Lecture 4: Regression Discontinuity Designs

Yu Bai

City University of Macau

Recap: Drink, Drank...

- MLDA (Minimum Legal Drink Age)
- Variation in state MLDA laws is easily exploited in a DD framework.
  - Alabama lowered its MLDA to 19 in 1975, but geographically proximate Arkansas has had an MLDA of 21 since Prohibition's repeal.
  - Did Alabama's indulgence of its youthful drinkers cost some of them their lives?

Question

Can we explore this in a DID framework?

Recap: Drink, Drank...

- What we do is not a standard DID.
- Remembered what we do?

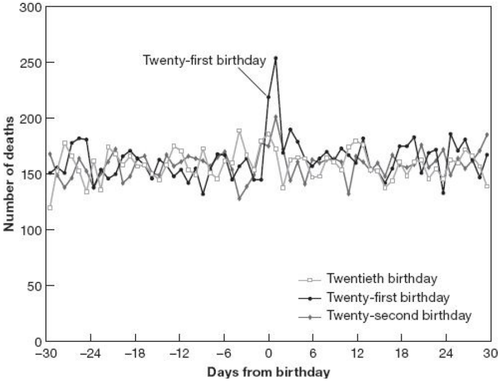
$$Y_{st} = \alpha + \delta_{rDD}\text{LEGAL}_{st} + \gamma_s + \lambda_t + e_{st},$$

where

  - LEGAL_{st}: the proportion of 18–20-year-olds allowed to drink in state s and year t
  - γ_s and λ_t: fixed effects (LSDV)

Birthdays and funerals

FIGURE 4.1
Birthdays and funerals



A new approach

- We are going to explore a new approach called Regression Discontinuity Designs (RDD).
- Main idea: As treatment intake may be determined by a running variable, we can quantify the causal effects by exploring the discontinuity around the cutoff value of that running variable.

In a nutshell

- Each unit has a score on a running variable which determines treatment.
- The cutoff is the value of the running variable at which treatment is assigned.
- There is a discontinuous change in probability of receiving the treatment at the cutoff.
  - Abrupt change in treatment probability can be used to learn the **local** treatment effect.

Using discontinuity to eliminate selection bias

Core intuition

Units with scores barely below the cutoff can be used as counterfactuals for units with scores barely above it.

- Heavily rely on knowing and having access to a running variable which determines treatment status
- Widely used in settings such as
  - elections
  - administrative programs
  - geographic boundaries

Sharp RDD

- Suppose that there is a binary treatment variable D_i which is completely determined according to

$$D_i = \begin{cases} 1, & \text{if } X_i \geq c \\ 0, & \text{if } X_i < c \end{cases}$$

  where
  - X_i is known as the "forcing" or "running" variable, which can be correlated with (potential) outcome.
  - c is a fixed cutoff point.

Sharp RDD

- Consider the usual independence assumption

$$Y_i(0), Y_i(1) \perp\!\!\!\perp D_i | X_i$$

- This condition should hold trivially, as conditioning on the covariates there is no variation in the treatment.
- However, it cannot be exploited directly, as it requires

$$0 < \mathbb{P}\left(D_i = 1 | X_i = x\right) < 1,$$

  for any values X_i can take, which is fundamentally violated.
- We need to restrict our attention on a particular set of values X_i can take.

Identification with Sharp RDD

- Suppose that we define τ_{SRD} as

$$\begin{aligned}\tau_{\text{SRD}} &= E[Y_{1i}|X_i = c] - E[Y_{0i}|X_i = c] \\ &= E[Y_i|X_i = c, D_i = 1] - E[Y_i|X_i = c, D_i = 0]\end{aligned}$$

Question

What exactly are we doing here?

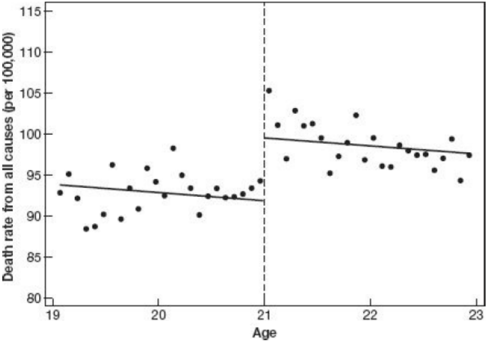
Identification with Sharp RDD

- Consider the interval $[c - \Delta, c + \Delta]$, where Δ is a small positive constant
- Further assume that
$$E[Y_i|X_i] = f(X_i),$$
is a continuous function.
- Then,
$$\begin{aligned} &\lim_{\Delta \to 0} \left(E[Y_i|c \leq X_i < c + \Delta] - E[Y_i|c - \Delta < X_i < c] \right) \\ =& E[Y_{1i} - Y_{0i}|X_i = c] \end{aligned}$$
- τ_{SRD} estimates the LATE for units (in a small neighbourhood) at the cutoff.
- $\tau_{\text{SRD}} = \tau_{\text{ATT}}$ **if and only if** every unit is affected by the treatment in the same way.

A sharp RD estimate of MLDA mortality effects

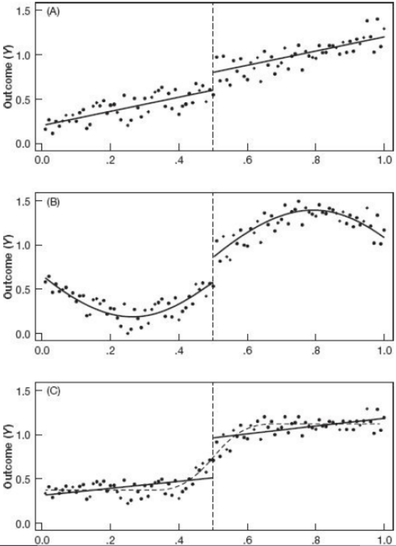
FIGURE 4.2

A sharp RD estimate of MLDA mortality effects



A sharp RD estimate of MLDA mortality effects

FIGURE 4.3
RD in action, three ways



Estimation

- Center the running variable by subtracting the cutoff: $\tilde{X}_i = X_i - c$
- Specify a regression model for $E[Y_i|X_i, D_i]$:
  - Linear, same slope for $E[Y_{0i}|X_i]$ and $E[Y_{1i}|X_i]$:

$$E[Y_i|D_i, X_i] = \alpha + \rho D_i + \gamma \tilde{X}_i$$

  - Linear, but different slope:

$$E[Y_i|D_i, X_i] = \alpha + \rho D_i + \gamma \tilde{X}_i + \delta \tilde{X}_i \times D_i$$

  - Polynomial:

$$E[Y_i|D_i, X_i] = \alpha + \rho D_i + \gamma_1 \tilde{X}_i + \gamma_2 \tilde{X}_i^2 + \delta_1 \tilde{X}_i \times D_i + \delta_2 \tilde{X}_i^2 \times D_i$$

Estimation

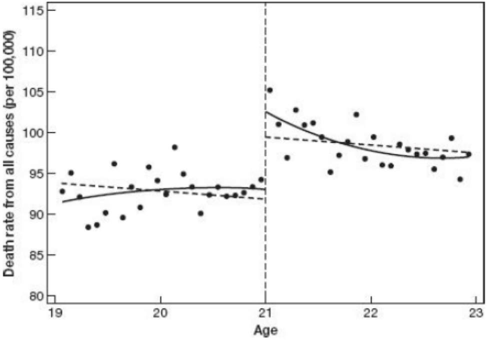
- Produce an RD plot, visualising the discontinuity
- Statistical inference via regression standard errors

Question

What are the associated treatment effects in the second and third cases?

Back to MLDA...

FIGURE 4.4
Quadratic control in an RD design



Back to the MLDA...

- Quadratic specification generates a larger estimate of the MLDA effect at the cutoff than does a linear model.
- The more elaborate model seems to give a better fit than the simple model:
  - Death rates jump sharply at age 21, but then recover somewhat in the first few months after a twenty-first birthday.

Back to MLDA...

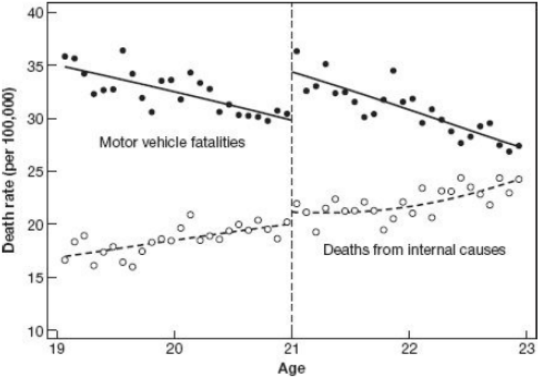
TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22 (1)	Ages 19–22 (2)	Ages 20–21 (3)	Ages 20–21 (4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Back to the MLDA...

FIGURE 4.5

RD estimates of MLDA effects on mortality by cause of death



- There a clear break at the MLDA cutoff, with no evidence of potentially misleading nonlinear trends.
- There isn't much of a jump in deaths due to internal causes.

Nonparametric RD

- Which model specification we should use?
  - Old school questions...
- Alternatively, we can go nonparametrically: estimate the model in a sample such that $c - b \leq X_i \leq c + b$
  - b describes the width of the window and is called a bandwidth.

Nonparametric RD

- The results in Table 4.1 can be seen as nonparametric RD with a bandwidth equal to 2 years of age for the estimates reported in columns (1) and (2) and a bandwidth half as large(that is, including only ages 20–21 instead of 19–22) for the estimates shown in columns (3) and (4).
- Model specification issue should matter little when both are estimated in narrower age windows around the cutoff.

How to choose the bandwidth parameter b?

- Can be technical but also very practically relevant
- The more information available about outcomes in the neighborhood of an RD cutoff, the narrower we can set the bandwidth while still hoping to generate estimates precise enough to be useful.
- Theoretical econometricians have proposed sophisticated strategies for making such bias-variance trade-offs efficiently, though here too, the bandwidth selection algorithm is not completely datadependent and requires researchers to choose certain parameters.
- The goal here is not so much to find the one perfect bandwidth as to show that the **findings generated by any particular choice of bandwidth are not a fluke**.

What to conclude from the MLDA example?

- The RD estimates generated by parametric models with alternative polynomial controls come out similar to one another and close to a corresponding set of nonparametric estimates.
- These nonparametric estimates are largely insensitive to the choice of bandwidth over a wide range.
- This alignment of results suggests the findings generated by an RD analysis of the MLDA capture real causal effects.

Fuzzy RDD: Intro

- In the previous example, treatment status shall change with certainty after the running variable passes the cutoff.
- However, cutoff may not perfectly determine treatment status, but might still create discontinuities in the probability of treatment intake.
- Similarly as in the **IV** setting, cutoff is like a *treatment assignment*, but only some units will comply.
- This implies that we can use such a discontinuity as an IV to produce an estimate of LATE.

Fuzzy RDD estimation

- Local estimator: only consider data above and below the cutoff $\pm b$
- Code the instrument according to $Z_i = \mathbb{1}\{X_i > c\}$
- Fit 2SLS:
$$Y_i = \alpha + \beta \tilde{X}_i + \tau \hat{D}_i + u_i,$$
where $\hat{D}_i$ is fitted value from a first-stage regression of D_i on Z_i and $\tilde{X}_i = X_i - c$.
- Of course, as before, we can have more flexible specification for $\tilde{X}_i$.

The Elite Illusion

- The Boston and New York City public school systems include a handful of selective exam schools.
  - Fewer than half of Boston's exam school applicants win a seat at the John D. O'Bryant School, Boston Latin Academy, or the Boston Latin School (BLS).
  - Only one-sixth of New York applicants are offered a seat at one of the three original exam schools in the Big Apple (Stuyvesant, Bronx Science, and Brooklyn Tech).
- We are interested in the consequences of the exam school treatment.

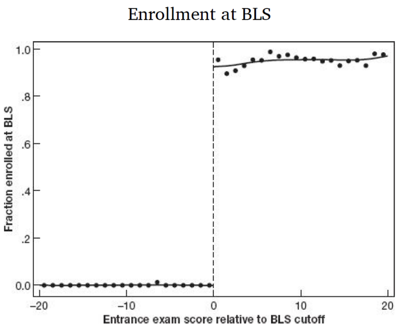
The Elite Illusion

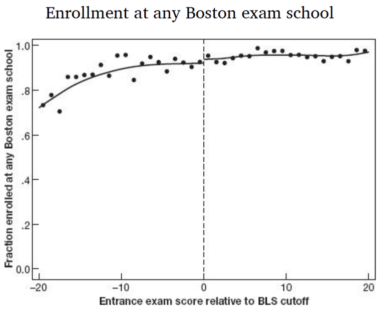
- The case for an exam school advantage is easy to make, but it's also clear that at least some of the achievement difference associated with exam school attendance reflects these schools' selective admissions policies: selection bias.
- Run an experiment and use LATE, as in The Oregon Trail? good idea, but...
- How can we hope to design an experiment that reveals exam school effectiveness?

The Elite Illusion

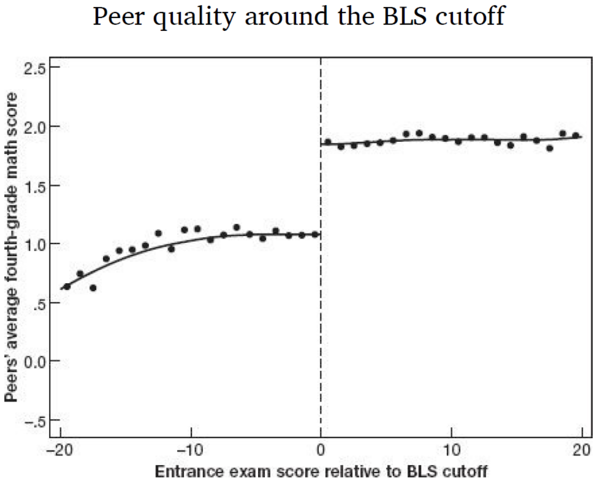
- The discrete nature of exam school admissions policies creates a natural experiment.
- Among applicants with scores close to admissions cutoffs, whether an applicant falls to the right or left of the cutoff might be as good as randomly assigned.
- However, not all students who score above the cutoff attend that school in the end.
- When discontinuities change treatment intensity, the resulting RD design is said to be *fuzzy*.

Fuzzy RDD


Enrollment at BLS
Fraction enrolled at BLS
1.0
.8
.6
.4
.2
0.0
–20
–10
0
10
20
Entrance exam score relative to BLS cutoff


Enrollment at any Boston exam school
Fraction enrolled at any Boston exam school
1.0
.8
.6
.4
.2
0.0
–20
–10
0
10
20
Entrance exam score relative to BLS cutoff

Fuzzy RDD


Peer quality around the BLS cutoff
Peers' average fourth-grade math score
2.5
2.0
1.5
1.0
.5
0.0
−.5
−20
−10
0
10
20
Entrance exam score relative to BLS cutoff

Fuzzy RDD

- Most but not all qualifying applicants enroll at BLS.
- Most students who miss the BLS cutoff indeed end up at another exam school, so that the odds of enrolling at some exam school are virtually unchanged at the BLS cutoff.
- Applicants who qualify for admission at one of Boston's exam schools attend school with much higher-achieving peers than do applicants who just miss the cut, even when the alternative is another exam school.

Fuzzy RDD is IV

- Naive regression approach:

$$Y_i = \theta_0 + \theta_1 \overline{X}_{(i)} + \theta_2 X_i + u_i,$$

where
  - Y_i: student i's seventh-grade math score
  - $X_{(i)}$: student i's fourth-grade math score
  - $\overline{X}_{(i)}$: average fourth-grade math score of i's seventh-grade classmates
- θ_1 estimate is around .25, but it is unlikely to have a causal interpretation for the simple reason that students educated together tend to be similar for many reasons.

Fuzzy RDD is IV

- First-stage:

$$\overline{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i}$$

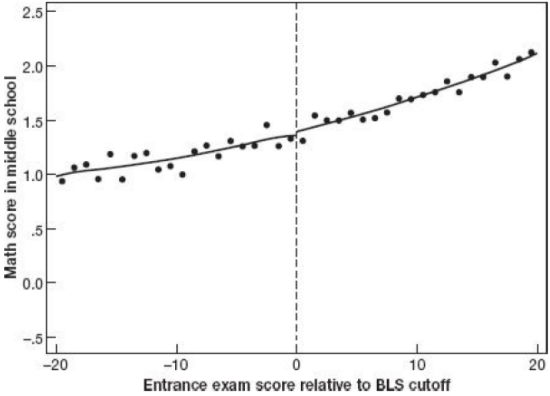
- Second-stage:

$$Y_i = \alpha_2 + \lambda \hat{X}_{(i)} + \beta_2 R_i + e_{2i}$$

- Exclusive restriction: D_i should only affect Y_i through $\overline{X}_{(i)}$.

Fuzzy RDD is IV

FIGURE 4.9
Math scores around the BLS cutoff



Fuzzy RDD is IV

- D_i is not included in the second-stage.
  - We've assumed that in the neighborhood of admissions cutoffs, after adjusting for running variable effects with a linear control, exam school qualification has no direct effect on test scores, but rather influences achievement, if at all, solely through peer quality.
- The 2SLS estimate of λ is -.023 with a standard error of .132.

Fuzzy RDD is IV

- As with any IV story, fuzzy RD requires tough judgments about the causal channels through which instruments affect outcomes.
- The causal journey never ends; new questions emerge continuously.
- But the fuzzy framework that uses RD to generate instruments is no less useful for all that.

Identification with Fuzzy RDD

- As before, introducing some notations
  - y_{1i}, $y_{0i} = \alpha_i$
  - $d_i = f(x_i)$
- So, we can express the observed outcome as

$$y_i = y_{0i} + d_i \cdot (y_{1i} - y_{0i}).$$

- Furthermore, assuming that treatment effects is constant over different individuals:

$$\beta_i = y_{1i} - y_{0i} = \beta$$

Identification with Fuzzy RDD

- Let $e > 0$ be an arbitrarily small number and x_0 be the cutoff value.
- Now we do some algebra

$$E[y_i|x_i = x_0 + e] - E[y_i|x_i = x_0 - e] = \beta\,(\mathrm{RD} + \mathrm{A1})\,, \tag{1}$$

where

$$\begin{aligned}
\mathrm{RD} &= E[d_i|x_i = x_0 + e] - E[d_i|x_i = x_0 - e] \\
\mathrm{A1} &= E[\alpha_i|x_i = x_0 + e] - E[\alpha_i|x_i = x_0 - e].
\end{aligned}$$

Identification with Fuzzy RDD

- Consider the following assumptions:

Assumption (A1)

$E[\alpha_i | x_i = x]$ is continuous in x at x_0.

Assumption (RD)

1. The limits $d^+ = \lim_{x \to x_0^+} E[d_i | x_i = x]$ and $d^- = \lim_{x \to x_0^-} E[d_i | x_i = x]$ exist.
2. $d^+ \neq d^-$.

- Of course, density of x_i has to be positive in the neighbourhood of x_0.

Identification with Fuzzy RDD

- Under Assumption (A1), we have

$$\lim_{x_i \to x_0^+} E[\alpha_i | x_i = x_0] - \lim_{x_i \to x_0^-} E[\alpha_i | x_i = x_0] = 0.$$

- Now, we take limit on both sides of (1) and rearrange terms:

$$\beta = \frac{\lim_{x_i \to x_0^+} E[y_i | x_i = x_0] - \lim_{x_i \to x_0^-} E[y_i | y_i = x_0]}{\lim_{x_i \to x_0^+} E[d_i | x_i = x_0] - \lim_{x_i \to x_0^-} E[d_i | x_i = x_0]},$$

  which is well defined as by Assumption (RD) the denominator is non-zero.
- Under Sharp RDD, $d^+ = 1$ and $d^- = 0$. So Assumption (A1) is enough to achieve identification.