Data analysis 3

9 March 2017

Term Project Assignment

This is an individual research project. The project has to investigate a relationship between two variables, with an aim to answer a business or a policy-relevant question. Most interesting analyses would try to estimate the effect of a policy (or a policy variable) on an important outcome, but investigating other associations is also fine if you can give sufficiently strong motivation.

Use a wide panel (that observe many cross-sectional units such as we did for health expenditure / CO_2 examples). Analyze features of the panel. Apply more than one panel method and argue for your choice. (Upon agreement, you may do a project with time series.)

Sources may include the World Development Indicators, Eurostat, OECD or any other source of data including scraped data.

Procedure

Topics need to be individual, i.e. no exact same topic may be presented by two students. This will be achieved by a first come, first serve (get approved) basis.

I set up a Google spreadsheet with and "approved?" column. Please describe the topic of your choice and the data you want to use, **by 17 February** the latest, for approval (the earlier the better). If you need consulting, let us know. Only approved topics may serve for the assignment.

The project results have to be written up in a paper of 5 pages at most (page limit will be strictly enforced). Turn in time for the paper is **Midnight Thursday 9 March**.

The project has to cover the following

Question. Explicit statement of the question. Why it is important (literature may help here but not compulsory). You may discuss the policy relevance of the question using your country as an example. (E.g., Should policy change in my country to achieve some outcome?) Eventually, you have to answer the question in an explicit way, and, if necessary, you have to qualify that answer.

Data. Exact source, sample selection, discarding outliers (if any), weighting or not weighting, logs or levels. Consider diagnostic tests if needed. Choice of variables (logs, binary, transformations to achieve stationarity etc.) Argue for your choices.

Identification strategy. The regression that identifies the effect you are after (or gets closest to that affect or the association you aim to measure). Argue for that regression and, if possible, show evidence to back up that argument.

Results. Show the main result and emphasize it. Interpret the parameter(s) of interest without further assumptions.

> Does it show the causal effect? Why or why not? If not, which way it is different from a causal effect?

> Robustness checks: if there are multiple "best" ways of estimating the effect tell about those and show the results of some of those.

> Interactions: your answer may be different for different countries or different times. You may show results showing such differences and discuss them.

Conclude for policy. If applicable, give policy recommendations to your own country based on the evidence you provided.

Paper

Your report should be of 5 pages maximum, including the most important figure and table(s). You may add an appendix with additional figures and tables. Include your code in another appendix.

Grading will be

- 0 to 20% points for thinking (question and answer)
  - why do the analysis, supporting evidence, what the results mean for the question
- 0 to 50% points for the analysis
  - Exploratory data analysis and data manipulation,
  - choice of statistical model and it details, and background analysis substantiating that choice.
  - robustness checks
- 0 to 30% for the quality of paper (documentation of data analysis steps, clarity of writing, structure, etc.)
- 0 to 20% points for something extra
- max is 100% even if sum of parts exceeds 100