Data Analytics II

Causal Econometrics

(MECON & MIQEF)

Johanna Kutz
Swiss Institute for Empirical Economic Research (SEW)
University of St. Gallen | Switzerland | Spring 2024

Self-study Kick-off

Organisation

- Self-studies count for **25% of the grade** and are completed **individually**
- The problem set and the dataset will be available on studynet by **March 1st**
- Support via:
  - Email (johanna.kutz@unisg.ch)
  - **(Optional)** Q&A session on **March 5th 18:15-20:00 -** Submit questions by **March 4th** via email
  - Q&A Session on **March 26th 18:15-20:00 -** Submit questions by **March 24th** via email

Submission deadline: 2nd Sunday after the break (**April 28th**) until midnight on StudyNet

A complete submission should contain…

1. A PDF with the documentation and results (details below)
2. One file with the main code
3. One file containing the functions

Learning goals

There are **two** main **learning goals** from the self-studies:

1. Improve your understanding of the **research designs** and **estimators** (you only really understood if you can code it)
2. Improve your **coding/googling skills**: It is probably the most important skill for Data Analysts to search and find online resources that help to achieve what they plan to do (highly individual, everybody might find other resources useful)

You have on purpose some degrees of freedom about what you implement and how you implement the details (like in a later job or research project)

Task

You will conduct an empirical case study

You get a *dataset and problem set*

Please take "your" dataset to work with. Datasets will be individualized (please do not copy results).

You have to follow the tasks on the problem set

Rules

1. Follow the advice on the Problem Set.
2. For some exercises, you have (on purpose) some leeway on the implementation of the estimators
   - Please state all decisions taken clearly!
3. Implementation must be done in **Python**
4. Implement the estimators from scratch (no library functions)
5. You may use libraries as inputs to your functions, e.g. Lasso implementation within a new Causal ML function

Projects that are identical will receive zero points

Rules

- *Implement the estimators from scratch (no library functions)*
- *You may use libraries as inputs to your functions, e.g. Lasso implementation within a new Causal ML function*

Example:

1. Calculate the mean difference in y for group A and B
   - Calculate the mean manually!
2. Run the following regression $y_i = \alpha + \beta x_i + \epsilon_i$
   - No need to calculate mean manually!
   - The estimator itself should be manually coded

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\,\bar{x}\,,$$

Documentation

Be precise! Not more than 6 pages

Structure of documentation:

1. Results: Compactly summarize your results in few tables or graphs and discuss your findings. Answer the questions on the Problem Set
2. Implementation: Describe how you implemented the estimators (if there is some flexibility)

Evaluation

Total of 25 points

3 points for the programming: Does the code run?, Is it well commented?, Is the coding efficient (**use of functions**)?, …

22 points for solving the Problem Set:

- Results
- Discussion and Interpretation of results
- Implementation