Session 1: Export Diversification and Gravity

Ben Shepherd

Principal, Developing Trade Consultants Ltd.

ARTNeT Capacity Building Workshop for Trade Research: Gravity Modeling

Thursday, August 26, 2010

Outline

Introduction

- Thus far, we have worked with a log-linearized version of the gravity model.
- What happens if bilateral trade between i and j is in fact zero?
- Taking logarithms effectively drops such observations from the sample, because log(0) is undefined.
- What are the consequences of dropping zero observations? What can we do to avoid dropping them? What can we learn from them (export diversification)?

How Common are Zeros?

- There is a good amount of recent evidence to the effect that zeros are in fact surprisingly common in the bilateral international trade matrix.
- Haveman and Hummels (2004) find that nearly 1/3 of the bilateral trade matrix is empty.
- Helpman et al. (2008) find that about half of the country pairs in their 158 country sample do not trade with each other at all.
- As we drill down to ever finer levels of product disaggregation, we can expect the problem to become more and more serious.
- The corrollary is that export diversification or extensive margin trade growth is an important phenomenon.

Potential Consequences of Zeros

1. Dropping zeros means we are getting rid of potentially useful information. We might be able to learn something about why some countries trade in some products, but others do not.
2. By only using a portion of the available data, we might be producing biased estimates of the coefficients we are primarily interested in.
3. Returning to the non-linear gravity model–which is what attention to zeros can imply–might indicate some ways in which we can make empirical models fit the data better.

Overview

- The recent literature has paid a good deal of attention to the “zero problem”.
- Three main approaches:
  - Ad hoc solutions;
  - The Poisson model;
  - Heckman’s sample selection model.
- Estimating the AvW gravity model using nonlinear methods is also a possibility. But it is cumbersome, relatively labor intensive (you need to program it), and thus rarely used in practice.

Outline

An Ad Hoc Solution

- Log(0) is undefined, but log(0+0.0001) is not.
- For all but very small numbers, $log(x + 0.0001) \approx log(x)$.
- Adding a small, positive number to all trade flows can be a sensible place to start, to see if including or excluding zeros appears to make much of a difference empirically.
- It is commonly used in the policy literature, but has no theoretical basis, and is approximate at best.

Outline

The Poisson Model

- Santos Silva and Tenreyro (2007) propose an econometric solution to the zero problem.
- The Poisson model commonly used for count data can in fact be applied more generally to non-integer variables, and is equivalent to (weighted) non-linear least squares. The estimator is consistent under weak assumptions, and the data need not be distributed as Poisson.
- Their point is in fact a very general one: the econometric estimates of log-linearized models can be misleading due to a particular, and noxious, type of heteroskedasticity. But we will focus on the gravity implications only.

The Poisson Model

- Poisson models are very easy to estimate in Stata:
  - poisson trade ln_dist [etc.] dummies, robust cluster(dist)
  - xtpoisson trade ln_dist [etc.], fe
- There is no robust estimator for xtpoisson, so install the external package xtpqml: just use the fe option, and it will give you robust and clustered estimates automatically.

The Poisson Model
What Difference does it Make?

Estimator: Dependent variable:	OLS $\ln(T_{ij})$	OLS $\ln(1 + T_{ij})$	Tobit $\ln(a + T_{ij})$	NLS T_{ij}	PPML $T_{ij} > 0$	PPML T_{ij}
Log distance	−1.347**	−1.332**	−1.272**	−0.582**	−0.770**	−0.750**
	(0.031)	(0.036)	(0.029)	(0.088)	(0.042)	(0.041)
Contiguity dummy	0.174	−0.399*	−0.253	0.458**	0.352**	0.370**
	(0.130)	(0.189)	(0.135)	(0.121)	(0.090)	(0.091)
Common-language dummy	0.406**	0.550**	0.485**	0.926**	0.418**	0.383**
	(0.068)	(0.066)	(0.057)	(0.116)	(0.094)	(0.093)
Colonial-tie dummy	0.666**	0.693**	0.650**	−0.736**	0.038	0.079
	(0.070)	(0.067)	(0.059)	(0.178)	(0.134)	(0.134)
Free-trade agreement dummy	0.310**	0.174	0.137**	1.017**	0.374**	0.376**
	(0.098)	(0.138)	(0.098)	(0.170)	(0.076)	(0.077)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9613	18360	18360	18360	9613	18360
RESET test *p*-values	0.000	0.000	0.000	0.000	0.564	0.112

The Poisson Model

- Poisson enables us to estimate a gravity model which includes the zeros: the dependent variable is trade, not log(trade).
- The independent variables still enter in logs, and the coefficients can still be interpreted as elasticities.
- For some fairly abstract reasons of statistical theory, Poisson is actually a very good workhorse estimator for gravity even if zeros are not a problem in the data: the type of heteroskedasticity that “the log of gravity” deals with seems very common.

Outline

The Heckman Sample Selection Model

- Poisson implicitly assumes that there is nothing “special” about zeros: the problem is just to get them into the estimation sample.
- An alternative approach to the problem is in terms of sample selection:
    - A first set of covariates determine the probability that two countries engage in trade at all (i.e., that country-pair’s trade propensity)
    - A second set of covariates determine the intensity of bilateral trade, subject to the existence of a trade relationship.
- The original Heckman models deals with this in two steps; HMR modify his approach to include a third step.

Conceptualizing Sample Selection

- One way of thinking of sample selection is as an omitted variables problem.
- Take the basic log-linear gravity model:

$$\ln\left(X_{ij}\right) = b_0 + b_1 \ln\left(GDP_i\right) + b_2 \ln\left(GDP_j\right) + b_3 \ln\left(d_{ij}\right) + e_{ij}$$

- By dropping zeros from the sample, our dependent variable is no longer really bilateral trade, but bilateral trade contingent on a trading relationship existing.
- An important variable left out of the model is the probability of being included in the sample, i.e. having a non-zero trade flow.
- To the extent that the probability of selection is correlated with GDP or distance, then it has the potential to bias OLS estimates.

The Heckman Model

- Heckman proposed an elegant solution to this type of problem:

1. Estimate a Probit (trade propensity) model in which the dependent variable is a 1/0 indicator of whether or not a given observation is in the sample;
2. Then estimate the main model by OLS, including a measure of the probability of being in the sample, derived from the Probit estimates.

- Conceptually, this is a two-step estimator. But in practice, it is common to estimate both stages simultaneously via maximum likelihood.

The Heckman Model

Trade Propensity versus Trade Flows

Variables	m_{ij}	(Probit) T_{ij}
Distance	-1.176^{**}	-0.263^{**}
	(0.031)	(0.012)
Land border	0.458^{**}	-0.148^{**}
	(0.147)	(0.047)
Island	-0.391^{**}	-0.136^{**}
	(0.121)	(0.032)
Landlock	-0.561^{**}	-0.072
	(0.188)	(0.045)
Legal	0.486^{**}	0.038^{**}
	(0.050)	(0.014)
Language	0.176^{**}	0.113^{**}
	(0.061)	(0.016)

The Heckman Model

The Combined Model

Variables	Benchmark	NLS	Firm heterogeneity	Heckman selection
Distance	-1.176^{**}	-0.798^{**}	-0.769^{**}	-1.214^{**}
	(0.031)	(0.039)	(0.038)	(0.031)
Land border	0.458^{**}	0.834^{**}	0.855^{**}	0.436^{**}
	(0.147)	(0.132)	(0.142)	(0.149)
Island	-0.391^{**}	-0.169	-0.164	-0.425^{**}
	(0.121)	(0.120)	(0.118)	(0.120)
Landlock	-0.561^{**}	-0.447^{**}	-0.433^{*}	-0.565^{**}
	(0.188)	(0.172)	(0.187)	(0.187)
Legal	0.486^{**}	0.387^{**}	0.381^{**}	0.488^{**}
	(0.050)	(0.048)	(0.049)	(0.050)
Language	0.176^{**}	0.023	0.023	0.223^{**}
	(0.061)	(0.062)	(0.060)	(0.061)

Estimation in Stata

- The Heckman model is easy to estimate in Stata:
  - heckman ln_trade ln_dist [etc.] dummies, select(ln_dist [etc.] dummies) robust cluster(dist)
  - The select option should include the list of variables that go into the first stage (trade propensity) equation.
- Use STATA's "canned" Heckman routines to get correct second stage standard errors. Estimating the two stages separately and doing some algebra results in correct coefficient estimates, but standard errors that are too small.

Estimation in Stata

- Include all gravity variables in BOTH stages of the model.
- It is greatly preferable to include at least one variable that enters only the selection (probit) equation, not the outcome (gravity) equation: i.e., it affects trade propensity, but not the amount of trade.
- Fixed costs are a good candidate. HMR (2008) use the "Doing Business" costs of domestic market entry.

Estimation in Stata

- It is easy to estimate simpler trade propensity models in Stata too:
    - probit trade_dum ln_dist [etc.] dummies, robust cluster(dist)
    - logit trade_dum ln_dist [etc.] dummies, robust cluster(dist)
    - xtlogit trade_dum ln_dist [etc.] dummies, fe
- For abstract statistical reasons, do not rely just on probit fixed effect results–they may be misleading. Try xtlogit results too, as they do not suffer from the same problem.

Summary

- Zero trade flows are surprisingly common.
- The proportion of zeros in the trade matrix increases in line with:
    - Country diversity and disaggregation;
    - Sectoral disaggregation.
- The presence of zeros tends to bias OLS gravity model estimates.
- It also potentially throws away some useful and interesting information.

Summary

- The literature is currently undecided as to the “best” way to deal with zeros.
- Options include:
  - Ad hoc solutions (add a small positive number);
  - The Poisson estimator;
  - Some version of the Heckman sample selection model.
- In practice, it can be a good idea to try all of these.
- In some cases, unfortunately, they can give quite different results.