

DA4 Term project

Procedure

- Topics need to be individual (or at pair), i.e. no exact same topic may be chosen twice. 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 for approval (the earlier the better). There will be a in-person consultation + email to TA, Danila.
- Only approved topics may serve for the assignment.

You may do it in a pair or alone

- You may do the project alone
- You may do it in a pair, but it will be harder

The assignment

- 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.

Project idea, data source

- Try find an interesting question and data source
- <https://gabors-data-analysis.com/data-source-ideas>

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.
 - You may do an event study project although it's difficult to find the right setup

Use panel data

- Start with a simple diff-in-diffs, as you define periods.
- Then do panel data methods and argue for choice

IF you do it in a pair.

- In addition to panel methods, you must also show results in an event study setup.
 - Change the data structure
 - Show event study as in class
 - Discuss possible problems due to heterogeneity of timing and effect

Output

To submit

1. A paper in pdf! Your report must be of 5 pages maximum (!), including the most important figure and table(s) excluding references
 1. Free to chose format
 2. No need for references (but you may have them)
 3. You may have an extra appendix if absolutely necessary.
2. It must be reproducible. 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 2-3 *main* results 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 based on the evidence you provided. (10%)
- Something extra (e.g. nice graphics, great data collected, very good discussion). (10%)

Q&A

1. *Can I use a report I wrote for another class?* No. You should do something new. But you may use data, ideas from other classes, projects.
2. *Is a zero result OK?* Yes! As long as you gave a good argument why x may cause y , that is fine.
 - Typically, we wanna see a cross sectional correlation though.

Q&A 2 panel data

- Great panel: $N = \text{thousands}$, $T = 8+$
- Good panel: at least 100 obs, at least 10 years.
 - More observations is always better.
- **Minimum: $N \geq 49$, $T \geq 6$**
- Unbalanced panel OK, but $N > 150$ and must check the nature of unbalance