ECON 626: Empirical Microeconomics

Regression Discontinuity Basics (real data)

Department of Economics
University of Maryland
Fall 2019

The data set `ozier_jhr_small.dta` contains data from Ozier (2018). The study examines the impact of secondary school (high school) on outcomes, exploiting a jump in the probability of completing secondary school at a particular test score. The data set contains six variables:

`test` - test score, centered so that scores of zero or higher experience the jump in secondary
`jump` - an indicator for having a centered score of zero or higher
`intTestJump` - the product (interaction) of the two variables above
`secondary` - an indicator for completing 12th grade (secondary school)
`rv` - Sum of nonverbal reasoning and vocabulary measures
`female` - an indicator for being female

In this activity, you will replicate parts of Table 2 and Table 3 from the published paper, checking the robustness of the results using up-to-date regression discontinuity packages in Stata.

1. Regress the indicator for secondary schooling (`secondary`) on the following variables: `jump`, `test`, and `intTestJump`, subject to the restriction that the test score (`test`) does not exceed 0.8 in absolute value, and clustering the standard errors by test score (`test`). Do your results agree with the first column of the published paper's Table 2?

2. Now use either the commands `rdrobust` or `rd` with just the arguments `secondary` and `test`, in that order. How do the results change?

3. Use the command `rdplot` with just the arguments `secondary` and `test`, in that order. Is the graph easily interpretable?

4. To get a set of results that agree, only use robust (not clustered) standard errors in the explicit regression form; alternatively, use the `rd` command again but with options `k(rect)` and `bw(0.8)`, overriding the default triangular kernel and Imbens-Kalyanaraman optimal bandwidth.

5. Next, estimate the reduced-form impact of crossing the discontinuity on the combined reasoning and vocabulary score (`rv`) rather than schooling. Start with the same bandwidth and specification as in the first part of this activity, then try the commands `rdrobust` or `rd`.

6. Now use instrumental variables (2SLS) to estimate the effect of secondary school on the combined reasoning and vocabulary score. That is, the outcome should be `rv`; the excluded instrument for `secondary` should be `jump`; `test` and `intTestJump` should also appear on the right; optionally, include `female` as an additional control. Again cluster the standard errors by test score (`test`). Do your results agree with the first column of the published paper's Table 3?

7. Now use the command `rd` with just the arguments `rv`, `secondary`, and `test`, in that order. How do the results change?