Causal Inference for Policy Evaluation – Spring Semester 2025

Lab Assignment 1

Due date: Friday, 04.04.2025 (at 23:59)

Instructions

Please send your solutions to mario.bernasconi@unibas.ch. The solutions should include two files. One file with your code saved in .R format. A second file with written answers, including tables and graphs, should be sent as a .pdf file and is limited to a maximum of 2 pages (12 points, single spacing). PhD student groups are allowed one more page as they have one more question. Any text or output beyond that will not be considered.

Unconfoundedness (20 points for MSc, 25 for PhD)

Download the file `Assignment_1.Rdata` from ADAM and open it using `load(.)`. Make sure that the data frame `raw.data` and the vector `covs` are loaded into your global environment. The data is similar to that used in the first and second lab sessions. In particular:

- 'treat' is a dummy for participation in the program
- 'date_start' and 'date_end' denote the beginning and end dates of the unemployment spells
- People below age 18 and above or equal to 60 were dropped. Variables 'agegr_2', 'agegr_3' and 'agegr_4' are dummies for age groups [30,40), [40,50), and [50,60), respectively.
- 'sex_1' is a dummy for being female, 'marits_1' is a dummy for being married, 'swiss_0' is a dummy for being Swiss
- 'region_1', 'region_2', etc. are regional dummies
- 'educ_0', etc. are dummies for educational levels
- 'full_time_0' is a dummy for working full-time

- 'lastj_occpt_1', etc. are dummies for occupation types in the last job
- 'lastj_fct_0', etc. are dummies for the function in the last job
- 'quarter_1', etc. are dummies for quarter of the year

In line with the argumentation in Wunsch and Lechner (2008), we may anticipate heterogeneity in treatment effects across observable dimensions. In this exercise, we investigate the effect of participating in a training program on the monthly employment probability. We are particularly interested in whether the treatment effect varies by age.

Questions

1. Generate a dummy variable equal to 1 if people were below age 40 at the start of unemployment, or 0 if they were 40 or older. Compute the share of people below 40 by treatment status. Report the standardized bias and the difference in means along with a p-value. Based on these metrics, briefly comment on the balancedness of the treatment and control group with respect to age and whether we should account for age differences in our analysis. **(4 points)**

2. We want to estimate the ATE on the employment probability for each of the first 24 months after the program start using inverse probability weighting, separately for people below and above 40. **(10 points)**
   You should follow these steps:

   (a) Create 24 dummy variables indicating the employment situation in the first 24 months after the program starts.

   (b) Estimate the propensity scores for the two samples (below and above 40) using the covariates in 'covs', and then trim the samples appropriately.

   (c) Estimate the ATE separately for people above and below 40. Use 100 bootstrap replications to obtain standard errors and 95% confidence intervals. You can use the 'treatweight' command from the 'causalweight' package (note that your sample is already trimmed). Report the results.

3. Using the results from point 2, plot the monthly ATEs with a 95% confidence interval for each age group. Interpret the resulting figure. Would the interpretation change if you estimate the ATET? If so, how? **(6 points)**

4. **This question is for PhD students only.** Imagine jobseekers who participate in the online application program are more likely to get a job, but at the expense of other unemployed workers with whom they compete in the labor market.

   (a) Can we still compare program participants to non-participants to estimate the ATE(T), or would the resulting estimate be biased? In which direction? **(2 points)**

   (b) Which identifying assumption is violated in this case and why? **(2 points)**

   (c) Can you think of other sources of violation for the same assumption which would drive the bias in the opposite direction? **(1 point)**

Difference-in-differences (5 points)

Read the paper for the next lab session: Hjort and Poulsen (2019), you can skip Sections IV.E and IV.F, Section V, and Section VI. Then answer the following questions:

Questions

5. Describe how the authors define (i) the treatment and control groups, (ii) the pre-treatment and post-treatment periods. **(2 points)**

6. Equation (1) in the paper (page 1045) reports the main specification used to estimate the treatment effect using a DiD framework.

   (a) Why don't the author control for individual fixed effects? **(1 points)**

   (b) The authors control for country-specific time period (quarter or year) fixed effects, $\gamma_{c(i)t}$, to control for within-country-location-invariant changes in employment outcomes. It would seem more flexible to control for location-specific time period fixed effect, to control for within-location changes in employment outcomes. Is this a good idea? **(2 points)**