

Exercise Sheet for the Oral Exam in the Course “Causal Machine Learning”

Date: April 16, 2021, 14:00-18:30pm

Registration for oral exam (required): Doodle <https://doodle.com/poll/3pydzbh2ynq33xs5>

Location: Online (Zoom)

<https://unilu.zoom.us/j/91401321997?pwd=b3NON3JkMnhMalM4ZXhUZ3dxWUJldz09>

Meeting ID: 914 0132 1997

Passcode: 847885

Procedure of the oral exam: Each oral examination lasts a total of 20 minutes. Before the oral examination, each student individually conducts the empirical project described below and prepares a presentation of the project. During the oral examination, each student has 12 minutes to present the empirical project (the time limit is strictly enforced). Following the presentation, there will be an 8-minute general question and answer session.

Empirical project: Low voter turnout can be a problem for the legitimatisation political representation. There are attempts to increase voter turnout with so-called get-out-the-vote (GOTV) calls. A GOTV call includes information about the election date and the importance of voting. In 2002, an experiment was conducted in Michigan and Iowa to examine the effect of GOTV calls on voter turnout.¹. Voters were randomly divided into two groups. One group with 60'000 voters received a GOTV call, but not all of them were able to answer the call. The other group of almost 2 million voters was the control group, which did not receive a call.

You find in the Github folder the data set `mobilization.RData`, which contains the following variables:

¹Arceneaux, K., Gerber, A. S. and Green, D. P. 2006. *Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment* Political Analysis, Vol. 14, No. 1, pp. 37-62

- *vote02*: voted in 2002, 0 = no, 1 = yes
- *treat_real*: 1 = received a GOTV call, 0 = did not receive a GOTV call
- *contact*: 1 = answered a GOTV call, 0 = not reached by telephone (or not called)
- *persons*: number of persons in the household
- *newreg*: newly registered voter
- *age*: age (in years)
- *female*: 1 = women, 0 = men
- *vote00*: voted in 2000, 0 = no, 1 = yes
- *vote98*: voted in 1998, 0 = no, 1 = yes
- *state*: 0 = Michigan, 1 = Iowa
- *competiv*: 1 = competitive district, 2 = uncompetetive district
- *comp_mi*: = $\text{competiv} * (1 - \text{state})$
- *comp_ia*: = $\text{competiv} * \text{state}$

Each student can choose one of the following four projects:

1. Use the post-double-selection procedure to investigate the effect of answering a GOTV call on the probability to participate in the 2002 election.
2. Use the double/debiased machine learning approach to investigate the effect of answering a GOTV call on the probability to participate in the 2002 election.
3. Provide a machine-learning based effect heterogeneity analyses of receiving a GOTV call on the probability to participate in the 2002 election.
4. Use optimal policy learning tools to target the GOTV call.

No matter which project is chosen, the following questions should be considered (in addition to the substantive questions related to the specific project):

- Is the data clean? Are there any data issues?
- What is the identification strategy? Are the assumptions plausible?
- What are the pros and cons of the chosen estimation method?