Econ 312: Problem Set 4

Professor Magne Mogstad

Due Thursday, April 30 before the lecture

The following folder:

`https://dataverse.harvard.edu/file.xhtml?fileId=1232173&version=1.0`

contains data on class size, average math and verbal test scores for 2,019 5th grade classes in 1,002 public schools in Israel, as well as enrollment data for these schools and percentage disadvantaged pupils. (You need to use the file final5.dta and to understand what variables in the data file mean you need to look at tables in the paper and in the code that replicates them in the folder)

In Israel, schools face a rule which states that classes cannot be larger than 40 pupils. When enrollment is 41, schools are supposed to open a second classroom, and then open a third classroom at 81 pupils etc. This causes discontinuous drops in class size at multiples of 40.

1. Estimate the effect of class size on math scores using OLS without any controls, and then by adding the percentage of disadvantaged students in the class and enrollment as controls. Interpret your results.

Start by limiting the sample to schools with enrollment between 20 and 60 students. Generate a (predicted) large class dummy based on the first discontinuity at 40 students.

2. Use OLS to estimate the effect of being in a large class on math scores assuming that you have a sharp RDD around this discontinuity. Control for the percentage of disadvantaged students in the class and a linear trend in enrollment.

3. Use Local Linear Regression (use the command -lpoly- in Stata) to get a point estimate of the effect of being in a large class on math scores assuming you have a sharp RDD. Finally, use a nonparametric bootstrap to estimate the standard error on your RDD point estimate. Compare these results to the estimates you obtained with OLS.

4. Estimate the effect of class size on math scores using fuzzy RDD. Control for the percentage of disadvantaged students in the class and a linear trend in enrollment.

5. (*) Use -rdrobust- to estimate the effect of class size on math scores and compare your results.

Now use the complete sample, and define the following variable *predicted class size* = *enrollment* $/(int((enrollment - 1)/40) + 1)$

6. Plot average class size as a function of enrollment. Add predicted class size to the plot.

7. Estimate the effect of class size on math scores using IV.

8. If the RDD is valid, then the coefficient of interest should not change significantly if we include or exclude covariates. Check whether this is the case.

Explore the validity of the design and the robustness of the results above using the following checks:

9. Manipulation: Plot the distribution of the assignment variable.

10. Misspecification 1: Present a graph using binned local averages of class size and math score against enrollment. Use bins of width 20 and make sure that the bins do not cover the discontinuities. Can you see the discontinuity in class size and math scores?

11. Misspecification 2: Superimpose a linear and quadratic trend on the previous graph. Does the polynomial approximation capture the non-linearities well?

12. Misspecification 3: Explore the sensitivity of the results in 7) to 1) bandwidths (restrict the estimation sample to intervals around the discontinuities), and 2) how you control for enrollment.

13. Placebo check: Conduct the RD analysis where your outcome is percentage disadvantaged pupils.