

HOMWORK 4

ECN 607 – Econometrics II: Applied Microeconometrics
Prof. Bruce Blonigen

Due Date: Thursday, Feb. 19

1) Maximum Likelihood Estimation.

A) Do parts a), b), and c) from questions 4 in chapter 17 of Greene's text (5th Edition), pp. 522-3.

For the rest of this homework, we will continue to use the USFDIMaster database and be estimating modified versions of the Carr, Markusen, and Maskus (CMM) model of FDI determinants using data on U.S. outbound FDI activities. We'll begin with the following log-log base specification from homework 3:

```
reg lrpos lsumgdp lgdpdifsq lskdiff lht_tcost lhm_tcost lht_beri ldistance if outbound==0;
```

2) Dynamic Panel Estimator.

A) Read Arellano and Bond's 1991 article in *Review of Economic Studies*, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." (Read it at least well enough to understand the discussion in the Stata manual on the command "xtabond" and be able to answer the following questions). Two copies have been put on the left table in the TERF room.

B) Estimate a one-way Arellano and Bond (AB) with two lags of the dependent variable. Before doing so it will be important that you "clean" the current data by 1) dropping observations with missing values for any of the variables, 2) dropping any observations of inbound activity, and 3) using the "tsset" command to designate your cross-sectional and time identifiers. Also, use the "set matsize" command to allow 250 columns. Interpret the overidentifying restrictions test. What does it say about your estimates and empirical model? Interpret the autocorrelation tests. What does it say about your estimates and empirical model? (The AB article will be helpful with this last question)

C) Compare your AB estimator results to a standard First Difference (FD) estimator on the same sample. Use the "if e(sample)" condition on the end of your regression command to ensure you are using the same observations. (You also have to make sure you do not run any other regression in between, since Stata only saves which sample you used in your last regression). How much does it matter to model the dynamics for this sample?

D) Run a two-way AB model with two lags of the dependent variable. Which estimator do you prefer. The one-way AB estimates in B) or the AB estimates here in D)? Explain.

3) Binary Choice Models.

A) We wish to estimate a model to explain a bilateral tax treaty formation between the U.S. and other countries. Thus, the model is:

$$y_i^* = x_i \beta + \varepsilon_i,$$

where $y_i = 1$ if $y_i^* > 0$ and $y_i = 0$ if $y_i^* \leq 0$

In the USFDIMaster database, the variable “deff” is an indicator variable of whether a bilateral tax treaty is effectively in place between the U.S. and the partner country. Give an interpretation of y_i^* .

B) Resample the USFDIMaster data, use the standard construction of variables used in homework 1 to get **levels** of the variables (not logs). Use the following variables as a regressor matrix (including a constant): sumgdp, gdpdifsq, ht_beri, distance, and a one-period lag of rpos. You will have to generate the latter variable (by cross-sectional unit, of course!). Estimate a **linear probability** model and explain the marginal (economic) effects of sumgdp and lagged rpos.

C) Generate maximum likelihood **probit** and **logit** estimates. Provide marginal (economic) effects of sumgdp and lagged rpos. Compare probit and logit estimates to the linear probability estimates. Do you get qualitatively similar answers?