is missing.
- Helpman et al. (2008) find that about half of the country pairs in their sample do not trade with each other at all.
- Despite the existence of a large number of zero trade in the trade datasets, the gravity equation was almost invariably estimated using trade data sets converted into logs.

Issue of zero trade flows

- What happens if bilateral trade between countries is zero and we estimate them using conventional Log linear model?
- Taking logarithms of such observations effectively drops them from the sample as log(0) is undefined.
- Dropping these observations causes reduction of observations. Which also causes loss of information.
- Moreover, using truncated samples in logarithms may yield misleading results.

Issue of zero trade flows

- The problem of zero trade was rarely emphasized because (Martin and Pham, 2015):
  - trade theories were silent on their causes
  - lack of recognition of their frequency
  - due to the convenience of the log-linear estimation
- Recently there has been growing recognition among trade economists that zero trade flows do not occur randomly
- There are some recent developments to tackle the issue of zero trade.

Solution to the Zero trade problem

- The solution to the zero trade problem could be sought in three ways:

i) Simplest way:

- One can add 1 to all the observations of the total trade flow. Ln (0) is undefined but, ln (1) is zero. Hence, we may be able to recur some of the potential information.
- However, a Tobit estimation is suggested as the OLS would provide biased results (distribution is censored at zero)
- This approach does not have any theoretical basis. This is mostly used in the policy literature.

Solution to the Zero trade problem

ii) Pseudo Poisson Maximum Likelihood estimator:

- An alternative and theoretically more sound approach could be the use of Pseudo Poisson Maximum Likelihood (PPML) estimator.
- Santos Silva and Tenreyro (2006) suggest that, PPML can be a solution to the zero trade problem.

  They also highlight that, in the presence of heteroskedasticity, the PPML is a robust approach.
- This method can be applied on the levels of trade, thus estimating directly the non-linear form of the gravity model and avoiding dropping zero trade.
- The dependent variable is trade, not log (trade) whereas, the explanatory variables are still may be in log forms.

Solution to the Zero trade problem

- **Pseudo Poisson Maximum Likelihood estimator (Cont..):**
- This approach has been used in a number of estimation of gravity equations.

- **Estimating in Stata:**
- Estimating PPML is stata is fairly easy.

  Stata command: poisson *dep_var exp_var*, robust

Solution to the Zero trade problem

III) **The Heckman Sample Selection Model:**

- Conceptually it is a two step estimator.
  - In the first step: estimate a Probit (trade propensity) model in which the dependent variable is a dummy indicating whether or not a given observation is in the sample;
  - Then estimate the main model by OLS, including a measure of the probability of being in the sample, derived from the Probit estimates.

Estimation in Stata:

Command : heckman *dep_var explanatory_var,* select *(list of variables that go into the first stage equation)* robust

**** Note: All gravity variables should be present in both stages of the model.***

Issues of Endogeneity

- The problem of endogeneity occurs when an explanatory variable is correlated with the error term.
- **Endogeneity can arise from:**

I. **Measurement error:** If the variable in consideration is measured with an error and the pattern of error is systematic, the variable will be correlated with the error term.

  - Let, $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$

- Now, if the variable X_1 is measured with an error, say w_i, it will be a part of the error term. Therefore, $\mu_i = w_i + v_i$, where, v_i is the true error. Here, as $cov(X_{1i,} w_i) \neq 0$, $cov(X_{1i,} \mu_i) \neq 0$
- It will lead towards a biased result if w_i follows a common pattern. If w_i is purely random, it will not cause endogeneity problem.
- However, measurement error in the dependent variable does not cause endogeneity.

Issues of Endogeneity

II. Autocorrelation:

If the error are correlated serially, there will be endogeneity problem.

- Let, Y_i is a variable that depends on its own past:
- $$y_{it} = \beta_0 + \beta_1 y_{it-1} + \beta_1 X_{it} + \mu_{it}$$
- Now, as Y_{it-1} consists of μ_{it-1}, if the error terms ($\mu_{it,}$ μ_{it-1}) are correlated, then there will be the problem of endogeneity.
- An Example of it could be that, if the export of current year depends on the exports of the previous year. Now if the errors are correlated there, then there will be problem of endogeneity.

Issues of Endogeneity

III. Omitted variable bias:

- Endogeneity can arise due to the omitted variable bias.
- Let's suppose: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$
- However, if we omit X_{2i} and run the model as:

$$y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

- Then, X_{2i} is captured in the error term ε_i. i.e. $\varepsilon_i = \beta_2 X_{2i} + \mu_i$. Therefore, if X_{1i}, X_{2i} are correlated then $cov(X_{1i},X_{2i}) \neq 0$. It will imply that, $Cov(X_{1i}, \varepsilon_i) \neq 0$. Therefore, there will be bias in the result.

- **Example:** If in the gravity model estimation, we exclude any influential variable, that can cause possible bias in the estimation.

Issues of Endogeneity

IV. Simultaneity:

- If the variable considered in the model itself is a function of the dependent variable, then it may cause endogeneity bias.
- Let, $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$
- Where, X_{2i} is a function of y_i. Then, X_{2i} will be correlated with μ_i and there will be the problem of endogeneity.

- **Example:** In gravity model estimation, we often omit the tariffs in the regression equation. The reason behind it that, there could be 'reverse causality' problem. **Do tariffs reduce trade or does more trade induce tariff reductions?**

Solution to Endogeneity Problem

- A widely preferred solution is the use instrumental variable. In this case, the 2SLS (two stage least square) estimator is applied.

- **Instrumental Variable:**

Let’s consider the following equation:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \qquad (1)$$

where, $E(u) = 0$, $cov(X_1,u) = 0$ but, $cov(X_2,u) \neq 0$. The reason behind this could be any of the previously stated issue.

Here X_2 is endogenous.

Solution to Endogeneity Problem

To use the IV approach for the endogenous variable X_2, we need an observable variable z, which is not present in the equation (1) and it must satisfies two conditions:

i) The variable must be uncorrelated to the error term (u):

i.e. $cov(z, u) = 0$

ii) The variable must be strongly correlated with the endogenous variable:

i.e. $cov(z, X_2) \neq 0$

If the variable (z) satisfies both of these conditions, then it is said that z is an instrumental variable (IV) for X_2.

Solution to Endogeneity Problem

- **Approach:**

  In the first step we run the reduced form regression: i.e. we regress X_2 over other explanatory variables and z:

  $$X_2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 z + v$$

  In the second step we add $\alpha_2 z$ in our original regression equation in (1) in the place of X_2:

  $$Y = \theta_0 + \theta_1 X_1 + \theta_2 z + v$$

  where, $\theta_2 = \beta_2 \alpha_2$
- Now, here, z is uncorrelated with the error term v, hence, we will get an unbiased estimation of the coefficients.