University of Florida - Spring 2018
ECO 7707: International Economic Relations
Gunnar Heins (gheins@ufl.edu)

Problem Set 1

Due: Monday, January 29 before the beginning of class.

Instructions

This homework can be submitted in groups of 2. Please send me your code and document it well. Also write down and describe your results carefully. While R is recommended for the computational parts, feel free to use Matlab or Stata instead as well as other functions as the one stated for R.

1 Data Collection and Gravity Equation

This homework asks you to collect trade data in R and estimate the gravity equation. In particular it asks you to evaluate the impact of the Euro on trade flows using both an OLS as well as a diff-in-diff approach. Finally, it asks you to discuss potential threats to identification.

1. Use the function get.Comtrade() which we discussed in class to download all trade flows between the 20 largest countries for the year 2014.[1] Collect only information on the total trade (i.e. use c = "TOTAL") but get information on both imports and exports.

2. Match each country pair to its bilateral distance using the file "distance.RData". In this file "o" and "d" denote the origin and destination country, respectively. Distance is reported in km between a set of the biggest cities of the two countries.[2]

3. Match each importing and exporting country to its respective country GDP in 2014 using the file "gdp_2014.RData"

[1] The largest countries and their comtrade codes are: United States (842), China (156), Japan (392), Germany (276), United Kingdom (826), France (251), Brazil (76), Italy (381), Russia (643), India (699), Canada (124), Australia (36), Korea (410), Spain (724), Mexico (484), Turkey (792), Indonesia (360), Netherlands (528), Saudi Arabia (682), Switzerland (757).

[2] Therefore the distance for pairs where both origin and destination country are the same is not 0 and for example bigger for geographically large countries.

4. Using ggplot2, plot the natural log of trade flows on the y-axis and ln(distance) on the x-axis. Is the plot consistent with the gravity equation? Repeat the same exercise but use the natural log of the exporting country's GDP on the x-axis instead.

5. Run the following gravity equation using the lm() function in R:

$$ln(x_{i,j}) = \beta_0 + \beta_1 \cdot ln(y_i) + \beta_2 \cdot ln(y_j) + \beta_3 \cdot ln(dist_{ij}) + u_{ij} \quad (1)$$

   where $x_{i,j}$ denotes a trade flow from country i to country j, y_i and y_j the countries' GDPs and $dist_{ij}$ the bilateral distance between the 2. Further, assume that u_{ij} is a mean-zero error term and uncorrelated with the regressors.

6. What do you find? Are your results comparable to the ones by Frankel et al (1995) that we covered in class?

A common argument for the Euro is that it facilitates and spurs trade. Suppose you want to look into this claim and decide to augment the above gravity equation to do so.

7. Create a dummy that is one if both countries use the Euro today and 0 otherwise.

8. Reestimate the gravity equation but also include the Euro dummy, i.e. estimate

$$ln(x_{i,j}) = \beta_0 + \beta_1 \cdot ln(y_i) + \beta_2 \cdot ln(y_j) + \beta_3 \cdot ln(dist_{ij}) + \beta_4 \cdot \mathbb{I}\{Euro\} + u_{ij} \quad (2)$$

   using OLS as before. $\mathbb{I}\{Euro\}$ is equal to 1 if both i and j use the Euro and 0 otherwise. What do you find regarding the coefficient on this dummy variable? How does the R^2 compare to the one for regression (1)?

9. Do you think your estimate of the coefficient β_4 reflects the causal impact of the Euro on trade? Why or why not? Can you think of reasons why the error term u_{ij} might be correlated with the Euro dummy?

One concern of the above equation is that e.g. historical ties between EU countries or the membership in the same customs union affect both trade directly as well as the decision to enter the Eurozone. Since historical ties are for example difficult to measure, they will typically be part of the error term u_{ij} and OLS would give us a biased estimate of β_4. Therefore, instead of the above gravity equation, you are considering to compare trade before and after the introduction of the Euro using a diff-in-diff regression.

10. Collect the same data on trade flows as before but now for the year 1995. Match this data to country GDP in 1995 using the file "gdp_1995.RData"

11. Add the 1995 data to the previous one for 2014. You should end up with 3 new columns: (1) Trade flows for 1995, (2) GDP of country i in 1995 and (3) GDP of country j in 1995.

12. In order to derive the diff-in-diff specification, first write down equation (2) for both years, explicitly keeping track of the respective period

$$ln(x_{i,j}^{(14)}) = \beta_0 + \beta_1 \cdot ln(y_i^{(14)}) + \beta_2 \cdot ln(y_j^{(14)}) + \beta_3 \cdot ln(dist_{ij}) + \beta_4 \cdot \mathbb{I}\{Euro^{(14)}\} + u_{ij}^{(14)}$$
$$ln(x_{i,j}^{(95)}) = \beta_0 + \beta_1 \cdot ln(y_i^{(95)}) + \beta_2 \cdot ln(y_j^{(95)}) + \beta_3 \cdot ln(dist_{ij}) + \beta_4 \cdot \mathbb{I}\{Euro^{(95)}\} + u_{ij}^{(95)}.$$

The superscript numbers in brackets represent the respective years, i.e. 1995 and 2014. Notice that the error term is also allowed to differ by year, i.e. also unobservables that determine trade flows might change over time.

Subtracting the second from the first line gives us a desired diff-in-diff version of the gravity equation:

$$\begin{aligned} ln(x_{i,j}^{(14)}) - ln(x_{i,j}^{(95)}) = \beta_1 \cdot [ln(y_i^{(14)}) - ln(y_i^{(95)})] + \beta_2 \cdot [ln(y_j^{(14)}) - ln(y_j^{(95)})] \\ + \beta_4 \cdot [\mathbb{I}\{Euro^{(14)}\} - \mathbb{I}\{Euro^{(95)}\}] + [u_{ij}^{(14)} - u_{ij}^{(95)}] \end{aligned}$$

which can be written more compactly as

$$\Delta ln(x_{i,j})) = \beta_1 \cdot \Delta ln(y_i) + \beta_2 \cdot \Delta ln(y_j) + \beta_4 \Delta \mathbb{I}\{Euro\} + \Delta u_{ij}$$

Estimate this equation again with the lm() function. What do you find? How do your results differ from those found in (8)?

13. Suppose the error term u_{ij} depends partially on factors that are time-invariant v_{ij}, e.g. the language spoken in a country, and partially on factors that might vary over time $v_{ij}^{(t)}$, e.g. tariffs and shipping costs:

$$u_{ij} = v_{ij} + v_{ij}^{(t)} \tag{3}$$

Suppose you are worried that sharing a common language affects both trade flows directly as well the decision to enter the Eurozone and hence $Corr(v_{ij}, \mathbb{I}\{Euro\}) \neq 0$. Would this be a problem in regression (2)? And in the diff-in-diff specification?

14. Would you now interpret the estimate for β_4 as the causal impact of the Euro on trade. Why or why not? Can you think of a case in which Δu_{ij} might still be correlated with

$\Delta\mathbb{I}\{Euro\}$?