Applied Quantitative Methods II

Lecture 8: Regression discontinuity design

Klára Kalíšková

Outline

Intuition

Example

Students and merit scholarship

- RQ: What is the effect of scholarsip on the grades of students?
- Selection bias: only good students receive it - perform even better
  - comparison with/without by OLS biased up
- Exogenous variation: threshold of selection
- Students **right below and right above threshold should be the same**
- Compare their outcomes -T and C groups to get causal effect
  - effect only around cutoff - local treatment effect

1. What is important / problematic?
2. Where else can we find these cut-off values?

Intuition

- Problems in this approach:

1. students don't know about merit scholarship
2. students cannot persuade teacher to give them "extra" points
    1. not precisely: every ind. same probability of getting T
3. students are not very different below/above cutoff
4. no other discontinuities
5. no other effects
    - drinking age of 18(21), but also gambling, smoking, ...
6. depends on extrapolation (no direct observation of T-effect)
7. arbitrary choices - how "close", functions ...

Intuition

- Other examples:

1. age - 18/21, but also 60/65
    1. drinking, but also health insurance, ...
2. poverty scores - eligibility to programs
3. income thresholds
4. weight, height
5. rating
6. vote share
7. ...

RD: Introduction

- Based on a special type of **natural experiments**
- Occurs when the **probability of participation** in treatment discontinuously changes with the continuous variable X (forcing variable)

Example

estimate impact of welfare benefit on employment (Y) using age (X) threshold for eligibility for this welfare benefit

- X itself may be associated with outcome, but this association must be smooth

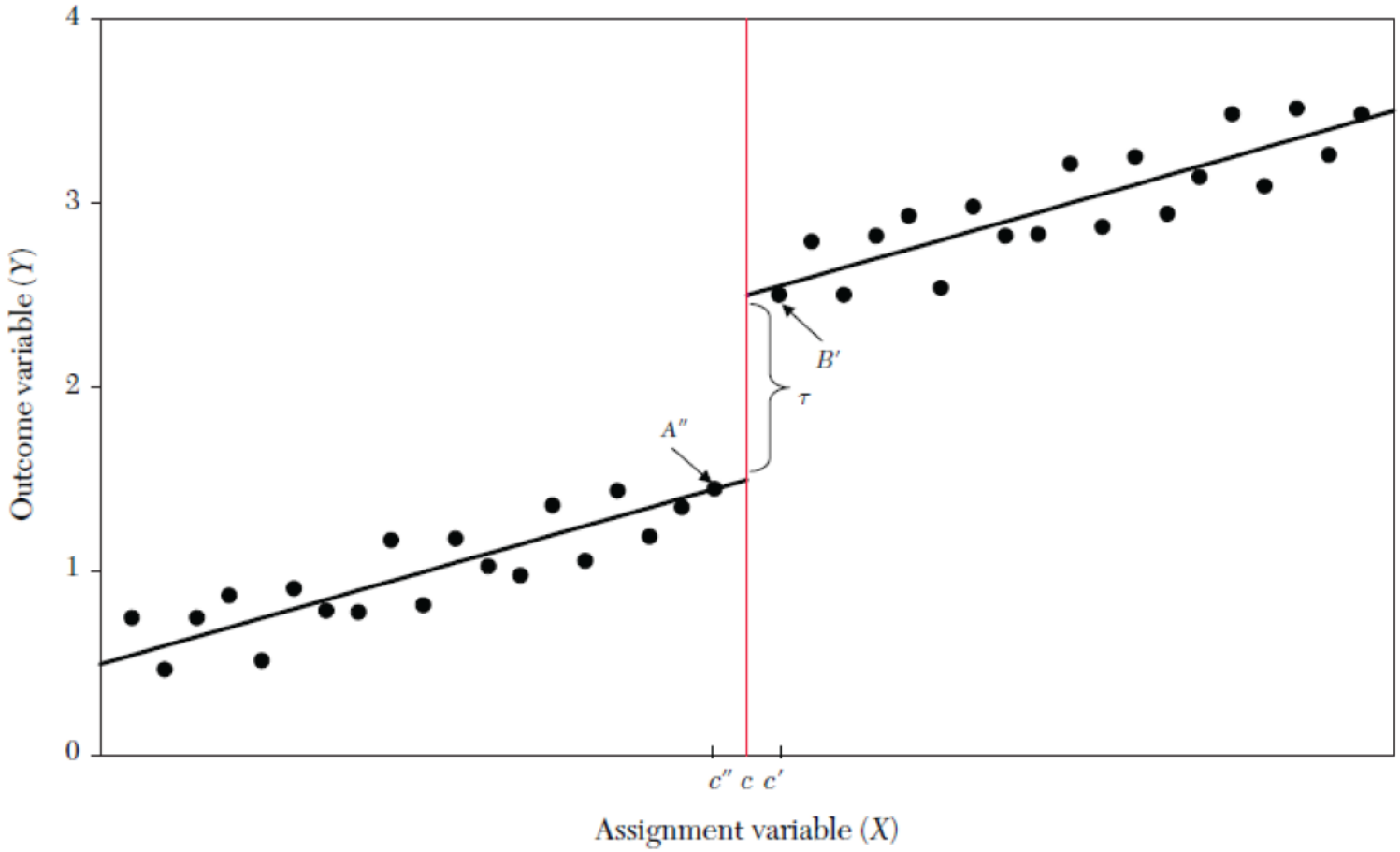
Example

Employment probability changes with age, but continuously (while eligibility for benefit changes with age discontinuously)

- Thus, any discontinuity of the outcome at the cutoff value is interpreted as evidence of a causal effect of the treatment.

RD: Introduction

- RD then compares people (cities, firms, countries) who are just affected by the rule and who are just not affected by the rule –> focus on those close to the threshold



RD: Introduction

- Assignment threshold may be 0/1 or in probability of treatment
- "Sharp" design - 0/1
- "Fuzzy" design - probabilities
  - then discontinuity used as an IV

Outline

RD - sharp design

Definition

- Probability of participation in treatment D changes from 0 to 1 at threshold $x*$

$$D_i = 1 \text{ if } x_i \geq x^* \text{ and } D_i = 0 \text{ if } x_i \leq x^*$$

- treatment status is exogenously determined by the level of threshold
  - not affected by individual decision

$$Y_i = \alpha + \beta X_i + \rho D_i + \epsilon_i$$

- we can estimate effect of the treatment around the threshold
  - usually centered at cutoff point: $X_i - c$, which changes interpretation of coefs!
  - non-treated on one side as counterfactuals of treated on other side
  - practical issue = manipulation (e.g. taxes, students persuading teachers)

RD - sharp design (definition)

Assumptions

- all unobserved determinants of Y are continuously related to X
- therefore we can use average outcomes just below threshold as valid counterfactuals
- cannot be directly tested - only suggestive evidence

Graphical example

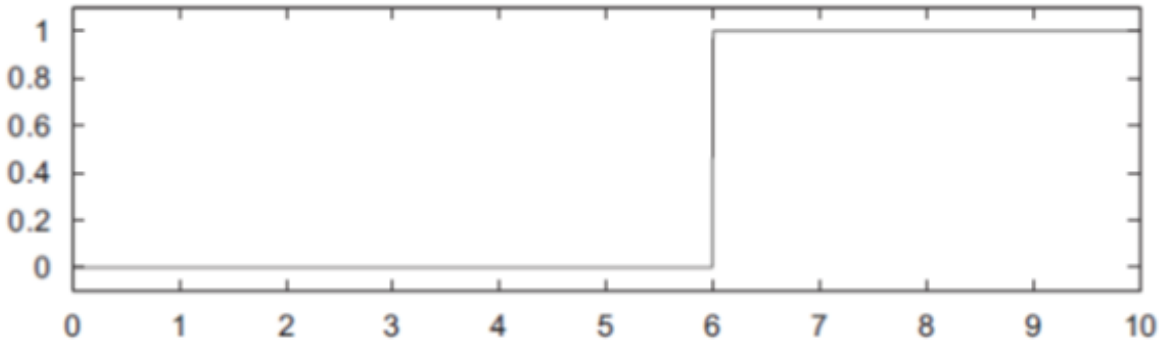


Fig. 1. Assignment probabilities (SRD).

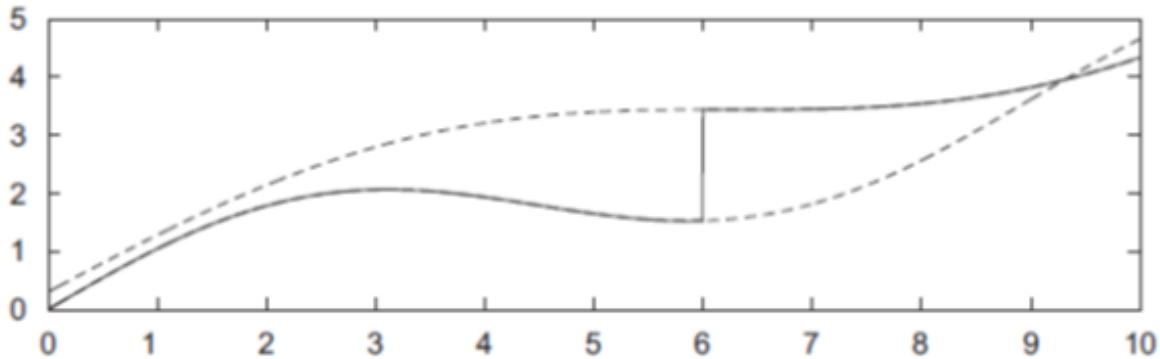
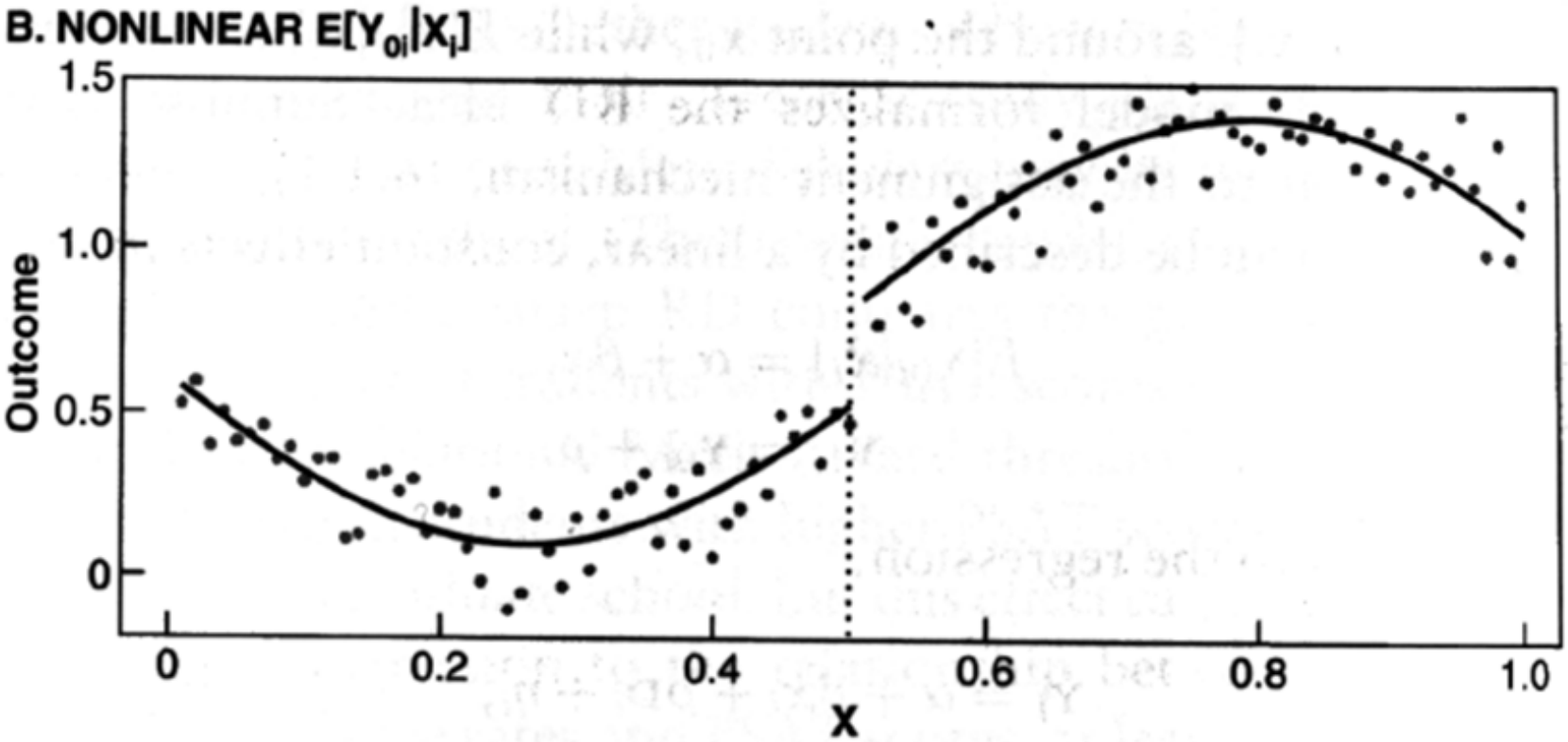


Fig. 2. Potential and observed outcome regression functions.

Graphical example - nonlinear case



RD - non-linear case

- The nonlinear relationship

$$E[Y_{oi}|x_i] = f(X_i)$$

- then we can estimate it by fitting

$$Y_i = f(X_i) + \rho D_i + \mu_i$$

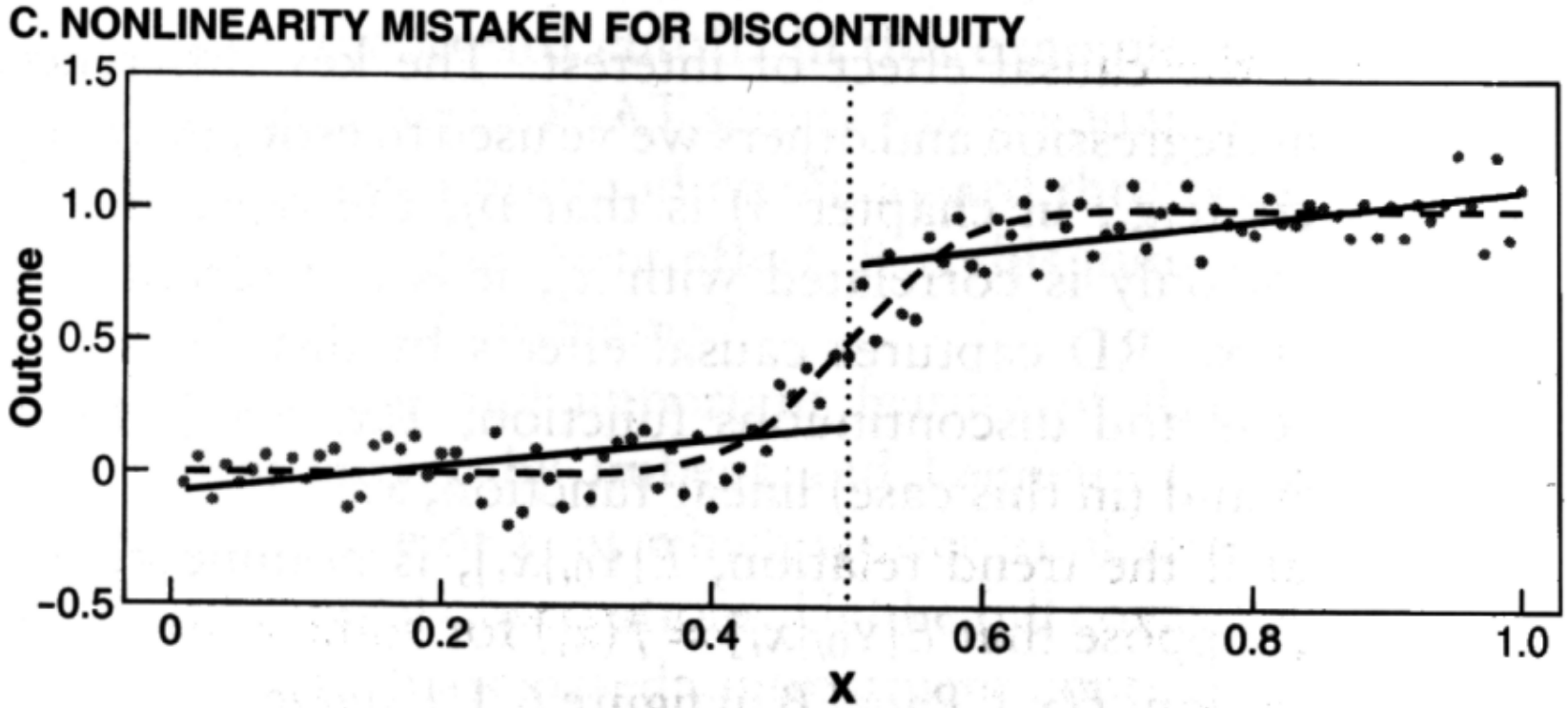
1. Non-parametric kernel estimates
   - choice of bins - then averages and weighting by kernel density
2. either p-th polynomial (possibly different on both sides - interaction with D)

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_p x_i^p + \rho D_i + \epsilon_i$$

   - Problem: you use data also far away from cutoff
   - good to use different functions - **robustness** of results
     - polynomials, different functions, as well as local OLS and bandwidth

RD - non-linear case

- Sometimes, what seems to be jump is non-linearity



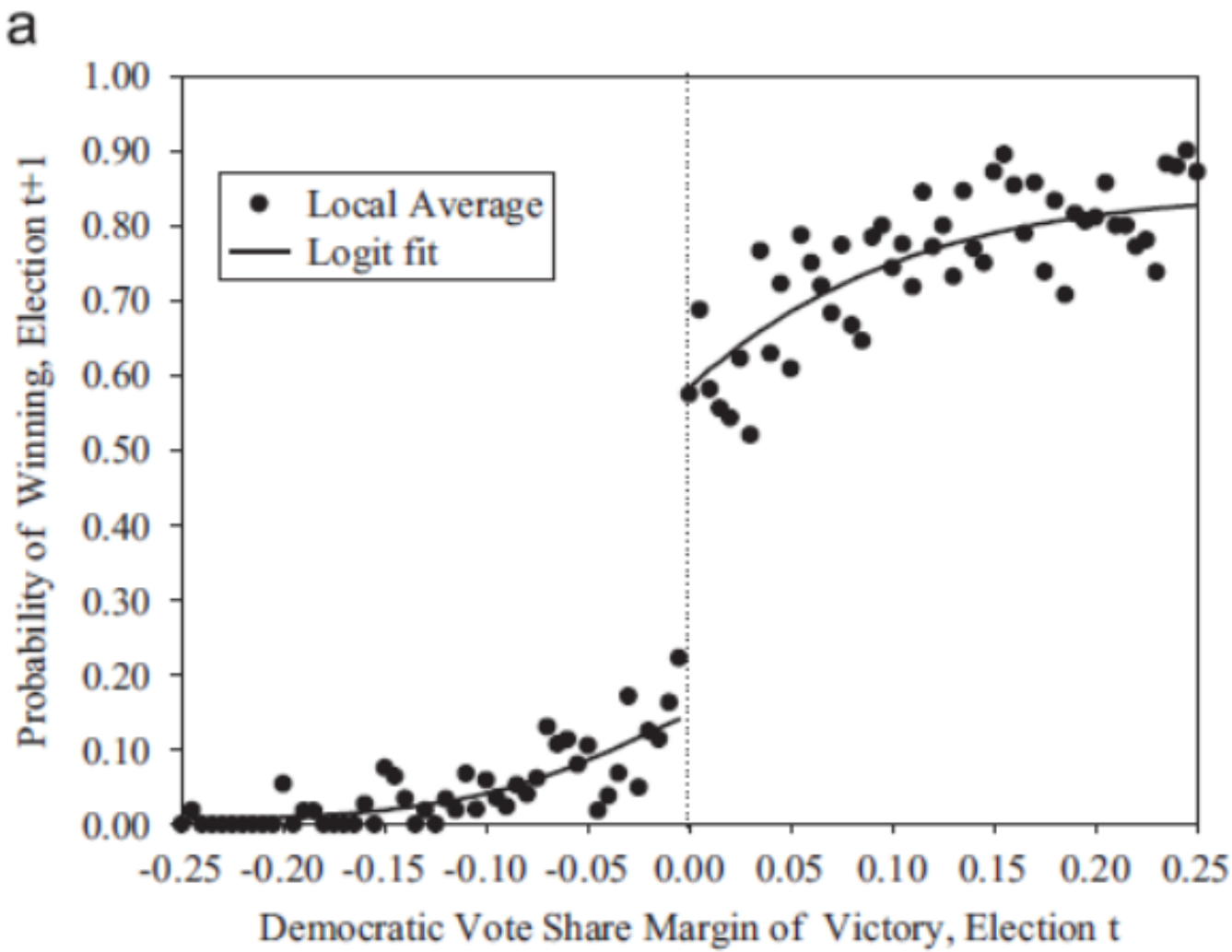
Example - US elections (Lee, 2008)

- Use of sharp RD design to estimate the effect of being incumbent politician on the probability of winning next elections
- Theory: incumbents may use resources and privileges of the office to gain advantage over opponents
- OLS is biased due to unobserved heterogeneity
  - incumbents already won and are different than opponents
- Data: 50 years, Democratic incumbents in US congressional elections
- Probability of winning the election in election year $t + 1$ by comparing candidates who just won compared to candidates who just lost the election in year t

Example - US elections (Lee, 2008)

- Regression discontinuity idea:
  - In elections where the result is very tight, the candidate who just barely won is very similar to the candidate who just barely lost in unobservable characteristics
  - But the one who barely won becomes an incumbent, while the other does not
- It is a sharp design, because the probability of becoming an incumbent shifts from 0 to 1
- Comparing election results of the next elections of those who barely won and those who barely lost in the previous elections

Example - US elections (Lee, 2008)



Outline

RD – Fuzzy design

- X is not the only factor that determines participation in treatment, there are other (unobserved) factors
- Probability of participation does not jump from 0 to 1 at the threshold, but maybe changes from 0 to 0.7 (or from 0.1 to 0.7)

Example

unemployed below certain age threshold are eligible for a training program, but not all of them take advantage of it

Example

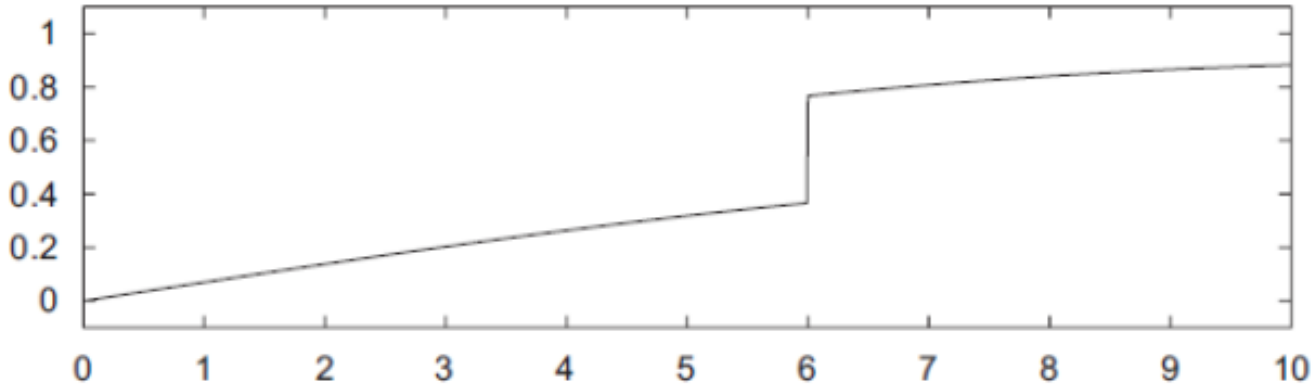


Fig. 3. Assignment probabilities (FRD).

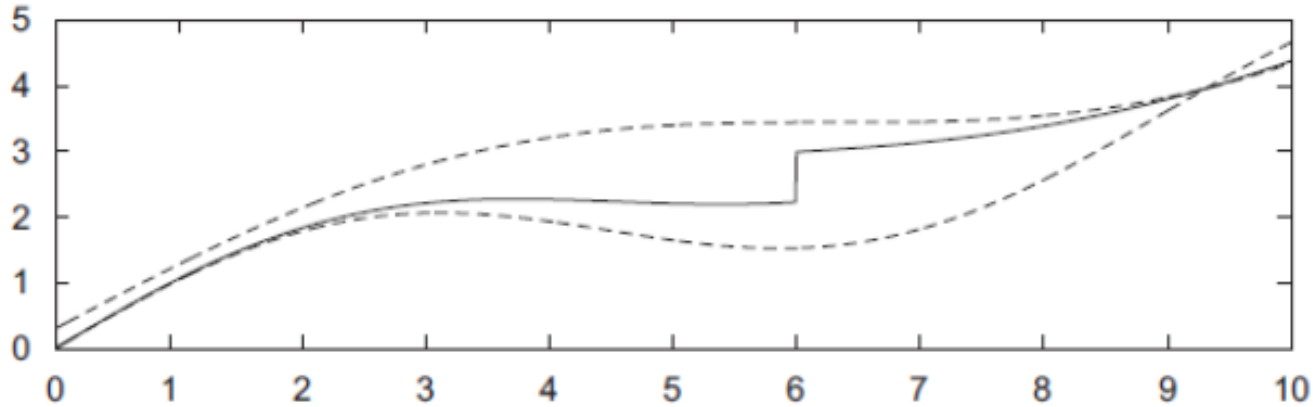


Fig. 4. Potential and observed outcome regression (FRD).

RD – Fuzzy design

- Fuzzy RD is equivalent to:
  - regressing outcome Y on treatment D using dummy for the forcing variable being above the threshold ($x > x^*$) as the instrument, applied on the subsample of x: $[x^* - h, x^* + h]$
- RD is less demanding about exogeneity:
  - IV: instrument must be exogenous (cannot affect outcome directly)
  - RD: x can affect outcome directly, but continuously (discontinuous effect must go through treatment, it cannot be direct)

RD – Fuzzy design - examples

Matsudaira (2007)

- Effect of remedial teaching program
- encouragement for students with score less than "c"
- The effect estimated only on compliers – those who are affected by the threshold "c" and *decide* to enroll

Van der Klaauw (2002)

- effect of financial aid on college admission acceptance
  - SAT (+other scores) determine eligibility for financial aid
- What is the effect of financial aid?
  - Causal effect: aid offer attracts more students to the college
  - Other effects: students with higher financial aid have usually better outside option from other schools
  - Other effects than discontinuity might matter too

Outline

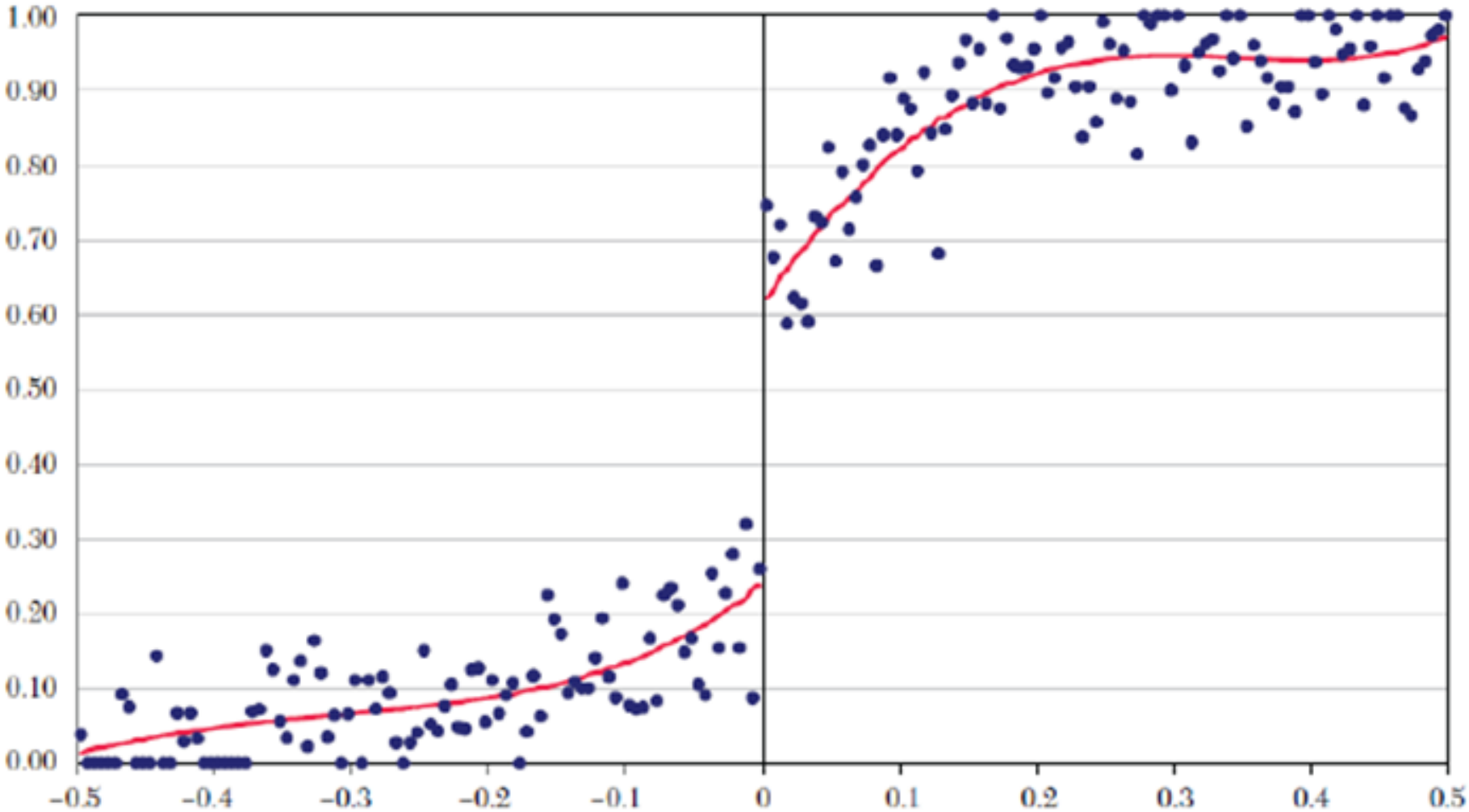
Implementation of RD

Step 1: graphical analysis

1. Plot d (treatment) vs. x (forcing variable): Is there a jump at x^* (cutoff)?
   - Important for fuzzy design, it is like the first stage for IV.
2. Plot Y (outcome) vs. x: Is there a jump at x^*?
   - If yes, the treatment effect is present
3. Plot other covariates vs. x: Is there NO jump at x^*?
   - There should be *no jumps in other covariates*
   - people around the threshold should be similar in characteristics (observable and unobservable)
4. Plot the density of x: Look out for clustering of people just above x^*
   - Clustering would indicate manipulation -> do not want that

Implementation of RD

Step 1: Outcome by forcing variable (Lee and Lemieux, 2010)



Implementation of RD

Step 1: Covariate by forcing variable (Lee and Lemieux, 2010)

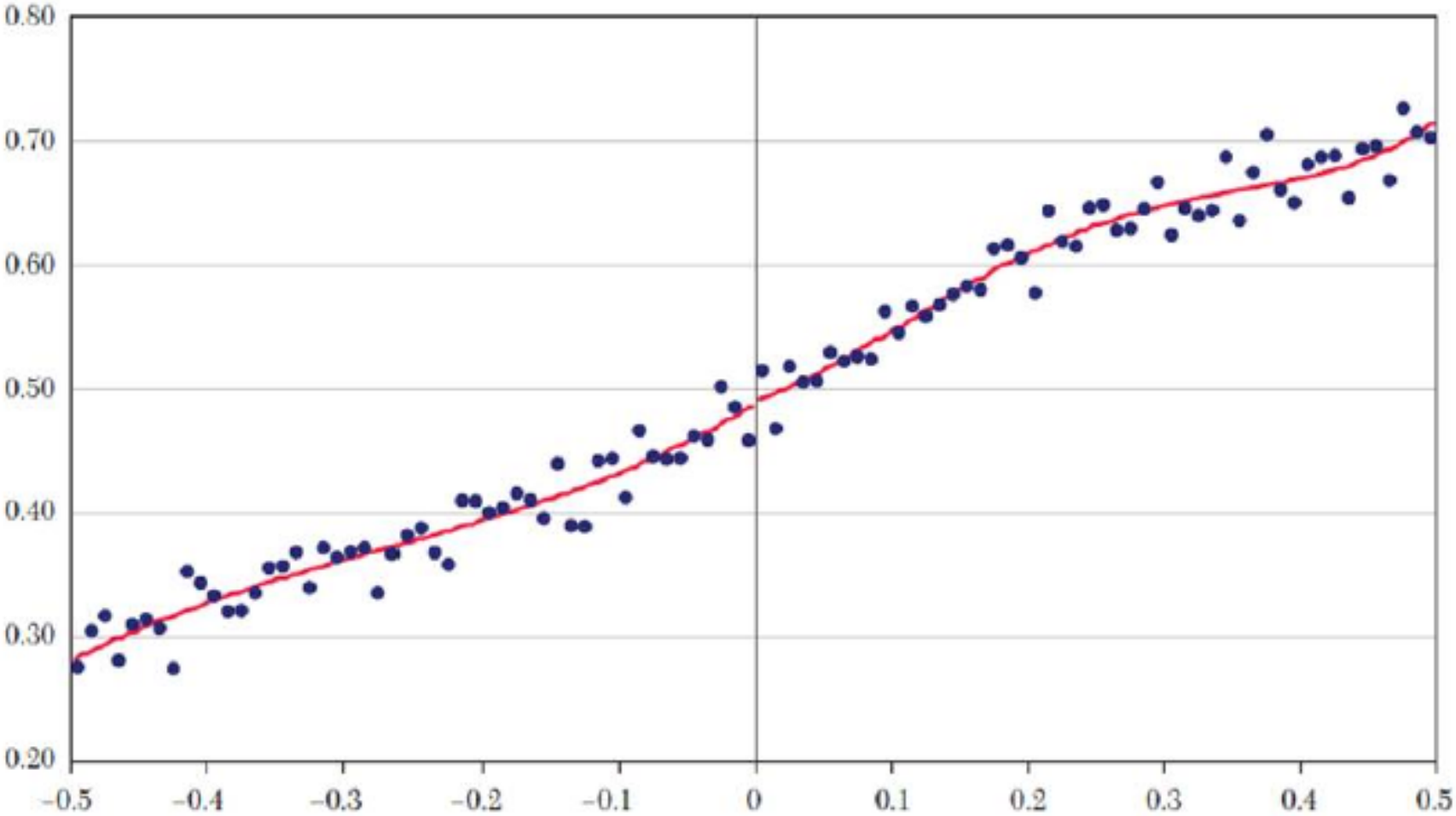
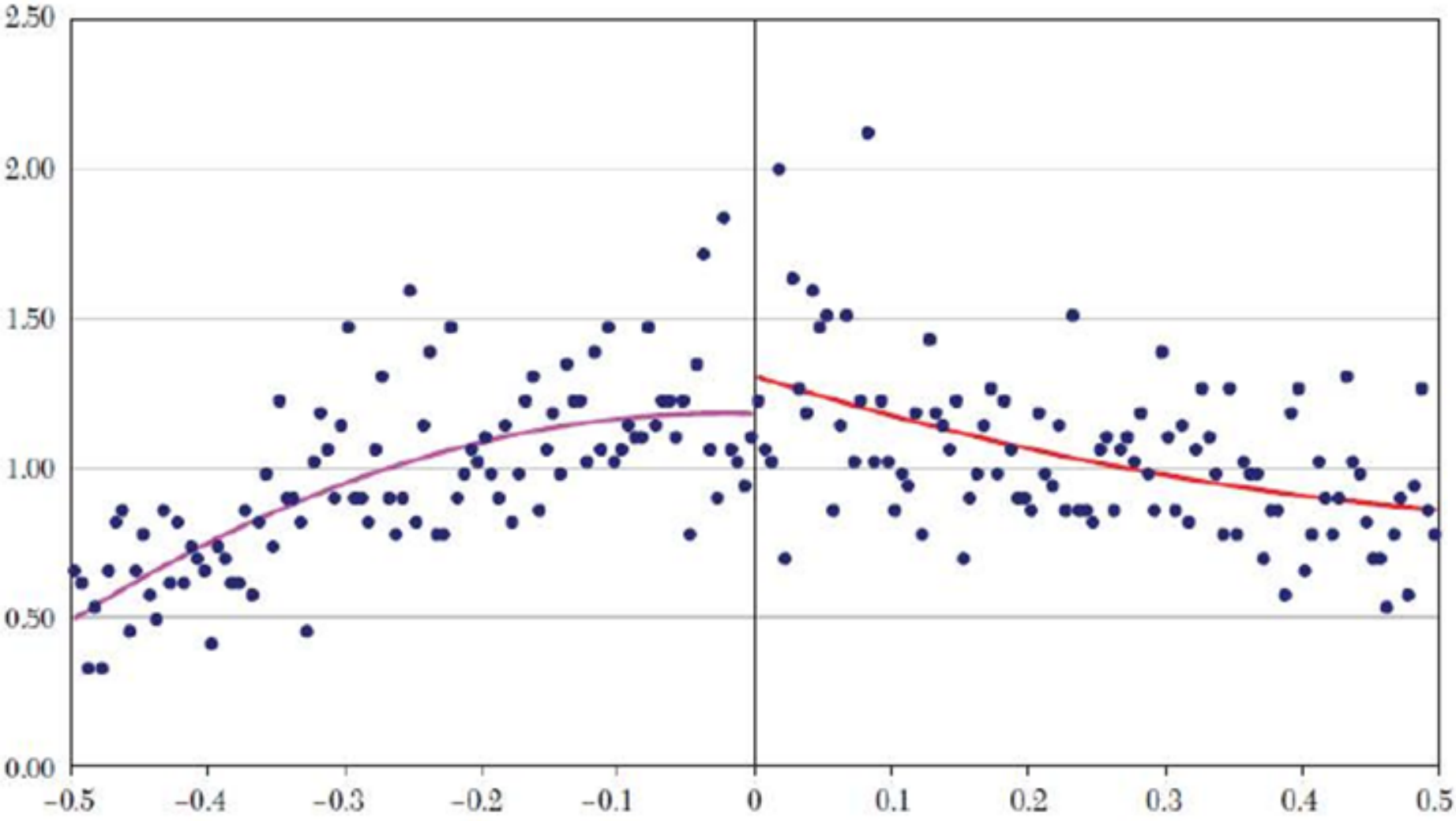


Figure 17. Discontinuity in Baseline Covariate (Share of Vote in Prior Election)

Implementation of RD

Step 1: Density of forcing variable (Lee and Lemieux, 2010)



Implementation of RD (sharp)

Step 2: Estimation

- Choose a window width h around the cutoff x^*
  - Calculate $E(Y|x^* - h \leq x < x^*)$ and $E(Y|x^* \leq x < x^* + h)$
- Calculate the difference in average Y between those just below and just above the threshold
- Use regression analysis (Kernel regression):
  - Run regression of Y on the dummy for being above the threshold ($x > x^*$) and other control variables for those around the threshold
- Problem: need lot of datapoints in the neighborhood of x^*

Implementation of RD (sharp)

Step 2: Estimation

- Problems:
  1. Choice of bandwidth
  2. Needs a lot of data in the neighborhood of x^*
- That can be solved by a polynomial method (estimate $f(x)$ as a function of higher order polynomials in x and use wider range of data)

RD-summary of issues

- Identifies only *local effect* restricted to the discontinuity point
- Once the design is fuzzy
  - The discontinuity applies only to compliers
  - Similar problems with *IV and LATE*
  - Unobserved factors can drive decision
  - If individuals can manipulate to which side of threshold belong based on their expectations, problems are even more severe

Outline

Example 1 - Angrist and Lavy (1999)

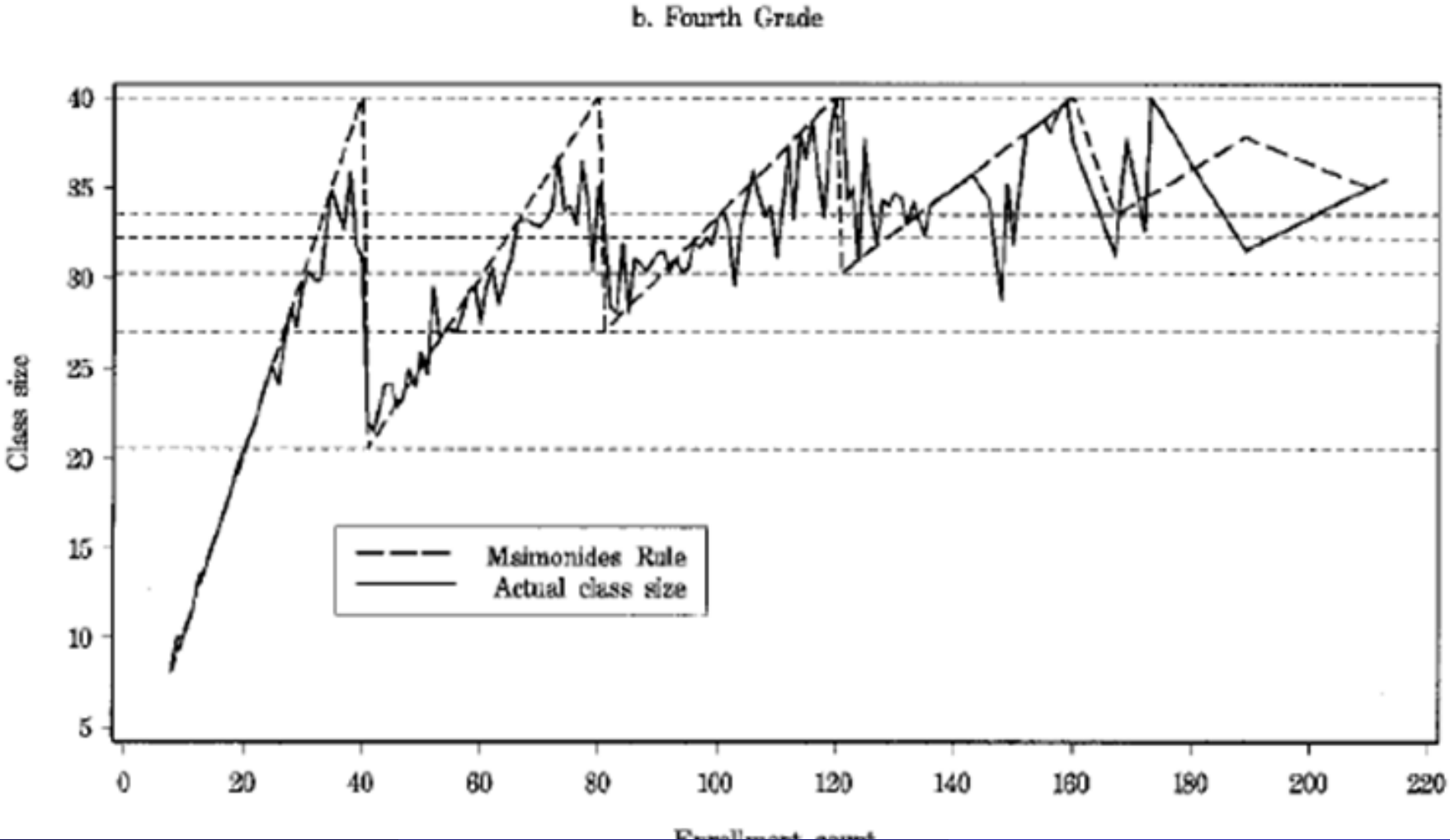
Motivation

- RQ: Effect of class size on students' test scores
- Data - info on schools / test outcomes, Israel
- Maimonides rule (from Talmud): classes should have max 40 students
  - school with 40 students $\rightarrow$1 class
  - school with 41 students $\rightarrow$2 classes (20 and 21)
  - above classical RD design: multiple discontinuities
- predicted class size in school with enrolment e_s is $m_s = \frac{e_s}{int(\frac{e_s-1}{40})+1}$
- discontinuities: 40-41, 80-81, ...

Example 1 - Angrist and Lavy (1999)

Comparison of predicted and actual class size

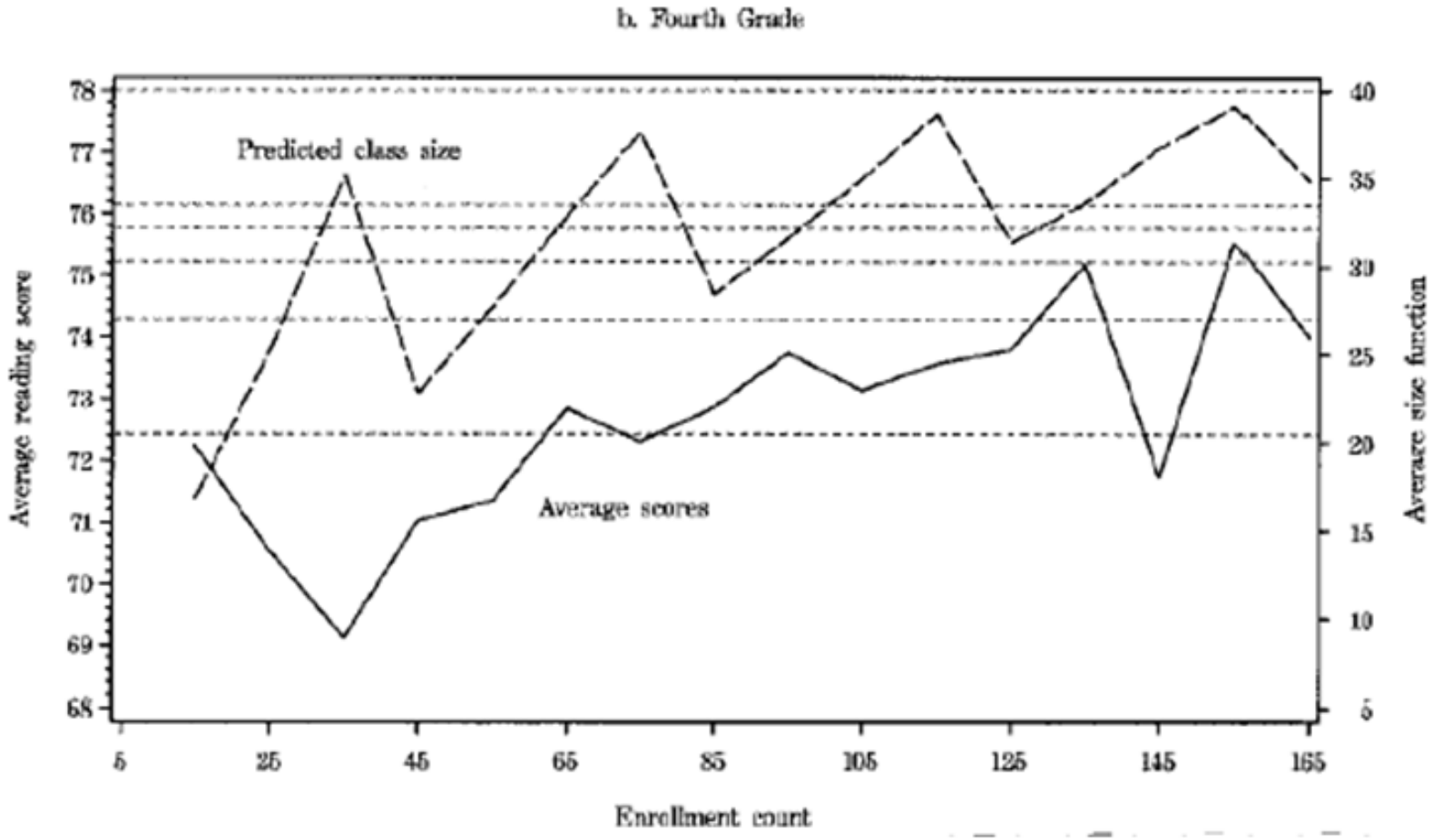
- Rule is not followed strictly => fuzzy design



Example 1 - Angrist and Lavy (1999)

Reduced form – test scores vs. enrollment

- Average test scores partly mirror predicted class size
- Students in schools with bigger enrollment do better on average



Example 1 - Angrist and Lavy (1999)

OLS results: biased

- Better socioeconomic background -> push for smaller classes
- Weaker students might be put in smaller classes

	5th Grade					
	Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean score		74.3			67.3	
(s.d.)		(8.1)			(9.9)	
Regressors						
Class size	.221	−.031	−.025	.322	.076	.019
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)
Percent disadvantaged		−.350	−.351		−.340	−.332
		(.012)	(.013)		(.018)	(.018)
Enrollment			−.002			.017
			(.006)			(.009)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30
R^2	.036	.369	.369	.048	.249	.252
N		2,019			2,018	

Example 1 - Angrist and Lavy (1999)

RD results

- 1.stage – regress class size on a predicted class size + other determinants
- 2.stage – regress test results on fitted class size
- Effect of class size is now significantly negative

	Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean score		74.4			74.5	
(s.d.)		(7.7)			(8.2)	
Regressors						
Class size	−.158	−.275	−.260	−.186	−.410	−.582
	(.040)	(.066)	(.081)	(.104)	(.113)	(.181)
Percent disadvantaged	−.372	−.369	−.369		−.477	−.461
	(.014)	(.014)	(.013)		(.037)	(.037)
Enrollment		.022	.012			.053
		(.009)	(.026)			(.028)
Enrollment squared/100			.005			
			(.011)			
Piecewise linear trend				.136		
				(.032)		
Root MSE	6.15	6.23	6.22	7.71	6.79	7.15
N		2019		1961	471	

Outline

Example 2 – Card, Lee, Pei and Weber (2012)

Regression kink design (RKD)

- RQ: Does the level of unemployment benefits affect the length of unemployment in Austria?
- Unemployment benefits are based on income
- Benefit formula exhibits 2 kinks:
  - There is minimum benefit level
  - Then benefits are 55% of earnings in the base period
  - There is maximum benefit level adjusted every year

Example 2 – Card, Lee, Pei and Weber (2012)

Regression kink design (RKD)

- People with dependents get small supplements
  - We see variation in unemployment benefits
- Not perfect correspondence to formula due to mistakes
  - not everyone receives benefits - mistakes in administrative data
- Is this fuzzy or sharp design?

Example 2 – Card, Lee, Pei and Weber (2012)

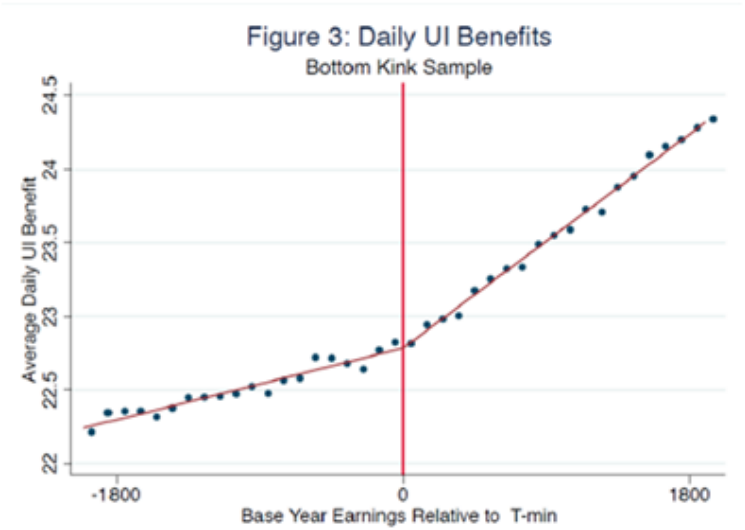
Regression kink design (RKD)

- One can use the kink in formula to identify the causal effect of the policy
- Instead of a jump in outcome one should expect a jump in the *first derivative*

Example 2 – Card, Lee, Pei and Weber (2012)

Regression kink design (RKD)

- Negative slope – people with higher earnings find job easier
- Kink – relationship becomes shallower once benefits change
  - Earnings x UI benefits



Example 2 – Card, Lee, Pei and Weber (2012)

Regression kink design (RKD)

- Earnings x Time to next job

- Negative slope – people with higher earnings find job easier
- Kink – relationship becomes shallower once benefits change

Conclusion

- When to use RD design?
- Types of RD design:
  - sharp RD
  - fuzzy RD
- Careful when centering on cutoff value - interpretation
- Estimation
  - parametric / non-parametric
- Problems of RD design
  - other effects
  - discontinuity of other characteristics
  - clustering of units above threshold

Outline

1. Know what you can trust

- Scientific knowledge - ranking of journals based on ... citations
- Metrics: Impact factor, Article influence score, Eigenfactor ...
- Impact factor:
  - average number of citations of articles published in recent time in journal
- Journals have to be included in a database that monitors citations
  - E.g. Scopus, Web of Science, RePeC, ...
- Fraud: citation manipulation possible
  - e.g. self-citations
- Important:$\Rightarrow$ ranking of journals
  - in science in general, in economics, in a field of economics
- Top-journals in economics: https://ideas.repec.org/top/top.series.simple10.html
- *Lesson1: Look at articles in journals that rank the best*

2. Get a general knowledge about your topic

- Read:
  - read a book
  - read a summary article
  - read a review article
  - read the latest articles from top journals
  - read those cited in that articles
  - read those citing that article
  - *Hint: easier to read about sth of your own interest*
- Read slow and fast
  - careful, slow vs. fast skimming through
- learn the important people in the field (names, research topics)
- get the general picture of what is going on now
- *Lesson 2: Read a lot*

2. Get a general knowledge about your topic

- How to find the articles and books?
- Go to library :-)
- Use scholar.google.com
  - very important
- Look into specialized databases / publishing houses
  - ScienceDirect, JSTOR, Springer, Wiley, ...
- Use *vpn* or other distance-access
- Download pdfs, but also supplementary material
- If cannot access fulltext, working-paper version fine
  - publishing a paper takes years in economics

3. Do not waste time

- be **organized** in your time and attitudes
  - less by-accident findings
  - imagine you lose memory/ you do it for a friend with no knowledge
- have your **data** organized
  - well described variables, values
  - well described procedures in do-files, even excel files
- have your **literature** organized
  - use a citation software (Mendeley, Zotero, EndNote, Bib-tex, ...)
  - use a pdf organizer (Mendeley, Zotero, Benubird, ...)
  - important papers - annotated bibliography
    - describe the paper in short but efficient way so that you don't need to read it again
  - use tags, keywords or groups of papers
- have deadlines and organized colleagues
- *Lesson 3: Get organized and don't waste time*

4. Get feedback

- uncertainty about your progress can be demotivating
  - actually, can be very frustrating
  - main reason of not completing PhD
- **talk** to your supervisor regularly
- talk to **other** relevant **faculty** members
- talk to your **friends** about the topic over lunch
- **present** your research
  - in class, conferences, lunch workshops, ...
- make someone read your papers
- when sure about your skills: try to explain it to your grandmother
- Remember: failures and mistakes are part of the process!
- *Lesson 4: Do not get fed-up with it*

Summary of successful research

- Know what you can trust - good journals
- Read a lot
- Be organized
- Don't get fed-up with it

Bibliography

- Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142(2), 675–697. doi:10.1016/j.jeconom.2007.05.004
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355. doi:10.1257/jel.48.2.281

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání