

HOMEWORK 3

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

Due Date: Thursday, Feb. 5

For 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 1:

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

which yields the following results:

Source	SS	df	MS	
-----+-----				Number of obs = 892
Model	1447.03983	7	206.719976	F(7, 884) = 74.85
Residual	2441.38939	884	2.7617527	Prob > F = 0.0000
-----+-----				R-squared = 0.3721
Total	3888.42922	891	4.3641181	Adj R-squared = 0.3672
				Root MSE = 1.6619

lrpos	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
lsumgdp	3.729948	.8214597	4.54	0.000	2.117709	5.342187
lgdpdifsq	-.669287	.3119789	-2.15	0.032	-1.281593	-.0569813
lskdiff	-.0924125	.1253975	-0.74	0.461	-.3385241	.1536991
lht_tcost	-.0681695	.0361272	-1.89	0.059	-.1390745	.0027356
lhm_tcost	1.062487	3.970349	0.27	0.789	-6.729923	8.854897
lht_beri	-2.67275	.2543142	-10.51	0.000	-3.17188	-2.17362

ldistance	-1.065663	.1085663	-9.82	0.000	-1.278741	-.8525855
_cons	1.42479	28.47055	0.05	0.960	-54.45296	57.30254

1) Endogeneity Concerns.

A) The variable `lht_beri` is a measure of the “costs of FDI” and is an index of how a country in a given year is perceived in terms of its business climate, legal protection of foreign assets, etc. Provide an explanation for why this variable may be endogeneous.

B) The USFDIMaster dataset contains variables connected with bilateral tax treaties between countries that determine tax rates on repatriated income from foreign affiliates. One of these is “ageeff”, the number of years that a country has had a bilateral tax treaty with the U.S., where “0” indicates that no treaty exists with the U.S. Explain how well you think this variable meets the criteria for an instrumental variable. Run 2SLS using this variable as an instrument and calculate

a Hausman test for endogeneity. Does the 2SLS coefficient estimate for “lht_beri” differ from the OLS estimate of this coefficient in the way you would expect given your explanation for the endogeneity of “lht_beri”? Explain. Is this a “weak” instrument? Explain.

C) An alternative instrument for “lht_beri” is “cnum”, the number assigned to each country in the dataset largely based on alphabetical order. Explain how well you think this variable meets the criteria for an instrumental variable. Run 2SLS using this variable as an instrument and calculate a Hausman test for endogeneity. Which instrument do you prefer and why? Is this a “weak” instrument? Explain.

2) Fixed-Effects Estimator

A) The outbound sample we are running has both cross-sectional variation and time variation. “Cnum” indexes countries, the cross-sectional units, while “year” indexes time periods. Why is it an unbalanced sample? Explain why it may be important to control for unobserved cross-sectional heterogeneity in the context of this FDI model.

B) Can one control for such unobserved cross-sectional heterogeneity using Stata’s “areg” command? If so, what type of estimator would it be?

C) Find and use a Stata command to generate the fixed effects estimator. Are the fixed-effects statistically significant? Compare the fixed-effects coefficients with the OLS pooled coefficients. Why does Stata drop “ldistance” out of the fixed-effects specification?

D) Use the “by”, “collapse” and “merge” commands to transform your data so that you can run the fixed effects estimator with a simple OLS command. Run that OLS on transformed data and verify that you get the same coefficient vector.

3) First-Difference Estimator

A) Transform your data into first-difference form. First differences can be generated in Stata using the command “gen xdif=x[_n]-x[_n-1]” after you have properly sorted the data by N then by T. Small “n” in the command is really referring to time periods. Be very careful to do this by cross-sectional unit, using the “by” command. Run the first-difference estimator and compare results with the fixed-effects estimator. Are they the same? Why or why not? Explain.

B) Now estimate the first-differenced data using a fixed-effects estimator. Explain how would one interpret the estimated fixed-effects in this context? Are the fixed-effects statistically significant?

4) Random-Effects Estimator

A) Estimate the outbound FDI data specification with a random-effects estimator. Compare the coefficient estimates with the fixed-effects estimators’ coefficients. Why isn’t “ldistance” dropped from this specification?

B) Run a Hausman test of the fixed-effects estimator to the random-effects estimator. Given the test, which estimates do you prefer and why?