Fiscal Policy and Inequality

21. Regression Discontinuity Designs

Elliott Ash

Monday, 26 November 2018

Regression Discontinuity Design (RDD)

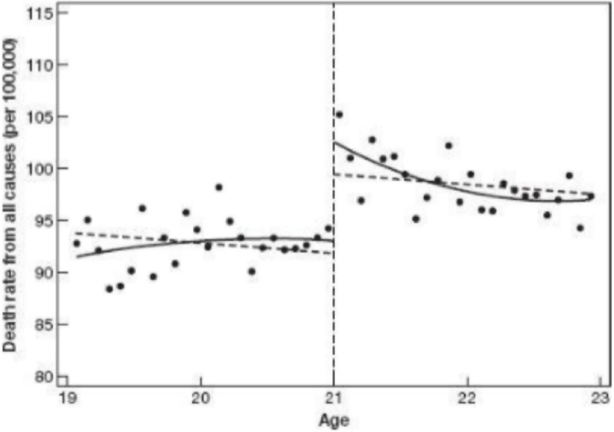
- Often rules are used to assign individuals to “treatments" which can be exploited for estimating causal effects.
  - in particular: threshold rules that are based on some ex-ante variable
- Example “running variables”:
  - Score in entry exams
  - Income for subsidy eligibility
  - Project quality score for public R&D subsidies
  - Age limit for alcohol consumption
  - Votes in an election
- The idea in RDD is to *exploit randomness around this threshold.*

Example: Effect of Minimum Legal Drinking Age on Death Rates

Carpenter and Dobkin (2009)

- outcome variable y_i : death rate
- treatment D_i : legal drinking status
- running variable x_i : age
- cutoff: at age 21, minors can suddenly drink legally

RDD Viz: Drinking Age and Death Rate



Running Variable

- The ex-ante variable is called the **running (forcing, assignment) variable**.
  - Selected threshold of the running variable divides individuals into "treated" and "not treated"
- Examples:
  - Score in entry exams
  - Income for subsidy eligibility
  - Project quality score for public R&D subsidies
  - Age limit for alcohol consumption
  - Votes in an election

RDD Conditions

- Two conditions for credible and precise RDD estimates:
  1. Variation in the treatment status near the threshold is as good as random
     1.1 How likely is this? Were there a way to anticipate the threshold and to manipulate the running variable?
  2. Sufficient number of observations around the threshold

RDD Estimation

- OLS regression:

$$Y_i = \alpha + \rho \mathbb{I}[x_i > c] + f(x_i)'\beta + \epsilon_i$$

- $f(x_i)$ includes polynomials in the forcing variable
  - generally linear or quadratic
  - also interact with being above or below the cutoff

Local Linear Regression

$$Y_i = \alpha + \rho \mathbb{I}[x_i > c] + \epsilon_i$$

- Limit sample to a given window (bandwidth) of width h around the cutoff point
- How to choose the bandwidth?
  - Trade-off: the closer you get the better it is for identification, but the less data you have...
- Use existent statistical algorithms for selecting the “optimal bandwidth" (e.g.: Imbens-Kalyanaraman 2011, Calonico, Cattaneo and Titiunik 2014).
  - Explore the robustness of results to different bandwidths

Testing the validity of RDD

- RD Design can be invalid if individuals can precisely manipulate the assignment variable x_i in order to get (or to avoid) treatment.
- Testing for validity:
  1. Density of the running variable should be continuous (McCrary test)
  2. Predetermined characteristics should have the same distribution just above and just below the cut off

Sharp vs. Fuzzy RD

- **Sharp RD:** treatment status (D_i) is a **deterministic** and **discontinuous** function of the running (assignment, forcing) variable (x_i):
  - $D_i = 1$ if $x_i > c$,
  - $D_i = 0$ i f$x_i \leq c$
- **Fuzzy RD:** being above threshold increases **probability** of receiving treatment, rather than deterministically changing treatment.
  - In this case, use RD as first stage in two-stage least squares estimation:

$$S_i = \alpha + \gamma \mathbb{I}[x_i > c] + \eta_i$$

$$Y_i = \alpha + \rho S_i + \epsilon_i$$

RDD in Python

- For sharp RD: estimate OLS (**linearmodels.PanelOLS**)
  - subsample around cutoff, and/or with polynomials in the running variable.
- For fuzzy RD: estimate 2SLS (**linearmodels.IV2SLS**)
  - can subsample around cutoff
  - instrument is a dummy variable for being above cutoff
  - endogenous variable is whether treatment is actually assigned.
  - include polynomials in running variable as covarates.

Electoral RD: Causal Effect of Incumbency

Lee, David (2008), 'Randomized experiments from non-random selection in U.S. House elections', Journal of Econometrics.

- Does a Democratic candidate for a seat in the U.S. House of Representatives have an advantage if his party won the seat last time?
- Exploits the fact that (previous) election winner is determined by majority rule in vote share:
  - forcing variable is margin of victory (difference between Democratic and Republican votes shares).
- Because D_i is a deterministic function of x_i , there should be no confounding factors.
  - in particular, no electoral fraud around the cutoff.

Viz: Electoral RDD

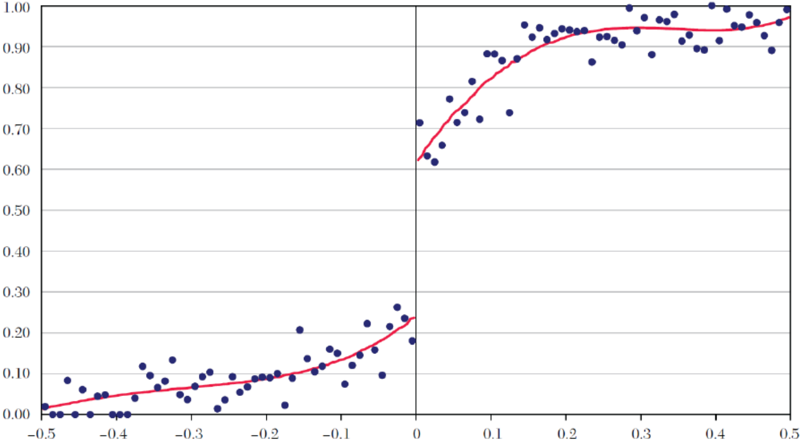


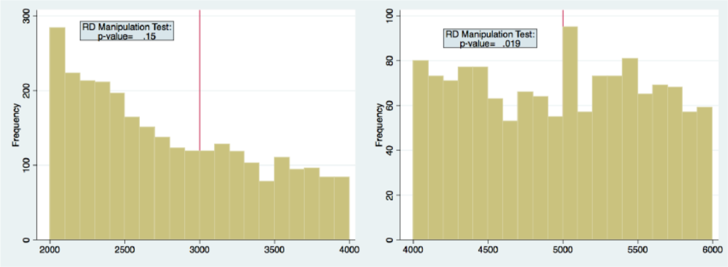
Figure 10. Winning the Next Election, Bandwidth of 0.01 (100 bins)

Electoral RDD: Notes

- Incumbency raises party re-election probabilities by 40 p.p.
- Checks:
  - Any discontinuities in predetermined covariates at the cut-off?
  - Is there bunching near the cutoff?
  - Are results robust to different bandwidths or functional form assumptions?

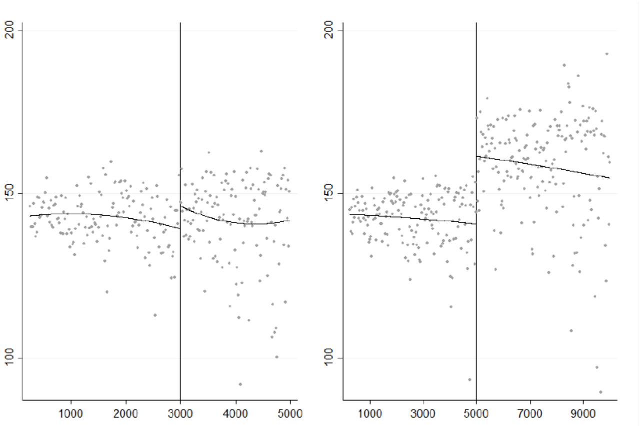
Manipulation Test: Density Around Cutoff

Bagues and Campa (2017): Histograms of Population Around Population Thresholds



Manipulation Test: Effect on Past Covariates

Bagues and Campa (2017): Federal Transfers Per Capita



RDD: Recap

- Useful method to analyze the impact of treatment when the assignment varies discontinuously due to some rules!
  - (test score, electoral results, income threshold, etc.)
- Graphical analysis is key, and can be very convincing
- Only feasible when the number of observations around the threshold is sufficiently large
- Condition for consistency: no precise manipulation at the threshold
  - may not hold if agents anticipate the threshold
- Make sure there are no other treatments at the same threshold
- RDD estimates treatment only for individuals around the threshold – effect could be different for others.