Advanced Microeconometrics 6SSPP393

Topic 5: Regression Discontinuity Design

Yonatan Berman

Semester 2, 2023/24

Causal inference

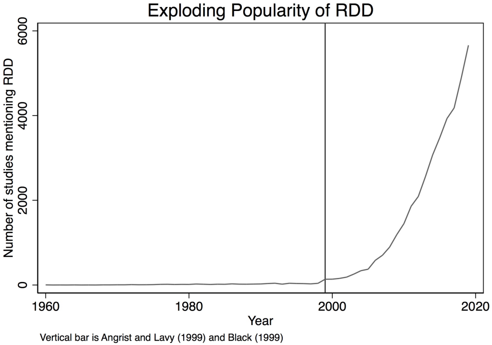
- In many cases we aim to attach a causal interpretation to the results of an empirical analysis
- Causal interpretation means that it captures the cause of some intervention
- The "golden standard" is a randomized controlled trial
- In most realistic cases it is impossible, so quasi-experimental designs are needed
- Difference-in-differences can be used if we observe units (or groups of units) over time
- But what can be done in the absence of time? How can we overcome unobserved heterogeneity? How can we identify causal relationship?

Regression discontinuity

- The idea is to use cases where there are thresholds that could be rather arbitrary
- Then, from one side of the threshold units receive a treatment, and from the other side no
- So, very close to the threshold from either side, the units are in principle very similar, to the extent that the assignment of treatment is approximately random

Regression discontinuity

- Donald Campbell, educational psychologist, invented regression discontinuity design but then it went dormant for decades
- Angrist and Lavy (1999) and Black (1999) independently rediscover it



Regression discontinuity – example 1

- Suppose we want to estimate the effect of attending a university with higher admission requirements (and therefore potentially "better" in various ways) on future earnings

- We could have used a simple regression such as:

$$y_i = \alpha + \beta x_i + u_i$$

  with y_i being earnings (say at the age of 35), and x_i being a binary variable that is 1 if attended the "better" university" and 0 otherwise

- Why is that a bad design?

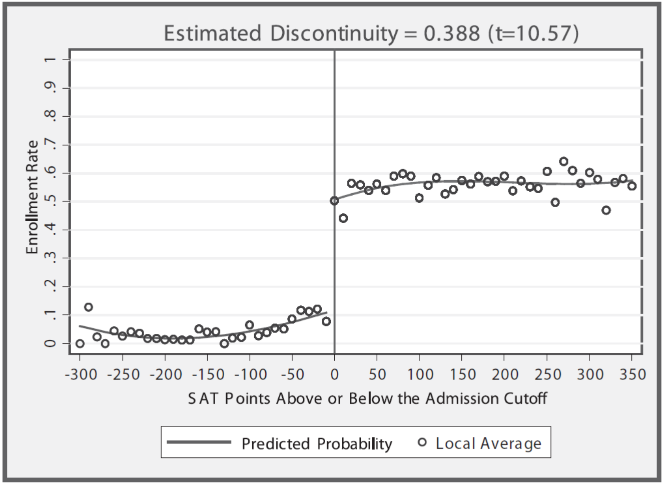
Regression discontinuity – example 1

- Unlike previous cases, it is not useful to use time, even if we could, to improve the design
- The reason is that students earnings before they attend a university are mostly 0, and if not, they typically unrelated to future earnings
- DiD design, for example, is unhelpful
- Because selection is so important in this example there needs to be another solution

Regression discontinuity – example 1

- Now let us assume that the higher admission criterion for the better university is represented by some threshold to a test score
- This creates discontinuity
- Above the threshold, admission to the better university is much more likely than below it, but that makes sense anyhow
- Yet, very close to the threshold from below and above, there is a jump in this probability
- This was used in Hoekstra (2009)

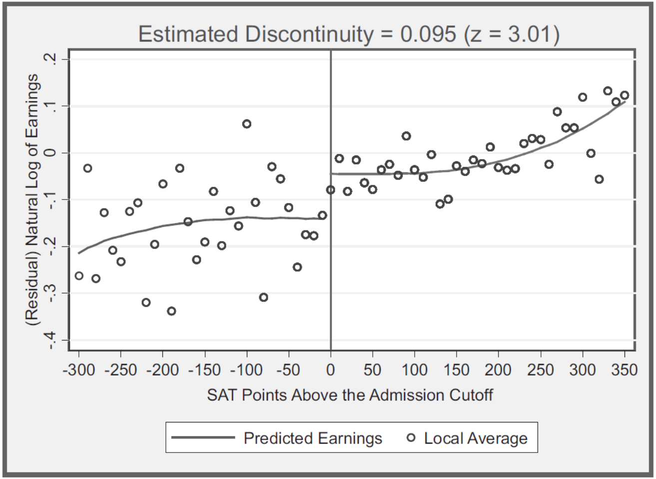
Regression discontinuity – example 1



Regression discontinuity – example 1

- Suppose the threshold SAT score was 1250
- A student with 1240 had a lower chance of getting in than a student with 1250
- Are these two students really so different from one another? If we had hundreds of students who made 1240 and hundreds more who made 1250 – are those two groups similar to one another on observable and unobservable characteristics?
- If the university is arbitrarily picking a reasonable cutoff, there might be reasons to believe they are also picking a cutoff where the natural ability of students jumps at that exact spot?

Regression discontinuity – example 1



Understanding RDD in more detail

- Goal: estimate some causal effect of a treatment on some outcome
- Problem: Selection bias (i.e., $E[Y^0|D=1] \neq E[Y^0|D=0]$)
- RDD basic idea: if treatment assignment occurs abruptly when some underlying variable X (the "running variable") passes a cutoff c_0, then we can use that arbitrary rule to estimate the causal effect even of a self-selected treatment

Arbitrary Thresholds and Treatment Assignment

- Firms, schools, and government agencies assign "things" based on arbitrary thresholds of continuous variables
- Consequently, probabilities of treatment will "jump" when that running variable exceeds a known threshold
- Examples of arbitrary thresholds:
  - Academic test scores: scholarships or prizes, higher education admission, certificates of merit
  - Poverty scores: (proxy-)means-tested anti-poverty programs (generally: any program targeting that features rounding or cutoffs)
  - Land area: fertilizer program or debt relief initiative for owners of plots below a certain area
  - Date: age cutoffs for pensions; dates of birth for starting school with different cohorts; date of loan to determine eligibility for debt relief
  - Elections: fraction that voted for a candidate of a particular party

Examples of Arbitrary Treatment Thresholds

- Think of these in light of a treatment where $E[Y^0|D=1] \neq E[Y^0|D=0]$:
  - Google Maps rounded a continuous score of ratings to generate stars
  - US targeted air strikes in Vietnam using rounded risk scores
  - Universal healthcare after age 65
  - When a newborn's birthweight is below 1500 grams, it gets intensive medical care

Types of RD Designs

- There are traditionally two kinds of RD designs:
  - Sharp RDD: Treatment is a deterministic function of running variable, X
  - Fuzzy RDD: Discontinuous "jump" in the *probability* of treatment when $X > c_0$ – cutoff is used as an instrumental variable for treatment

- Fuzzy is a type of Instrumental Variables strategy and requires explicit IV estimators like 2SLS – *we will focus on sharp RDD*

Deterministic Treatment Assignment in Sharp RDD

- In Sharp RDD, treatment status is a deterministic and discontinuous function of a covariate, X_i:

$$T_i = \begin{cases} 1 & \text{if } X_i \geq c_0 \\ 0 & \text{if } X_i < c_0 \end{cases}$$

- where c_0 is a known threshold or cutoff. In other words, if you know the value of X_i for a unit i, you know treatment assignment for unit i with certainty.

Definition of Treatment Effect

- The treatment effect δ, is the discontinuity in the conditional expectation function:

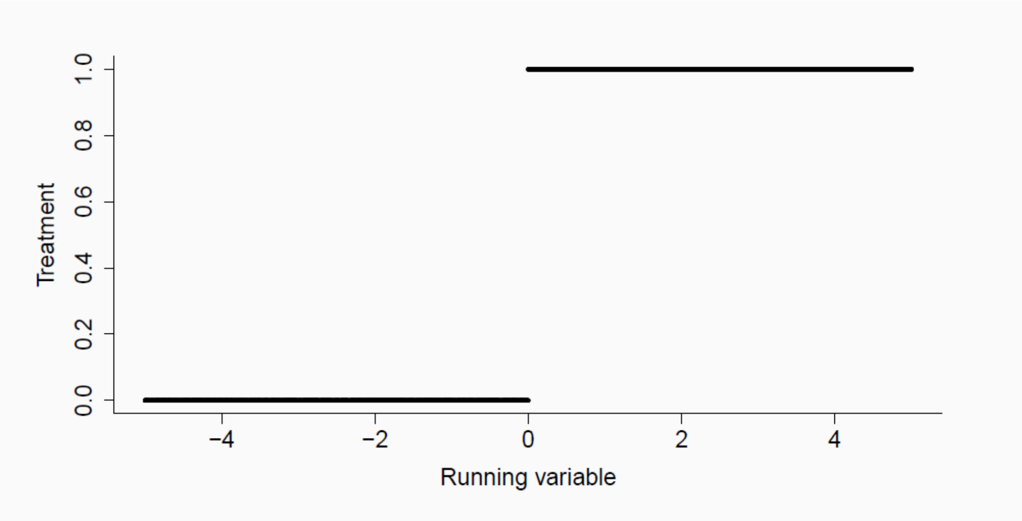
$$\delta = \lim_{X_i \rightarrow c_0^+} E[Y_i^1 | X_i = c_0] - \lim_{X_i \rightarrow c_0^-} E[Y_i^0 | X_i = c_0]$$

- Average causal effect of the treatment at the discontinuity

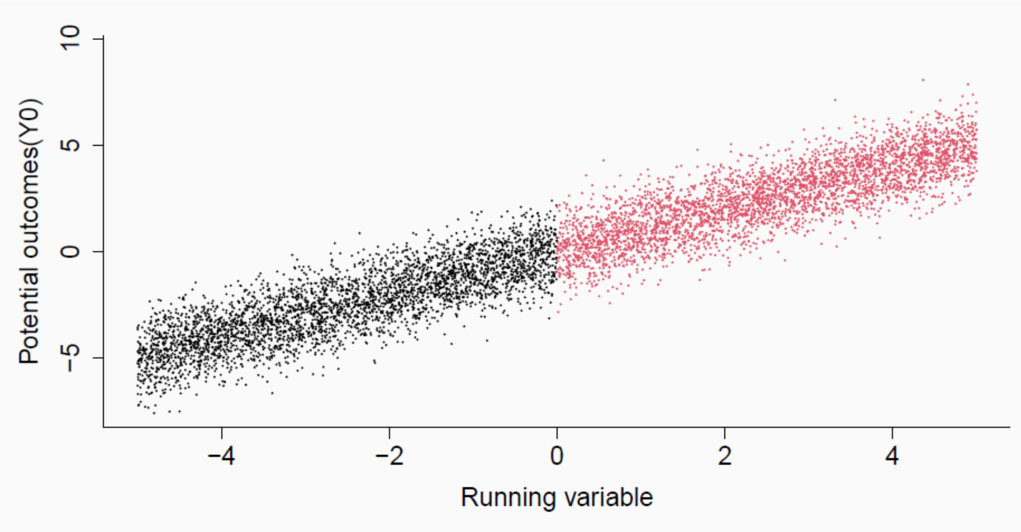
$$\delta_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c_0]$$

- T is correlated with X and deterministic function of X; overlap only occurs in the limit and thus the treatment effect is in the limit as X approaches c_0

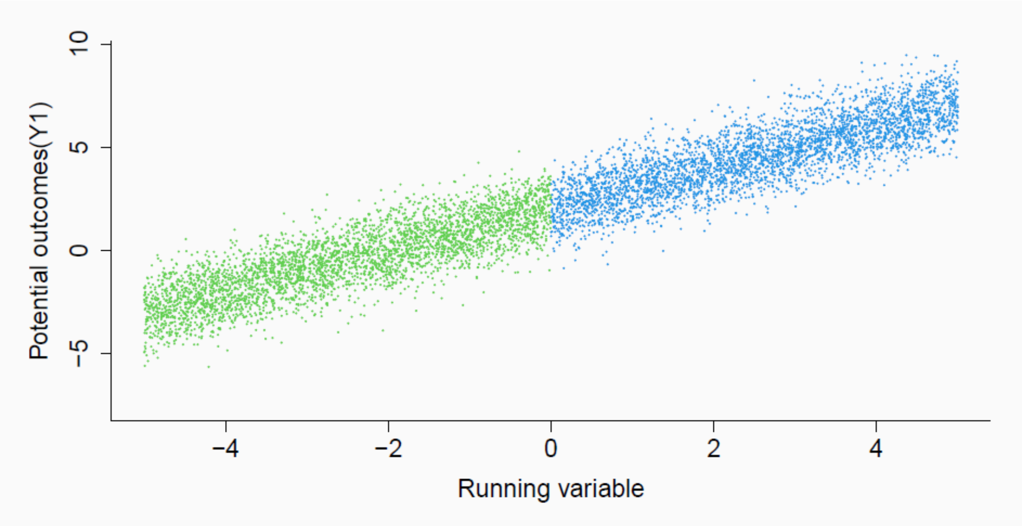
Basic RD Model in Stata – simulation



Basic RD Model in Stata – simulation

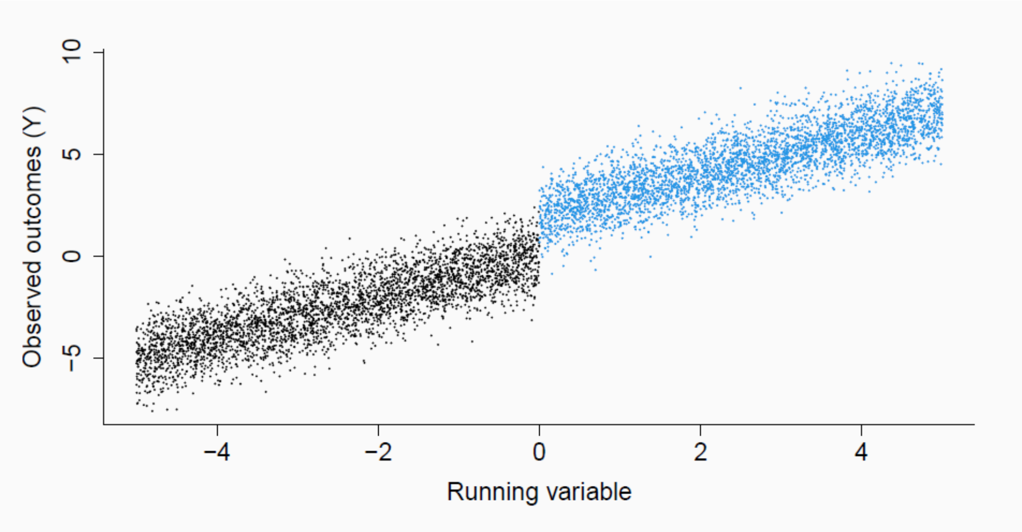


Basic RD Model in Stata – simulation

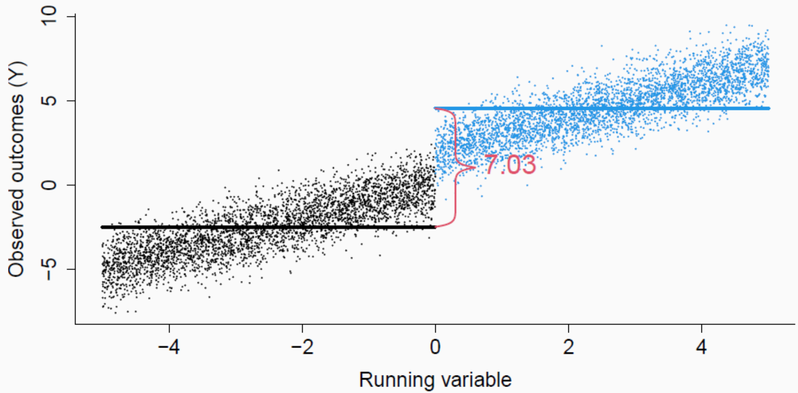


Basic RD Model in Stata – simulation

Basic RD Model in Stata – simulation

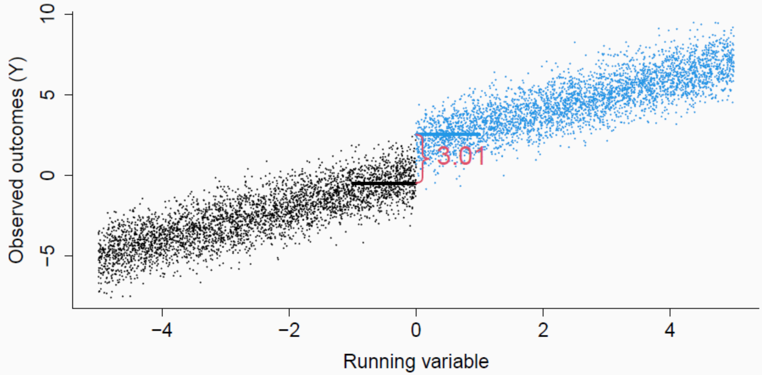


Basic RD Model in Stata – simulation



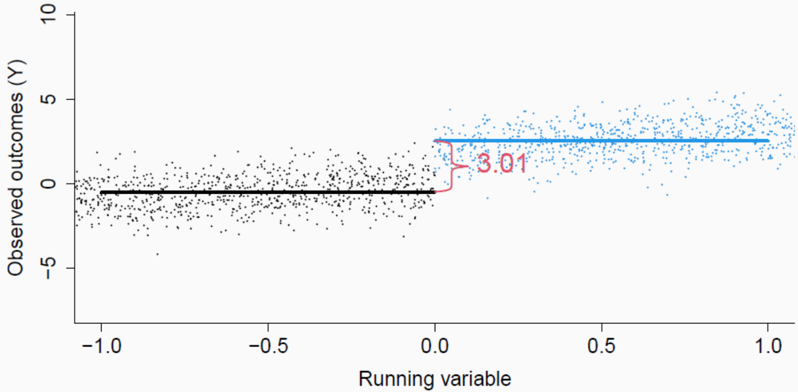
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

Basic RD Model in Stata – simulation



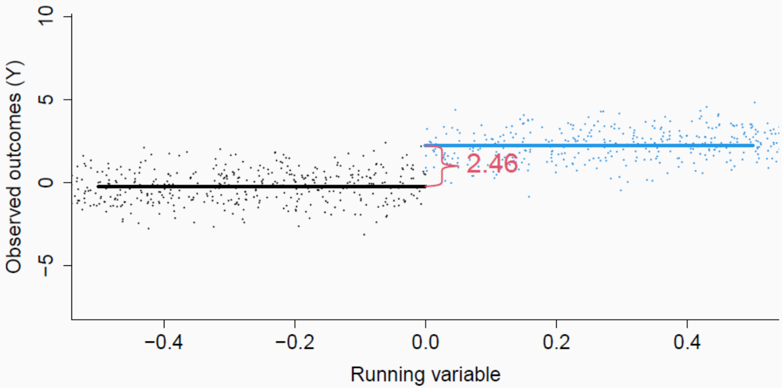
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-1 \leq X_i \leq 1$ via OLS

Basic RD Model in Stata – simulation



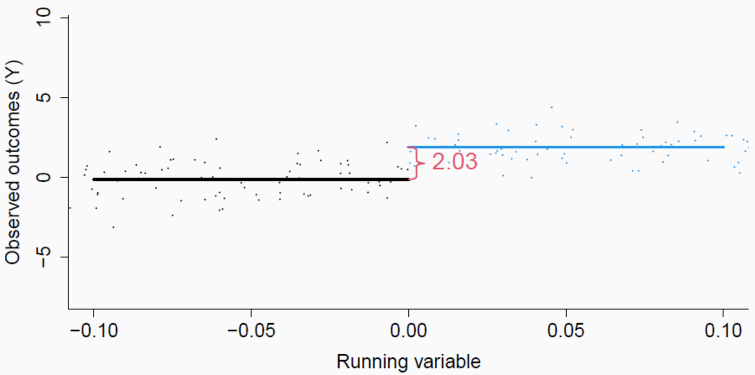
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-1 \leq X_i \leq 1$ via OLS

Basic RD Model in Stata – simulation



Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-0.5 \leq X_i \leq 0.5$ via OLS

Basic RD Model in Stata – simulation



Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-0.1 \leq X_i \leq 0.1$ via OLS

Basic RD Model in Stata – simulation

- There is still a problem with such estimation
- What can be a solution?
- Extrapolating in either side of the threshold
- Instead of using

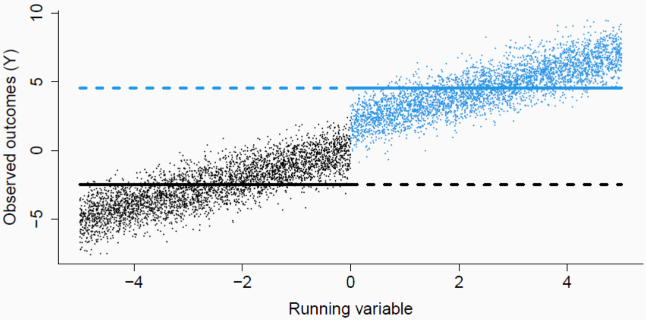
$$Y_i = \alpha + \delta T_i + \epsilon_i$$

  with T_i being either 0 or 1, if below or above the threshold
- We make explicit use of the value of X_i, the running variable

Extrapolation in Sharp RDD

- In RDD, the counterfactuals are conditional on X
- We use *extrapolation* in estimating treatment effects with the sharp RDD because we do not have overlap:
  - Left of cutoff, only non-treated observations, $T_i = 0$ for $X < c_0$
  - Right of cutoff, only treated observations, $T_i = 1$ for $X \geq c_0$
- The extrapolation is to a counterfactual

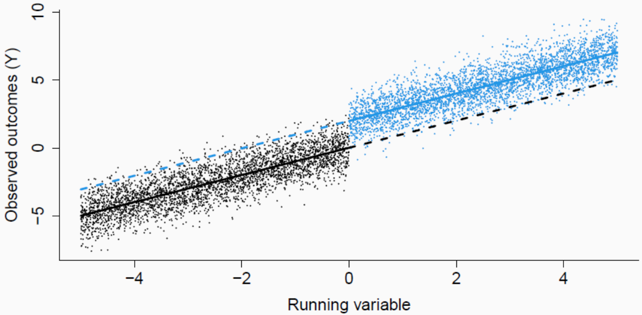
Extrapolation in Sharp RDD



Dashed lines are extrapolations

Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

Extrapolation in Sharp RDD



Dashed lines are extrapolations

Equivalent to estimating $Y_i = \alpha + \beta X_i + \lambda X_i * T_i + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

Smoothness in RDD

- A key assumption in the RD design is smoothness of conditional expectation functions (Hahn, Todd and Van der Klaauw 2001)
- $E[Y_i^0|X = c_0]$ and $E[Y_i^1|X = c_0]$ are continuous (smooth) in X at c_0
  - Potential outcomes not actual outcomes
  - If population average *potential outcomes*, Y^1 and Y^0, are smooth functions of X through the cutoff, c_0, then potential average outcomes won't jump at c_0
  - The cutoff is considered exogenous; implying no other factors influencing potential outcomes at c_0 change
  - Unobservables are also assumed to change smoothly through the cutoff

Smoothness Assumption in RDD

- The smoothness assumption allows us to use average outcome of units right below the cutoff as a valid counterfactual for units right above the cutoff
- Extrapolation is allowed if smoothness is credible, and extrapolation is nonsensical if smoothing isn't credible
- Why not directly testable? Because potential outcomes are not observable

Estimation in practice

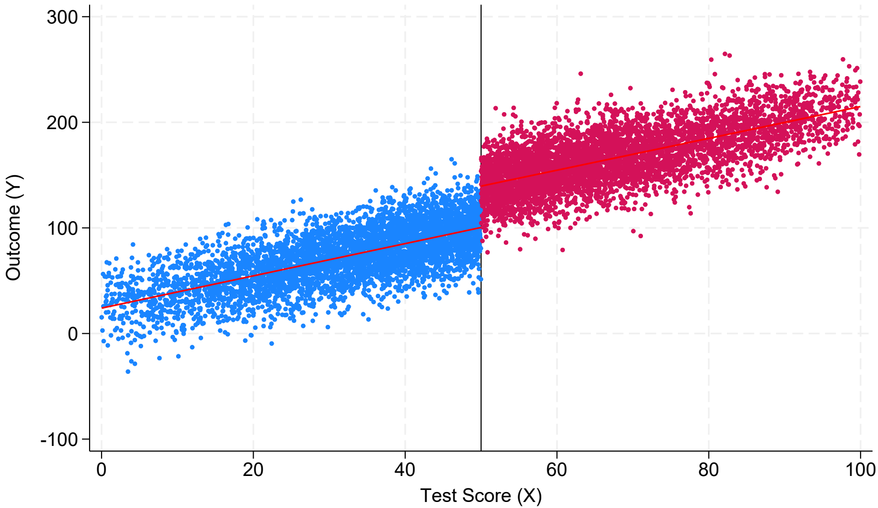
- In practice, RDD estimation is simple:

$$Y_i = \alpha + \beta(X_i - c_0) + \lambda(X_i - c_0) \times T_i + \delta T_i + \epsilon_i$$

- c_0 is used to transform the data with the threshold – it would not change the estimation, but only the intercept
- We can now run a simulation in Stata and see how it is done in practice
- We simulate

$$Y_i = 25 + 1.5 \cdot (X_i - 50) + 0 \cdot (X_i - 50) \times T_i + 40 \cdot T_i + \epsilon_i$$

Estimation in practice



Source	SS	df	MS
Model	25001092.8	3	8333697.59
Residual	3769513.31	9,544	394.961579
Total	28770606.1	9,547	3013.57558

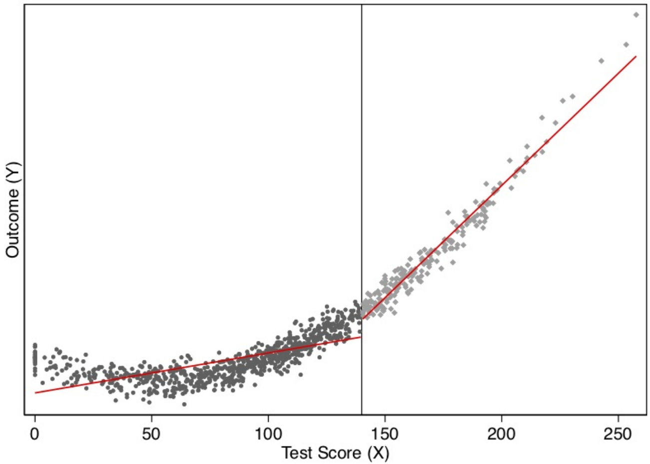
Number of obs	=	9,548
F(3, 9544)	=	21100.02
Prob > F	=	0.0000
R-squared	=	0.8690
Adj R-squared	=	0.8689
Root MSE	=	19.874

y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
tx	1.561328	.0226443	68.95	0.000	1.516941	1.605716
txT	-.0299213	.0322773	-0.93	0.354	-.0931917	.0333491
T	38.76193	.7109734	54.52	0.000	37.36827	40.15559
_cons	101.1867	.4987637	202.88	0.000	100.209	102.1644

Nonlinearity

- So far we made the assumption that the outcome is linear in the running variable
- We allow for a change in the slope at the threshold, as well as a treatment effect, but that might be two restrictive
- Imposing linearity when things are not linear, can lead to biased estimates

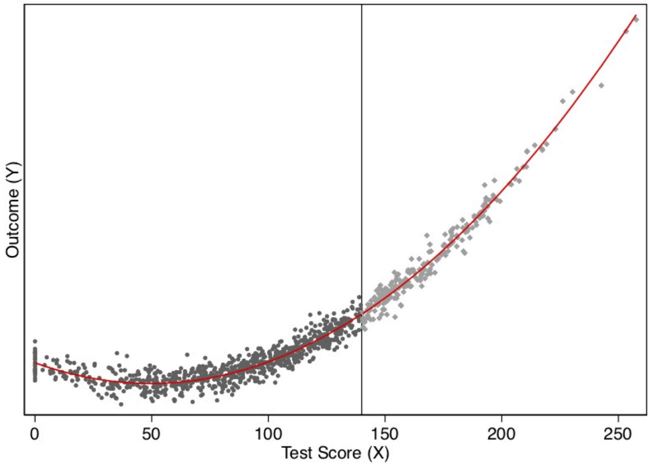
Nonlinearity



Nonlinearity

- To account for the possible nonlinearity, we need to allow for both sides of the threshold to be more complicated, typically second-order or third-order (cubic) polynomials
- In practice, this means fitting the data below the threshold and above the threshold separately to some polynomial and check whether there are statistically significant differences

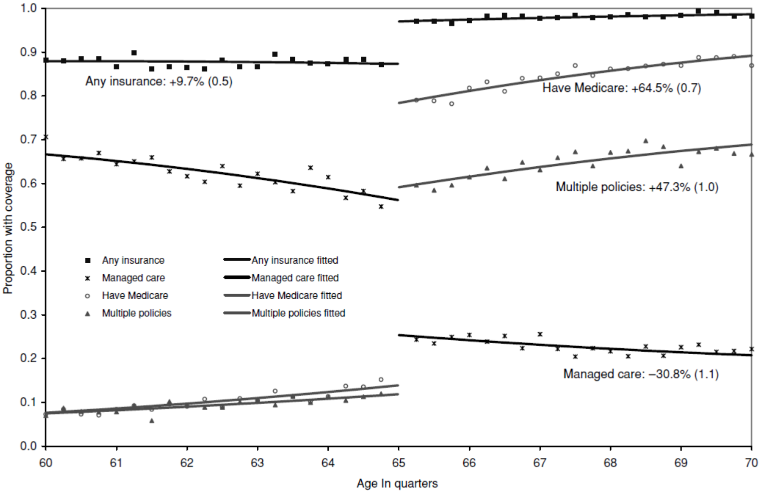
Nonlinearity



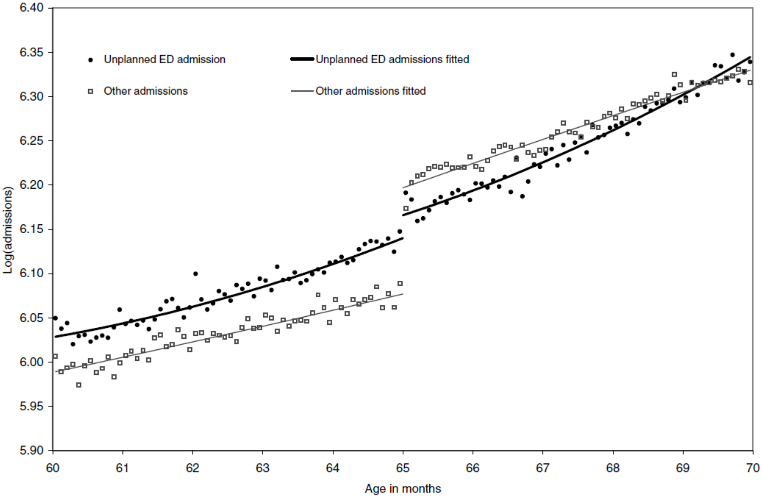
Classic example – Card, Dobkin, Maestas (2009)

- Medicare is the state-funded health insurance program in the US for people aged 65 and above
- It is clearly very costly, so it is worth checking if it saves lives and how much
- By the age cutoff, it is possible to use RDD to answer this question

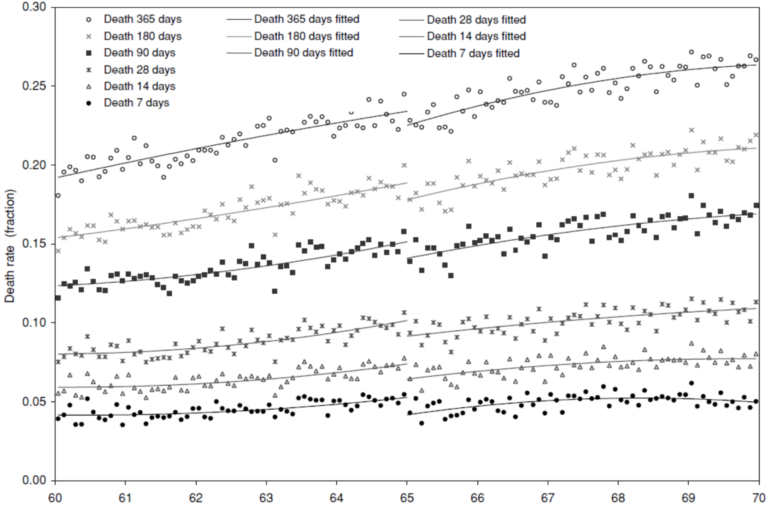
Card, Dobkin, Maestas (2009)



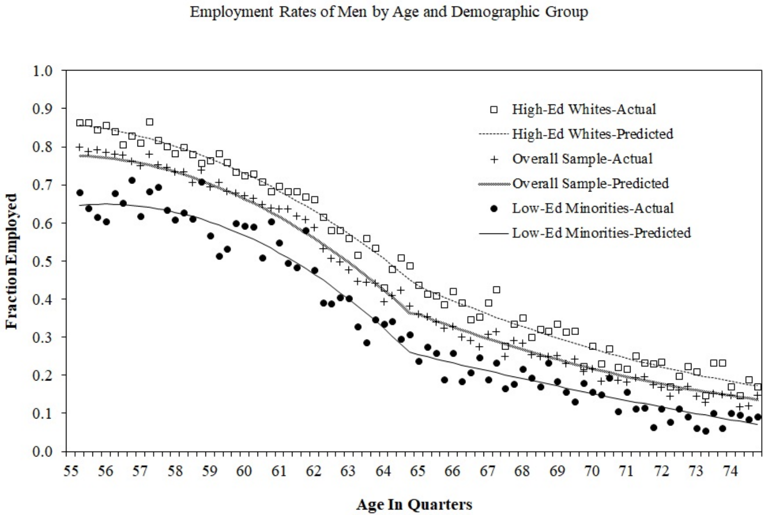
Card, Dobkin, Maestas (2009)



Card, Dobkin, Maestas (2009)



Card, Dobkin, Maestas (2009)



Takeaway points

- Regression discontinuity is a very useful tool to test causal effects in various contexts
- It requires an exogenously determined cutoff that can in practice be interpreted is creating randomly assigned treatment
- It is important to test whether indeed below and above the threshold the treated and untreated are similar in dimensions that are not the outcome variable and are not strongly correlated with it
- It is a good tool when using cross sectional data, or when we do not observe the pre-treatment period

For after reading week

- Read Card, Dobkin, Maestas (2009)
- Go over the simulation code
- Exercises will be posted before the next seminar