Regression Discontinuity: Extensions

INFO/STSCI/ILRST 3900: Causal Inference

31 Oct 2024

Learning goals for today

At the end of class, you will be able to:

1. Explain the smoothness (continuity) assumption for RDD and why it's important
2. Compare sharp versus fuzzy regression discontinuity
3. Discuss the connection between fuzzy RDD and instrumental variables
4. Explain manipulation and why it poses a problem for regression discontinuity

After today's class, read sections 20.2.2-20.2.4 from Huntington-Klein

The big idea

- Treatment of interest depends only on whether a **running variable** is above or below a **cutoff** c
- We don't require (conditional) exchangeability
- **Smoothness (Continuity)** Assumption: Potential outcomes $\mathrm{E}(Y^a \mid R = r)$ vary smoothly at the cutoff
  The only thing that should change at the cutoff is the treatment.
- Consistency above and below the the cut-off
  $\mathrm{E}(Y^a \mid R = r) = \mathrm{E}(Y \mid R = r, A = a)$
- We can only estimate a local average treatment effect (LATE)
  The average treatment effect for individuals at the cutoff.

The big idea

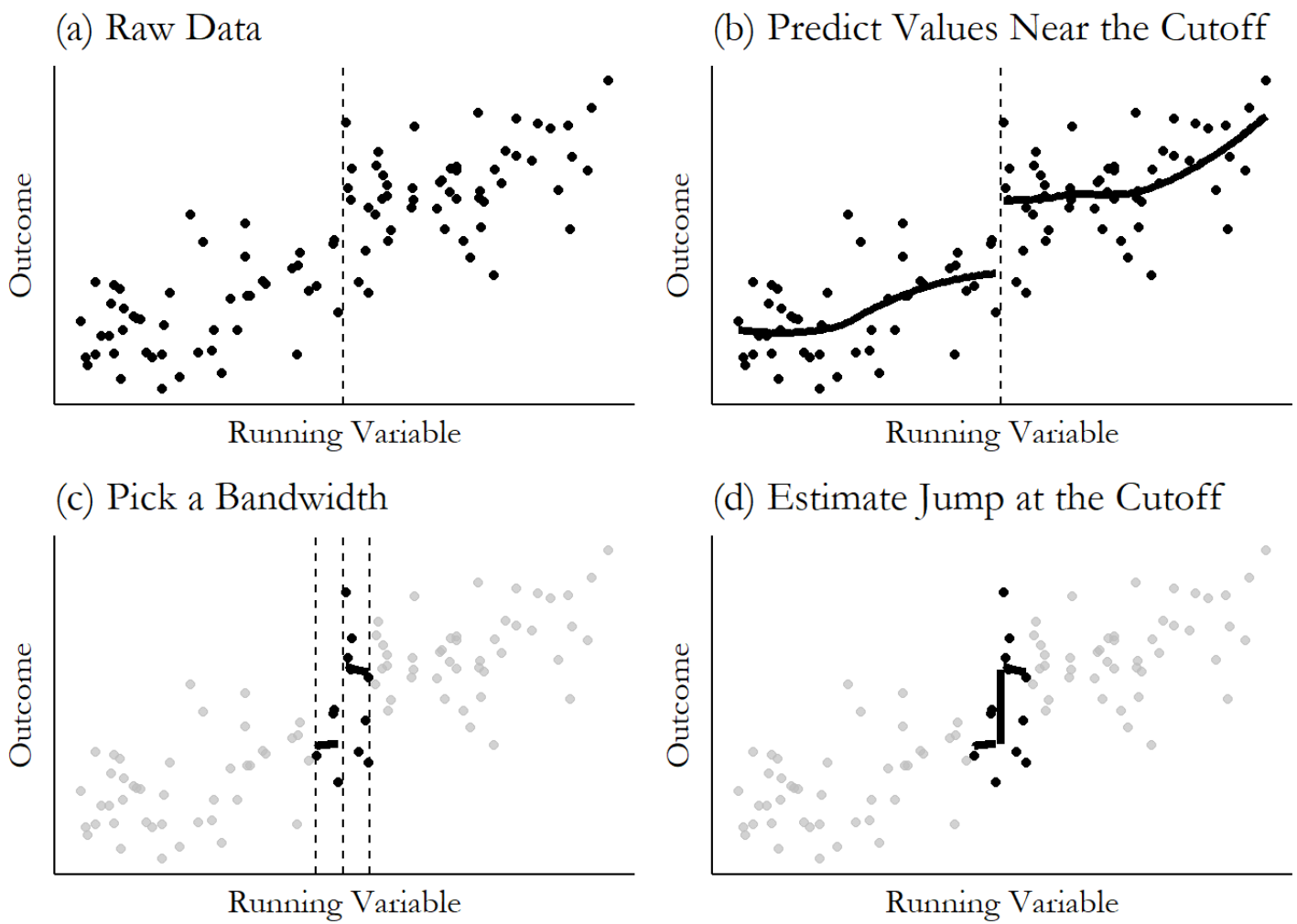


Figure: Huntington-Klein, Nick. The effect: An introduction to research design and causality. Chapman and Hall/CRC, 2021.

The big idea

- Bandwidth selection: How far away from the cutoff are we willing to look?

- Smaller bandwidth: less bias but more variance
- Larger bandwidth: more bias but less variance
- Simpler (less flexible) model: more bias but less variance
- Complex (more flexible) models: less bias, but more variance; more likely to overfit to your data

- In practice:
  - The more data, the smaller the bandwidth
  - Stay away from higher degree polynomials in regression

PollEv: The Continuity Assumption

In which of the following scenarios is the smoothness/continuity assumption violated?

- The only thing that changes at the cutoff is treatment
- The expected potential outcomes are not continuous at the cutoff
- Many factors that affect the potential outcomes change at the cutoff, not just treatment
- The expected potential outcomes are smooth at the cutoff

Join by web:
PollEv.com/causal3900

What can go wrong?

- **Other discontinuity**: Something other than treatment also jumps at the threshold
- **Fuzzy RDD**: Some units are treated on either side of threshold
- **Manipulation**: Units have control over over their running variable

Other discontinuities

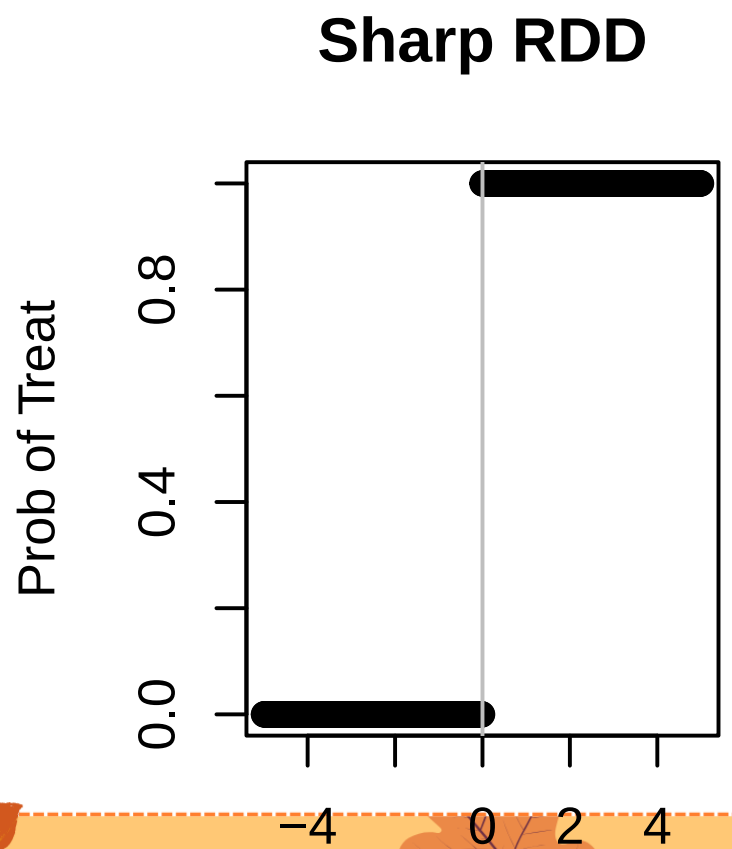
- **Continuity/Smoothness Assumption**: Potential outcomes are smooth (continuous) around the cutoff
  The only thing that changes at the cutoff is treatment
- **Other discontinuity**: If something else "jumps" at the cut-off, then we can't distinguish between effect of treatment and the other thing

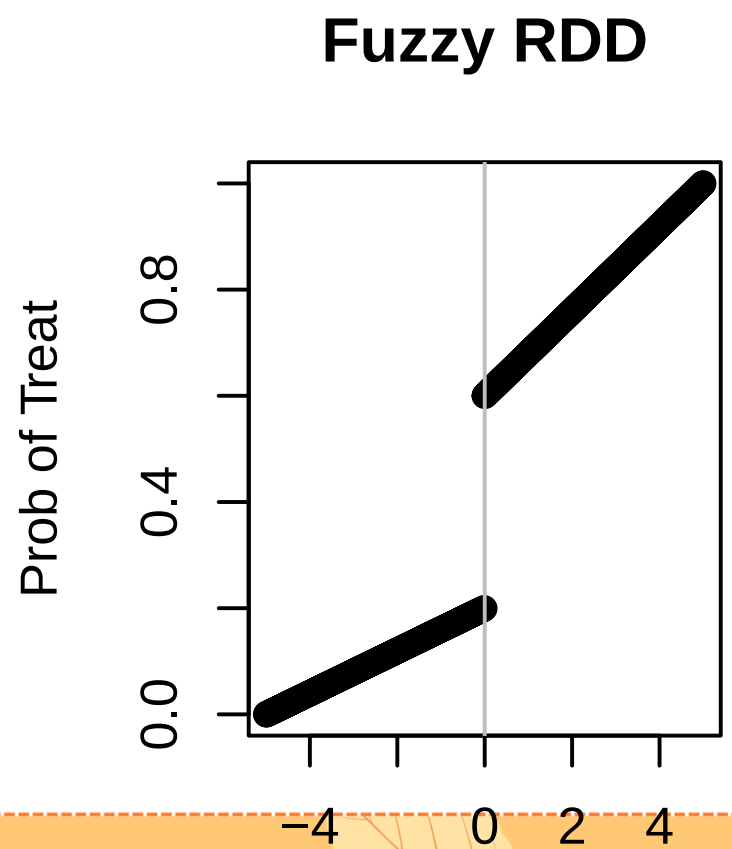
- National Merit example:
  - Suppose students above cut-off also receive tutoring on how to write better personal statements
  - Cannot distinguish between effect of Certificate of Merit and tutoring

- Requires knowledge about problem context
- Can check with placebo tests

Fuzzy RDD

- **Standard setting**: (Sharp RDD) Everyone with running variable above cutoff is treated, everyone with running variable blow cutoff is not treated
- **Fuzzy setting**: Probability of receiving treatment jumps at the cutoff





Fuzzy RDD: ART and retention in care

What is the effect of immediate (vs deferred) anti-retroviral therapy (ART) on retention in care? [1]

- HIV care and treatment program in rural South Africa
- Patients were assigned to immediate versus deferred ART eligibility, as determined by a CD4 count < 350 cells/μl
- **Treatment**: ART (immediate versus deferred);
- **Outcome**: Retention (follow-up) in care
- **Running variable**: CD4 count in blood, **Cutoff**: 350
- Something that is essentially random (being above or below cutoff), encourages treatment uptake... sound familiar?

[1] Bor J, Fox MP, Rosen S, Venkataramani, A, Tanser F, Pillay D, et al. (2017) Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. PLoS Med 14(11): e1002463

Fuzzy RDD

- Around the cut-off, being above/below is like an instrumental variable
- Effect of being above cutoff on outcome

$$\lim_{r \to c^+} \mathrm{E}(Y \mid R = r) - \lim_{r \to c^-} \mathrm{E}(Y \mid R = r)$$

- Effect of being above cutoff on treatment

$$\lim_{r \to c^+} \mathrm{E}(A \mid R = r) - \lim_{r \to c^-} \mathrm{E}(A \mid R = r)$$

- Dividing gives us the **local ATE for compliers**:

$$\frac{\lim_{r \to c^+} \mathrm{E}(Y \mid R = r) - \lim_{r \to c^-} \mathrm{E}(Y \mid R = r)}{\lim_{r \to c^+} \mathrm{E}(A \mid R = r) - \lim_{r \to c^-} \mathrm{E}(A \mid R = r)}$$

Manipulation

- **Standard setting** For people close the cutoff, being above or below the cutoff is essentially random
- **Manipulation**: People choose to be above/below cutoff
- We sometimes have control over our running variable (to some extent)
- Manipulation is a problem when units can choose precisely to be above/below cutoff
- Treatment near the cutoff is no longer "like random"

Manipulation Example: Hiring Discrimination

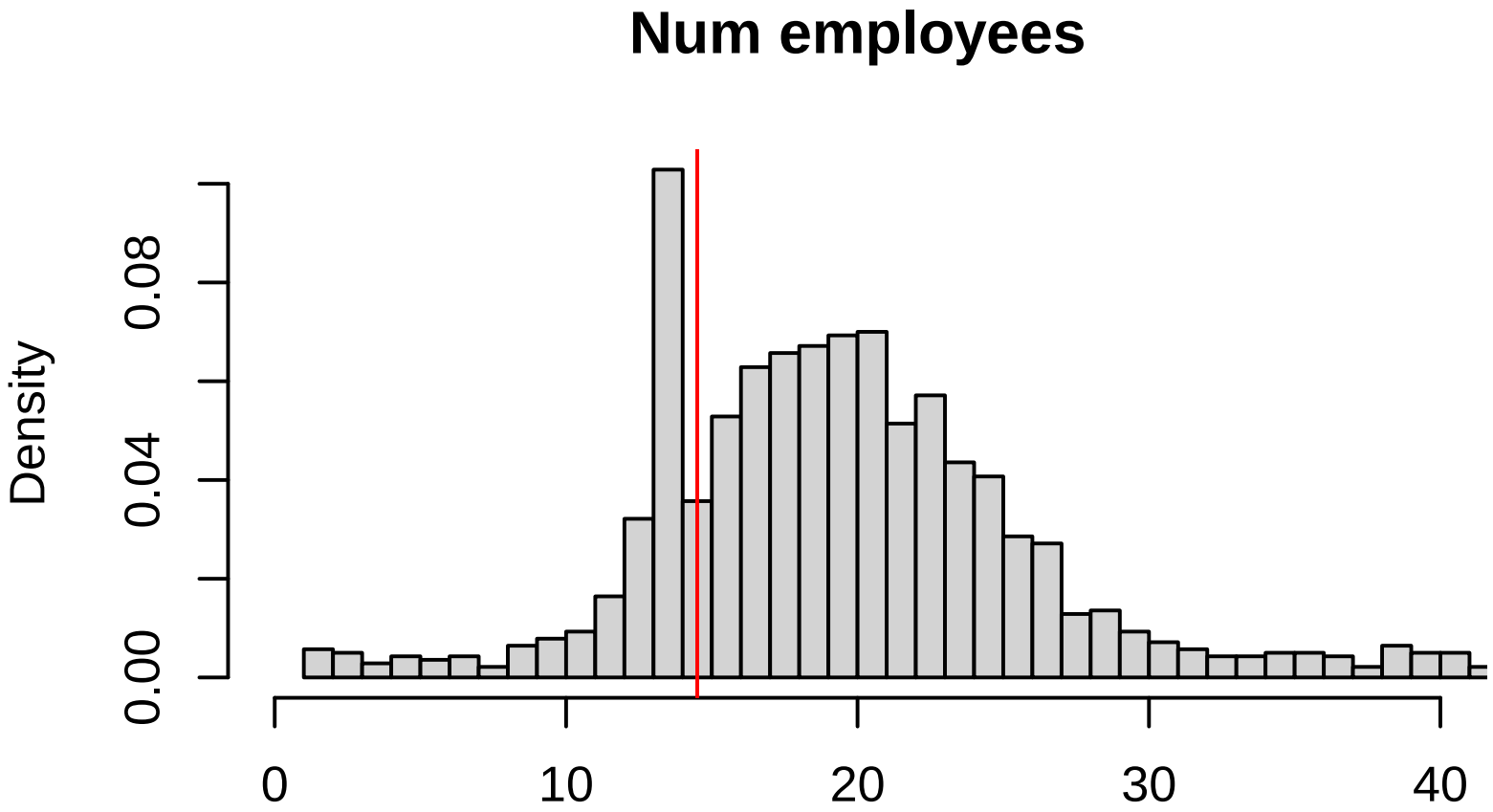
What is the effect of the 1964 civil rights act on hiring discrimination?[2]

- Federal EEOC law prohibits discrimination and applies to firms with 15 or more employees
- Firms with 14 should be essentially the same as firms with 15 employees
- Firms have direct control over how many employees they hire
- Those wanting to avoid EEOC law may decide to stay under 15 employees

[2] Evaluating the effect of an antidiscrimination law using a regression-discontinuity design. Hahn et al. (1999)

Manipulation: what can we do?

- Hard to test for manipulation directly
- Check for balance in covariates
- If manipulation is occurring, we would expect to see "heaping" on one side of the cut-off



Coding Example