Advanced Applied Econometrics
OECD, November 2010
Thijs van Rens

Regression Discontinuity Design

Exercise 1. Sharp and fuzzy RD The Stata dataset `maimonides.dta` contains data on class size, average math and verbal test scores for 2024 5th grade classes in 1003 public schools in Israel, as well as enrollment data for these schools. These are the data used in Angrist and Lavy (1999). The purpose of this exercise is the replicate some of the results in that paper.

1. Generate a table of summary statistics and compare to Table I in the paper. Notice that there are some (very small) differences between the dataset used here and the one used in the paper.

2. Plot average class size as a function of school enrollment. Verify you can replicate Figure Ia.

3. In this exercise we focus on match scores. Estimate the conditional correlation between class size and math scores, by running a OLS regression. Control variables include the percentage of disadvantaged students in the class and school enrollment. Verify you can replicate columns 4, 5 and 6 in Table II.

4. Estimate the effect of class size on math scores using the sharp RDD.

   (a) First, use only the first discontinuity in class size at enrollment of 41 students. Generate a dummy for large classes using this discontinuity. Then estimate the effect of being in a large class on math scores. Limit the sample to schools with enrollment between 20 and 60 students and control the percentage of disadvantaged students in the class and a linear trend in enrollment.

   (b) Now use all discontinuities by defining a variable for predicted class size, which is defined as

   $$f_{sc} = \frac{\text{enrollment}}{\text{int}(\frac{1}{40}(\text{enrollment} - 1)) + 1}$$

   Verify you can replicate column 6 in Table III.

5. Estimate the effect of class size on math scores using the fuzzy RDD.

   (a) First, use only the first discontinuity as in part 4a.

   (b) Now, use all discontinuities and verify you can replicate column 8 in Table IV.

6. 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 here.

Excercise 2. Checklist for implementation of RDD Using the same dataset as in the previous exercise, we now explore the validity of the design and the robustness of the results using Lee and Lemieux's (2010) checklist for implementation of RDD.

1. Plot the distribution of the assignment variables. Is there evidence for manipulation?
2. Present the main RD graph using binned local averages. Can you see the discontinuity?
3. Superimpose a linear and quadratic trend on the graph from point 2. Does the polynomial approximation capture the non-linearities well?
4. Explore the sensitivity of the main result to a range of bandwidths, and a range of orders of the polynomial trend.
5. Conduct a parallel RD analysis on of the baseline covariates (placebo experiment). Does the treatment have an effect on this variable? What does that mean?
6. Explore the sensitivity of the results to the inclusion or exclusion of covariates (see point 6 of the previous exercise).