Introduction to R Programming
Slide Set 8: Application 3 - International Trade

Maria Ptashkina

Barcelona GSE ITFD

September 2021

Learning Objectives

- Content
  - Building up a gravity dataset
  - Running regressions: OLS, fixed effects model, random effects model
  - Regressions with binary dependent variable
- R-specific learning objectives
  - Cleaning data
  - Merging data
  - Running regressions
  - Exporting regression output

Study Design

- In this session we will work extensively with real world messy data
- We will create our own dataset to estimate a gravity model, by merging data on
  - Bilateral imports
  - WTO accession
  - GDP data
  - Gravity variables (distance, etc.)
  - Religion

- We will estimate three panel data models (simple linear model, fixed effects model, and random effects model) to assess the correlation between the volume of trade flows and gravity variables

- Finally, we will estimate the probability of concluding a free trade agreement using two models with binary dependent variable

Task 1: Import, Inspect, Clean and Merge

- Import trade flows from file 'tradeflows.csv', inspect the data, declare a data frame called `tf`
- Import WTO accession data from file 'joinwto.txt', inspect the data, declare a data frame callled `jw`
- Since we will combine many datasets, in real data there are often discrepancies which you would have to fix (but first find!)
- In the `jw` dataframe replace `BEL` or `LUX` with `BLX`; and `COD` with `ZAR`
- Drop duplicates
- Order the dataset by country

Task 1: Import, Inspect, Clean and Merge (cont.)

- Import GDP from file 'GDP.csv', inspect the data, declare a data frame called `gdp`
- To correct for Belgium and Luxembourg values, we need to sum the GDPs of two countries: organize the data in such a way that instead of two rows of there is one corresponding to the sum of GDPs of the two countries (there are many ways to do that, I am showing to you probably the least elegant one)
- Import dataset with gravity variables from file 'dist_cepii224.dta'
- Rename variables `country` to `exporter`, `partner` to `importer`, `repnum` to `exporternum`, `partnum` to `importernum`
- Replace `BEL` or `LUX` with `BLX`

- Check the number of unique reporters and unique exporters and the size of the trade matrix that should result from those numbers
- Since there are some observations that repeat, collapse the variables into a right size matrix, and fix the names of the columns
- Create all possible country-pairs-year combinations

- Reshape the GDP dataset to a long format and merge with the bilateral imports dataset. Note that the GDP dataset varies by a single country, and trade varies at the exporter-importer level. So first create two datasets in order to merge twice

- Repeat the same operation for the WTO accession data, and merge the dataset from the previous step with the WTO accession data

- Next, merge the resulting dataset with gravity dataset

Task 1: Import, Inspect, Clean and Merge (cont.)

- Import the dataset 'Religion.dta', inspect the data, declare a dataframe `rel`
- Merge the dataset you created before with the religion dataset
- Replace `NA` values in religion with 0
- Generate dummy variables for the WTO membership: `onein` equal 1 if one of the country pair is member of the WTO, 0 otherwise; `bothin` equal 1 if both countries is member of the WTO, 0 otherwise; `nonein` equal 1 if none of the country pair is member of the WTO, 0 otherwise (note that they are going to be perfectly collinear)
- Compute the log of the variables imports, GDPs and distance, and susbstitute minus infinity in imports with `NA`
- Generate panel id
- Create labels for all the variables
- Export dataset as `csv`

Task 2: Estimating the Gravity Model

- Subset the dataset for *computational issues*: choose the observations after year 2000, and with import flows larger than 139900
- WARNING! Never ever do an arbitrary reduction of your data like this!
- Run OLS estimation with country fixed effects, produce robust standard errors
- Run a fixed effects model, produce robust standard errors
- Run a random effects model, produce robust standard errors

Task 3: Exporting Regression Output

- Export the linear regression model output into a LaTex table using `stargazer()` function

Task 4: Binary Dependent Variable

- Import CEPII gravity dataset with data on bilateral preferential regional trade agreements (RTAs), declare a dataframe `ptas`
- Select variables you will need for the analysis: countries, year, contiguity, bilateral distance, common official language, common colony, common legal system, GDP per capita of origin and destination, WTO and EU membership, RTA, and trade values
- Compute the log of the variables imports, GDPs and distance, and susbstitute minus infinity in trade with `NA`
- Declare all the factor variables correctly
- Run a probit model of RTA on all the covariates
- Run a logit model of RTA on all the covariates
- Estimate and plot marginal effects

References and Resources

- Using R for Trade Policy Analysis ▸ Tutorial
- Trade Policy Analysis ▸ Data