Ec2147, Spring 2021

Professor Bryan Graham

Problem Set 1: part (b)

Due: March 30th, 2021

Problem sets are due by 5PM EST. You may work in groups (indeed I encourage you to all work together), but each student should turn in their own write-up.

This exercise utilizes the National Longitudinal Survey (NLS) extract from Card (1995). You can download this dataset from David's website here.

Using this dataset your goal is to estimate the earnings benefit of completing at least four years of college (16 or more years of schooling). The target population consists of individuals who have graduated from high school (12 or more years of schooling). Your instrumental variable should be one of the college proximity variables in the dataset. Your analysis should condition on family background characteristics (such that covariate adjustment is required).

1. Fit a textbook linear instrumental variables model as in Angrist and Pischke (2008). Under what conditions will the coefficient on the college variable estimate a LATE? Are these conditions plausible?

2. Implement one of the semiparametric procedures discussed in class. Discuss the assumptions underlying your analysis and discuss your results.

3. Implement a model based using the EM-Algorithm. Your may find the following Jupyter notebook helpful. You do not need to calculate standard errors.

References

Angrist, J. D. and Pischke, J.-S. (2008). *Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics: An empiricist's companion.* Princeton University Press, Princeton, NJ.

Card, D. (1995). *Aspects of Labor Market Behavior: Essays in Honour of John Vanderkamp*, chapter Using geographic proximity to estimate the return to schooling, pages 201 – 224. University of Toronto Press, Toronto.