DA4 Term project

2021

The 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. The analysis should focus on answering a causal question: how a causal variable affects an outcome.

- Sources may include the World Development Indicators, or any other source of data including webscraped data.

Setup

- Y is outcome
- X is causal variable
  - It should be related to an intervention or policy
- Z some confounder(s)

- You may do a difference-in-differences or a panel data project.

Difference-in-differences

- If using a two periods data for a diff-in-diffs

- Make sure there is a clear definition of
  - the intervention
  - the before and after period

- You will need at least several hundred observations on both ends

Panel data

- If using a panel data:
- Good panel: at least 100 obs, at least 10 years. More observations is always better.

- Minimum:
  - N should be at least 49
  - T should be at least 6

- Unbalanced panel OK, but number of observations>300 and must check the nature of unbalance

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.

- In the DA4 signup sheet has a term project sheet, with an "approved?" column. Please describe the topic of your choice and the data you want to use, **by 26 March** the latest, for approval (the earlier the better). We will offer consultations (on 25 March + email).

- Only approved topics may serve for the assignment.

Output

To submit

1. A paper in pdf! Your report should be of 5 pages maximum, including the most important figure and table(s).

2. Please submit a zipped folder with code and data you used OR: give a public repository (such as github) link in the paper

Structure 1/2

- Question and setup (40%)
  - 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. (10%)
  - Data. Exact source, sample selection, treating missing values, extreme valued observations (if any), weighting or not weighting, logs or levels. Feature engineering / choice of variables (logs, binary, transformations to achieve stationarity etc.) Argue for your choices. (20%)
  - Identification strategy. Decide upon your main regression that you hope would identify the effect you are after (or gets closest to that affect or the association you aim to measure). Explain what you are going to estimate and why. (10%)

Structure 2/2

- Results and their discussion (60%)
  - Present your results. Show the *main* result and emphasize it. Interpret the parameter(s) of interest of your main model. You may estimate a few versions of your model and argue for the choice of preferred specification. (20%)
  - Discussion - Does it show the causal effect? Why or why not? If not, which way it is different from a causal effect? (10%)
  - Robustness checks: if there are multiple good ways of estimating the effect tell about those and show the results of some of those. Could be related to decisions on the data or the statistical model. (10%)
  - Heterogeneity: your answer may be different for different countries or different times. You may show results showing such differences and discuss them. (for instance, by cutting the sample into two groups) (10%)
  - Conclude for policy. If applicable, give policy recommendations to your own country based on the evidence you provided. (10%)
- Something extra (e.g. nice graphics, great data collected, very good discussion). (10%)