Causal Inference and Observational Studies

Mandatory Assignment 4, Advanced Applied Statistics

In the fourth mandatory assignment you will analyze and discuss causality in relation to observational data. More specifically, you will conduct an empirical analysis of Danish citizens' welfare preferences. You will have to pick one of the two studies, Slothuus (2010) *or* Petersen et al. (2014) (i.e. **not** both), and discuss design features and analyze the data using matching techniques. Both studies use publicly available data from the Danish National Election Survey (2005 or 2007) and both studies use a quasi-experimental research design to study how exogenous factors shape welfare preferences.

If you pick Slothuus (2010): Focus on the results reported in Model 1 in Table 2. The model is estimated as a logistic regression. You do not have to reproduce the exact parameter for the treatment variable! Use `dnes2005.csv`.

If you pick Petersen et al. (2014): Focus on the results reported in Table 1 in Study 2. The model is an OLS regression. You do not have to reproduce the exact parameter for the treatment variable! The `education` variable included in the data is not identical to the one used in the study. Use `dnes2007.csv`.

Variables to be used

You can find descriptions of the variables in the codebooks. The `welfare` variable is the outcome variable and `treatment` is the treatment variable. The other variables are covariates.

Variable	Slothuus (2010)	Petersen et al. (2014)
`welfare`	policy perception (`V207`)	welfare recipient attitudes (`V0227`, `V0237`)
`treatment`	treatment	treatment
`female`	gender (`V310`)	gender (`V0012`)
`age`	age in years (`V357`)	age in years (`V0013`)
`education`	education (`V338`)	education (`V0508`)
`income`	income (`V333`)	income (`V0427`)
`polaw`	political awareness (`V291-V307`)	-
`emp_priv`	-	private sector employment (`V0447`)
`emp_pub`	-	public sector employment (`V0447`)
`pubtransfer`	-	public transfer (`V0447`)
`preschool`	-	preschool children (`V0383`)
`ega`	-	egalitarianism (`V0238`, `V0221`, `V0233`)

Content and tasks

1. Introduce and describe the identification strategy used in the study and the assumptions required to make causal inferences. Focus in particular on why we can make the move from correlation to causality not only possible but also plausible in the study. (hint: describe why it is reasonable to assume that assignment to treatment is as-if random)
2. Reproduce the key finding, i.e. a statistically significant average treatment effect in the expected direction (this does not have to be the *exact* coefficient from the model).
3. Use nearest neighbor matching (or other matching procedures) to test the robustness of the estimate.
4. Make an overall conclusion with a discussion of the internal and external validity of the inferences.

Requirements and submission

A front page must be included containing a title, name of the course, exam number and total number of characters.

Furthermore, the assignment must contain two appendices: An appendix containing your do-file from Stata and/or R-script, and an appendix containing operational definitions + descriptive statistics on all variables.

The assignment must be no longer than 5 pages if written by one person, and no longer than 8 pages if written by two persons. A page has of 2.400 units including space between words. Furthermore, tables and figures are included in this but front page, bibliography and appendices are not.