5th Week, EOD

BE 887 Econometric Exercise

September 19, 2018

Instructions: For this exercise you will replicate some of the main findings in Helpman, Melitz and Rubinstein (2008) using the data available on Marc Meltiz website. In addition, you will get experience implementation some alternative econometric specifications of the gravity equation. Specific steps and details appear below. You might need to read some other papers and websites to fully understand and implement the methodology. You should turn into to the me following.

1. A STATA do file that take that runs all the regressions. Please comment this file liberally so that I can read it.

2. A log file containing the output requested. If you are so inclined you can also turn in a spreadsheet with tables, but a log file is sufficient as long you comment you do-file well enough for me to follow.

3. Answers to the discussion points below where requested.

The objective of this assigment is to give you some experience with gravity equation estimation routines. In addition to using the HMR selection model and control function methods, you will use poisson quasi-maximum likelihood, and old tricks like adding a +1 to any zero trade flows.

1 HMR replication

Download the data from Marc Melitz's website here

`http://scholar.harvard.edu/files/melitz/files/gravity_data_share.zip`This is a set of almost ready to use STATA files. The main one you want to start with is data1980s_share.dta However, you will need to read some details in the HMR paper and make a few transformation of variables and define several sets of dummy variables.

1. Replicate Table 1, column 1 for 1986. Be careful defining landlocked and island as in the appendix. These are dummy variables equal to 1 if BOTH the importer and exporter are not islands or not landlocked. how to dyads sc's?

2. Replicate Table 1, column 2 for 1986. This is the probit regression. You'll need to define an indicator variable for whether a country pair has positive trade. Note that HMR report marginal effects of controls at the means of controls. After running the probit you'll need to use the command

   ```
   margins, dydx('controls') atmeans
   ```

   where 'controls' is a local variable containing all your controls.

3. Replicate Table 1, column 3 for the full 1980s panel. This is the panel trade flow regression only. Pay attention to the notes on the set of fixed effects and standard error assumptions. You do not need to replicate the probit for the panel. It is computationally intensive to run that regression and compute marginal effects. You may not be able to replicate the standard errors depending due to differences in degrees of freedom adjustments across STATA commands, but they should be close.

4. Merge in the regulation data from "data_regulation_share.dta", which you should have downloaded already. Note that you need to merge it twice: once on the importer id and and again on the exporter id. This measure is undefined for many observations and will cut the sample in half.

Replicate the probit regression in Table 2 column 1. You'll need to construct the regulation variables yourself from ind_cost and ind_prodcdays. I was unable to get the total observation count exactly the same, but it should be close to the paper.

5. Next, using the same data as above, use the STATA command

   heckman

   to construct the two step Heckman selection correctoin. Make sure you exclude the reguation variables from the 2nd stage. This will automatically run a selection model and construct the inverse mills ratio of the second state and adjust the standard errors according for the generated regressor.

6. Then using the probit first stage from part 4, replicate Table 2, column 4. You will need to generate the index function of the probit $\hat{z}_{ij}^*$ outcome variable. This is the index function of the unit normal distribution for the probit equation. Be careful you predict the correct outcome. Construct the 3 polynomial regressors in $\hat{z}_{ij}^*$ in powers of 1, 2, and 3. You will also need to construct the inverse mills ratio. Look up the formula and construct it yourself using $\hat{z}_{ij}^*$ and STATA's built in Normal CDF and PDF functions, or, use "predict" with the correct options after the heckman command to recover the mills ratio (use "help heckman" for details). You do not need to correct the standard errors for the generated regressors, just make sure you can run the second stage with correct regressors. There are some poorly described additional trimming and transformation employed by HMR on the polynomial measures and predicted probabilities. You might not get the exact same results, but they should be reasonably close.

2 Alternative Specifications and Tests

1. Santos-Silva and Tenreyro (REStat, 2006) propose a different estimator that can handle zero trade flow in their paper "The Log of Gravity". Transform your log trade dependent variable into levels, replacing missing trade flows with zero. Use the poisson quasi-maximum likelihood estimator suggested by Santos-Silva and Tenreyro on the same set of controls and the same 1986 cross-section data in HMR from Table 2, column 1 that you used in part 4. That is, include all the zeros that you used in the first stage probit and the regulation controls. Discuss why the coefficient on distance or any geographic coefficients is different. Do you prefer one estimator over the other?

2. Alternatively, use the poor man's zero correction of adding 1 to all trade flows. So zero trade flows become log(value+1)=log(1)=0. Regress this on the same set of controls and sample as the poisson regressoin you just ran . What happens to your coefficient on distance? Provide an explanation for why the coefficient is different from OLS on the log(value), the HMR methodology, and the poisson approach.

3. Return to the full panel dataset using log trade values. In most of the regressions so far the FTA dummy has been positive and significant, but it varies substantially dependon on the set of fixed effects, the time period, and whether the HMR selection correction is employed. Regress log trade flows on the FTA dummy only using the sets of fixed effects below. Describe why you think the coefficient on the FTA dummy varies substantiall across specification. Think about what the omitted variables are and when and why countries join FTAs. Think about where the identifying variation for the FTA dummy comes from as the fixed effects change.

   (a) importer + exporter+year, as in HMR

(b) importer*year + exporter*year

(c) importer*year + exporter*year + importer*exporter