University of Oregon
Department of Economics

# HOMEWORK 1

ECN 607 – Econometrics II: Applied Microeconometrics
Prof. Bruce Blonigen **Due Date: Thursday, Jan. 15**

Read the following articles "Estimating the Knowledge-Capital Model of the Multinational Enterprise," by Carr, Markusen and Maskus in the June 2001 issue of the *American Economic Review* and the comment on this paper by Blonigen, Davies, and Head in the June 2003 issue of the *American Economic Review.* For fun, you should read the reply to the comment by Carr, Markusen and Maskus in the same June 2003 issue.

Download the "USFDIMaster" Stata data set from the course website and make note of the name of the path and directory in which you place the dataset on your harddrive. Likewise, download the "Basic FDI.do" file. In the "Basic FDI.do" file, make sure to appropriately change the path name in the command file that pulls the USFDIMaster file into Stata. Run the "Basic FDI.do" file in Stata and verify that you get the following regression results:

```
Source |          SS       df      MS                Number of obs =    2382
-------------+------------------------------         F(  9,  2372) =  113.78
   Model | 3.2476e+11     9   3.6085e+10             Prob > F      =  0.0000
Residual | 7.5224e+11  2372   317133838              R-squared     =  0.3015
-------------+------------------------------         Adj R-squared =  0.2989
   Total | 1.0770e+12   2381   452333559             Root MSE      =   17808
```

| rpos \| | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| sumgdp \| | 8.303777 | .3567082 | 23.28 | 0.000 | 7.604285 | 9.003269 |
| gdpdifsq \| | -.0003635 | .0000384 | -9.47 | 0.000 | -.0004388 | -.0002883 |
| skdiff \| | 409.591 | 153.98 | 2.66 | 0.008 | 107.6416 | 711.5404 |
| skdgdpd \| | -.4244529 | .0479491 | -8.85 | 0.000 | -.5184793 | -.3304265 |
| ht_tcost \| | -101.8304 | 17.69576 | -5.75 | 0.000 | -136.5312 | -67.12965 |
| hm_tcost \| | -22.80593 | 10.89434 | -2.09 | 0.036 | -44.16935 | -1.442515 |
| ht_beri \| | -238.5731 | 63.57979 | -3.75 | 0.000 | -363.2508 | -113.8953 |
| htskd2 \| | 1.5626 | .3746681 | 4.17 | 0.000 | .8278894 | 2.297311 |
| distance \| | -1.634716 | .1589369 | -10.29 | 0.000 | -1.946386 | -1.323046 |
| _cons \| | 504.5629 | 2995.648 | 0.17 | 0.866 | -5369.796 | 6378.922 |

Beautiful results, right? Well, let's see about that. This homework involves completing the following programming tasks and answering the accompanying questions. At the beginning of class of the due date, I need your final Stata "do" and "log" file that completes all tasks on this homework, as well as a separate document that provides answers to questions posed to you on this homework. In your Stata "do" file you must put descriptive lines that demarcate the programming for each individual question on the homework.

**1) Economic Significance and Interpretation of Base Results.**

**A)** This is an expanded data set to that employed by CMM's 2001 AER article. In addition, the variable "ht_beri" is an alternative measure of investment cost in the host country and the current data set uses average years of schooling to measure skill differences, rather than labor force data

used by CMM. Since this is a different data set, discuss whether there are any sign or statistical significance differences with the CMM results.
**B)** What's the marginal effect of the variable sumgdp. In particular, what's the change in the dependent variable if this variable goes up by $1billion? Is this an economically significant effect? Explain why or why not.
**C)** What's the marginal effect of skdiff on the dependent variable? In particular, what's the change in the dependent variable when the parent country's population over 25 has an extra 2 years of schooling on the host country? Evaluate this at the means of the data (Summarize the data using the following command: "summ if e(sample)"). Is this an economically significant effect? Explain why or why not. What bothers you about this calculation considering the 2003 AER comment by Blonigen, Davies and Head on the 2001 AER paper by Carr, Markusen and Maskus.

**2) Addressing the Blonigen, Davies, Head (BDH) Critique.**

**A)** In these data, an easy way to incorporate the BDH critique (and the part of the CMM reply to the BDH comment) is to separately estimate the model for <u>inbound</u> FDI data and <u>outbound</u> FDI data. The data set has an indicator variable for inbound and outbound observations. Summarize the SKDIFF variable for inbound and outbound data. The following command does it for inbound: summ skdiff if inbound==1 & e(sample). This command tells STATA to give descriptive statistics for the skdiff variable if the inbound variable takes the value of "1" AND if the observation is in the sample of the previous regression just run. These conditional statements are very useful in STATA!! You will get to know them! From the descriptive statistics, explain why splitting the inbound and outbound data are sufficient to address the BDH comment.
**B)** Run the CMM regression specification separately for inbound and outbound data (use the conditional statement "if inbound==1", etc., here). Explain how results for these two samples differ from the results of the pooled sample above and why.

**3) Regression Diagnostics**

<u>**For the rest of this homework, we will use the outbound regression only.**</u>
**A)** Generate a residual versus fitted plot for this estimation. Conduct a Ramsey RESET test. What does this diagnostic say about your regression?
**B)** Generate variance inflation factors. Are you worried about multicollinearity with one or more variables? Besides the VIFs, do you see other signs of multicollinearity mentioned by Greene, p. 57.
**C)** Compare the fitted (or predicted) value for U.S. outbound FDI to Brazil in 1990 to the actual value. Do likewise for U.S. outbound FDI to France in 1990. Is there any pattern in the residuals for rich countries (ht_rgdpch>10000) than for poor countries (ht_rgdpch<10000)? Provide relevant analysis and discuss.
**D)** Perform one other diagnostic test that the STATA manual discusses under "Regression Diagnostics", explain the test and what the results suggest.

**4) Further Specification Tests**

**A)** An alternative model for explaining FDI patterns is a gravity-type equation, where FDI is assumed to be a function of the parent and host countries GDPs and the distance between them. Assume that SUMGDP, GDPDIFSQ, and DISTANCE characterize the regressor matrix for an alternative gravity model of FDI. Test the CMM model versus this gravity model and explain the test, its appropriateness for the problem at hand, and the results.

**B)** Suppose that the constant, SUMGDP, GDPDIFSQ, SKDGDPD, HTSKD2, HT_BERI, and DISTANCE are "free" or "important" variables, SKDIFF is your focus variable and HM_TCOST and HT_TCOST variables are "doubtful". Perform an Extreme Bounds Analysis (EBA). Then perform another EBA assuming that the constant is "doubtful" and not a "free" variable.

**5) Functional Form Change**

**A)** One obvious feature of these data is the obvious skewness of a number of the variable. (For example, for many the standard deviation is larger than the mean of the variable) As an alternative, let's consider taking the natural log of the data. One problem is that logging the data will mean that interaction terms, SKDGDPD and HTSKD2 will be perfectly multicollinear with other variables. <u>Show why this will be true</u>. Thus, we will drop these two variables from the specification. The second issue is that RPOS (our dependent variable) and HT_TCOST have observations where they take nonpositive values and obviously cannot be dropped. Rather than dropping these observations, create new variables that truncate these nonpositive values to 0.1. To do this, create a new variable identical to these variables and then read up on the STATA "replace" function. Now take the natural log (ln) of the data and run this CMM log model.
**B)** Show that the marginal effects of the coefficients in a log-log model are elasticities.
**C)** Rather than estimate a log CMM model, estimate a Box-Cox transformation using what STATA terms its "lambda" model. Again, do not estimate with SKDGDPD and HTSKD2 in the regressor matrix. Interpret the results in terms of whether a linear or log-log model should be used.
**D)** Calculate and display the marginal effects of the coefficient estimates from the Box-Cox model at the means of the data. How do the magnitudes compare to those found in the log-log specification.
**E)** Do the same diagnostic tests for the log-log specification as you did in parts **3A)** and **3B)** above. Compare to the results with the linear model.

**6) General Question**

**A)** Explain two other potential problems with the data that you think may need to be addressed besides those addressed in this homework.