TA Session 2
Treatment Effects II: RD and DiD

Microeconometrics II with Joan Llull
IDEA, Fall 2024

TA: Conghan Zheng

November 15, 2024

Overview

1. Regression Discountinuity

2. Differences-in-Differences

Regression Discountinuity

Regression Discountinuity

- Regression continuity (RD) research designs exploit precise knowledge of the rules determining treatment (some rules are arbitrary and therefore provide good experiments).
- RD comes in two styles: fuzzy and sharp.
  - The sharp design can be seen as a selection-on-observables story.
  - The fuzzy design leads to an instrumental variables (IV) type of setup.

Sharp RD

- Sharp RD is used when treatment status is a deterministic and discontinuous function of a **running variable** z_i. Suppose, for example, that

$$D_i = \begin{cases} 1, & \text{if } z_i \geq z_0 \\ 0, & \text{if } z_i < z_0 \end{cases}$$

  where x_0 is a known threshold or cutoff.
- *Deterministic*: once we know z_i, we know D_i.
- *Discontinuous*: no matter how close z_i gets to z_0 (from the left, in this example), treatment is unchanged until $z_i = z_0$.
- Example: a standardized test for college entrance, sharp RD compares the post college performance of students with scores just above and just below the threshold.

Sharp RD

- Consider the following regression that formalizes the RD idea,

$$y_i = f(z_i) + \rho D_i + \eta_i$$

  where ρ is the causal effect of interest, and $D_i = \mathbb{1}(z_i \geq z_0)$.
- As long as the **control function** $f(z_i)$ is continuous in a neighborhood of x_0, it should be possible to estimate this model.

Sharp RD

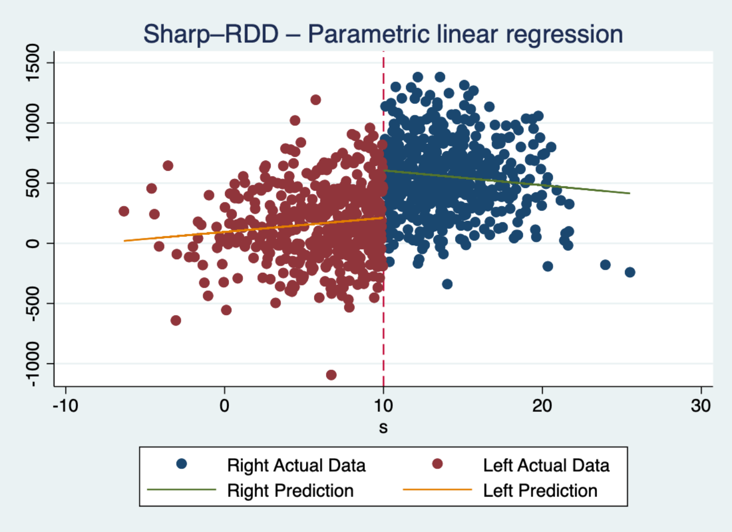
- For example, consider modeling $f(z_i)$ with a pth-order polynomial,

$$y_i = \underbrace{\alpha + \beta_1 z_i + \beta_2 z_i^2 + ... + \beta_p z_i^p}_{=f(z_i)} + \rho D_i + \eta_i$$

- A generalization of RD model allows different trend functions $f_0(z_i)$ for $\mathbb{E}(y_{0i}|z_i)$ and $f_1(z_i)$ for $\mathbb{E}(y_{1i}|z_i)$.
- Calculate the treatment effect:

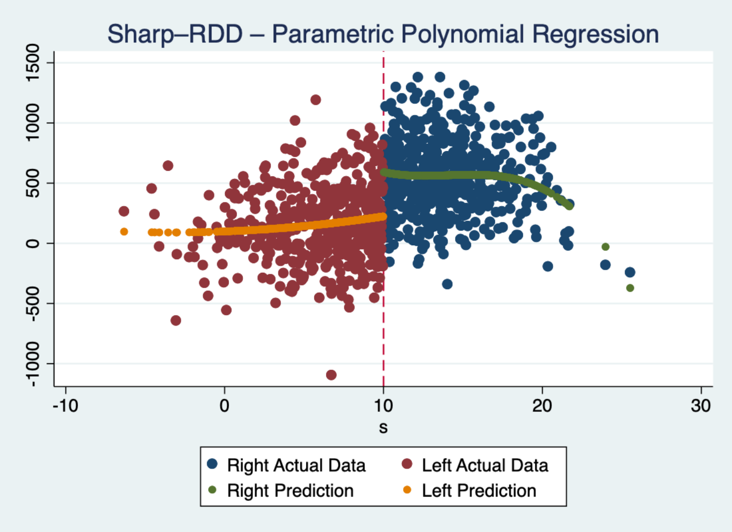
$$\alpha_{ATE,RD} = \underbrace{\lim_{z \to z_0^+} \mathbb{E}[y_i|z_i = z]}_{\simeq \mathbb{E}[y_{1i}|z_i]} - \underbrace{\lim_{z \to z_0^-} \mathbb{E}[y_i|z_i = z]}_{\simeq \mathbb{E}[y_{0i}|z_i]}$$

Sharp RD



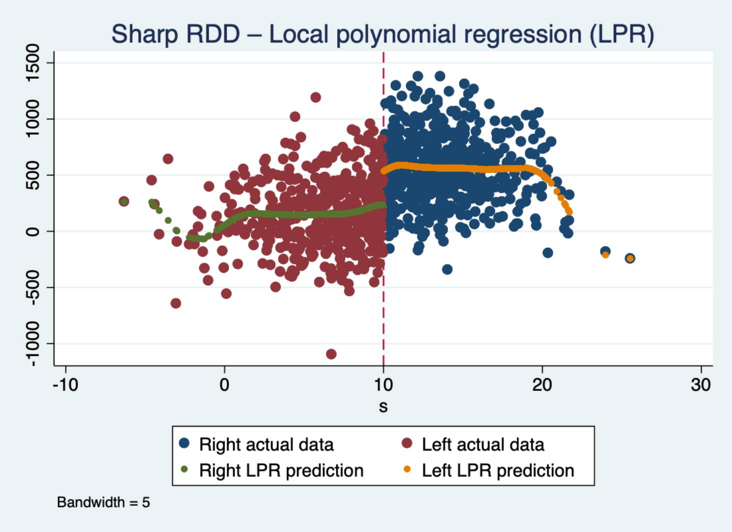
Linear $\mathbb{E}(y_{0i}|X_i)$

Sharp RD



Nonlinear $\mathbb{E}(y_{0i}|X_i)$: third-degree polynomials

Sharp RD



Local polynomials for f_0 and f_1
(you can set different kernels, bandwiths and degrees for them)

Identification

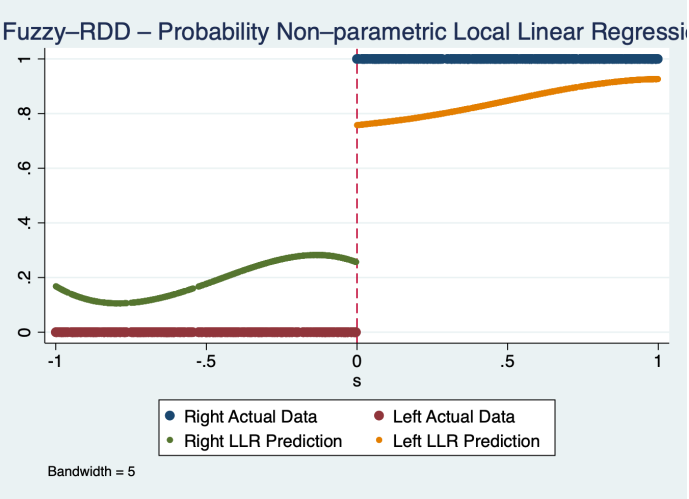
- The regression discontinuity design identifies the conditional ATE at the treatment cut-off.
- What we need for identification is **the continuity of $f_1(z)$ and $f_0(z)$**: which means that the conditional expectation of the untreated and treated outcome are continuously affected by the running variable.

Example

Ursprung and Zigova (2020)

- **Context**: You are interested in studying the effect of an artists death on the price of their artwork.
- **Data**: You have data on the auction sales of a number of renowned artists through their life-time and after their death. Each observation is an artwork sold (e.g. a painting sold at $10 000 when the artist was 35 years old, another sold at $20 000 two years after the same artist's death etc.).
- **The main design**: How would you estimate the effect using regression discontinuity?
  - What is your running variable (z)?
  - Illustrate how your data would look like if the death of an artist causes prices to increase, in particular plot y and D against z.
  - Is this a sharp or fuzzy design?
- **The covariates**:
  - You also have some variables X on the characteristics of the painting (e.g. size, medium, motif etc.). Does it makes sense to include these in the estimation? and how?
  - How do you expect X to behave around the z cut-off if the RD design is valid?

Sharp RD vs. Fuzzy RD



Discontinuity in D

Fuzzy RD

- Now, there is a jump in the probability in the probability of treatment at z_0, such that

$$\mathbb{P}(D_i = 1|z_i) = \begin{cases} g_1(z_i), & \text{if } z_i \geq z_0 \\ g_0(z_i), & \text{if } z_i < z_0 \end{cases}$$

  where $g_1(z_0) \neq g_0(z_0)$. We assume $g_1(z_0) > g_0(z_0)$ so that $z_i \geq z_0$ makes treatment more likely.

- Nonparametric estimation (Kernel, Wald, etc.) of limit $\alpha = \frac{y^+ - y^-}{D^+ - D^-}$: can be applied to both sharp RD and fuzzy RD, identifies treatment effects only locally at the point of discontinuity.

Fuzzy RD

- IV estimation: the discontinuity becomes an instrumental variable for treatment status instead of deterministically switching treatment on or off.
- The ATE at the cutoff:

$$\alpha_{ATE,RD} = \frac{\mathbb{E}(y|z_0^+) - \mathbb{E}(y|z_0^+)}{\pi(z_0^+) - \pi(z_0^-)}$$

In sharp RD, $\pi(z_0^+) - \pi(z_0^-) = 1 - 0 = 1$.

Example

Oreopoulos (2006)

- **Effects of interest**: returns to (compulsory) schooling
- **Context**: UK increased the minimum school leaving age from 14 to 15 in 1947
- **Why fuzzy?** The constraint is only binding for who would have left school at 14 without the change.
- **LATE or ATE**: with about half the students in UK around 1947 leaving school as soon as possible, the LATE from raising the school leaving age should come close to the ATE.
- **Running variable**: z_i, calendar year (47 in data means 1947); $z \geq 47$ (one is aged 14 at or after 1947) fully predicts that the minimum school-leaving age equals 15, and $z < 47$ (one is aged 14 before 1947) fully predicts that the minimum school-leaving age equals 14.
- **Treatment**: whether child attends school at age 15 ($D = 1$) or leaves at age 14 ($D = 0$)

Example

Oreopoulos (2006)

- The treatment variable D_i has conditional density:

$$f(D_i|z_i) = \begin{cases} g_1(z_i), & z_i \geq 47 \\ g_0(z_i), & z_i < 47 \end{cases} \quad , \quad g_1(47) > g_0(47)$$

where $z_i \geq 47$ makes the treatment more likely. Instrument variable:

$$S_i = \begin{cases} 1, & z_i \geq 47 \\ 0, & z_i < 47 \end{cases}$$

- It follows that

$$\begin{aligned} \mathbb{E}(D_i|z_i) &= \int D_i f(D_i|z_i) dD_i \\ &= \int D_i \Big[g_0(z_i) + \Big(g_1(z_i) - g_0(z_i) \Big) \cdot S_i \Big] dD_i \\ &= \mathbb{E}(D_i) \Big[g_0(z_i) + \Big(g_1(z_i) - g_0(z_i) \Big) \cdot S_i \Big] \end{aligned}$$

Fuzzy RD is IV

Oreopoulos (2006)

- We repeat the last line here

$$\mathbb{E}(D_i|z_i) = \mathbb{E}(D_i)\Big[g_0(z_i) + \Big(g_1(z_i) - g_0(z_i)\Big) \cdot S_i\Big]$$

The dummy variable S_i indicates the point of discontinuity in $E(D_i|z_i)$.

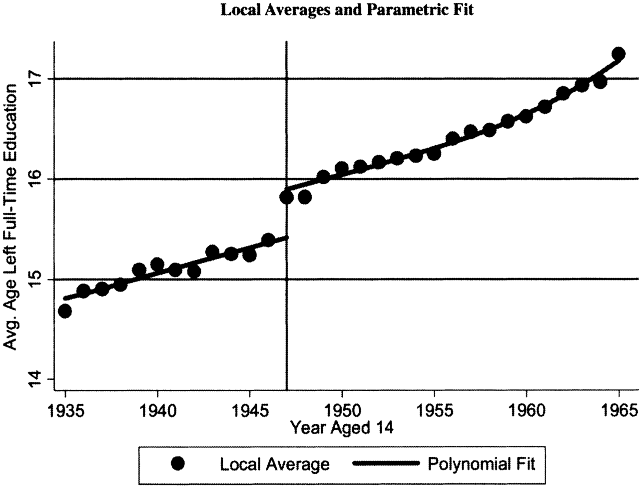
- To capture the non-linearity of the trend, we assume $g_1(z_i)$ and $g_0(z_i)$ each be some reasonably smooth function, for example, a p-th order polynomial:

$$\begin{aligned} E(D_i|z_i) = \mathbb{E}(D_i)\Big[&\beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \cdots + \beta_p z_i^p \\ &+ \Big(\beta_0^* + \beta_1^* z_i + \beta_2^* z_i^2 + \cdots + \beta_p^* z_i^p\Big) \cdot S_i\Big] \end{aligned}$$

- From this (the relevance condition) we see that S_i as well as the interaction terms $\{z_i S_i, z_i^2 S_i, \ldots, z_i^p S_i\}$ can be used as instruments for D_i.

Fuzzy RD

Oreopoulos (2006)



Discontinuity in D

Fuzzy RD is IV

Oreopoulos (2006)

1. The first stage of the 2SLS: discontinuity in D

$$D_i = \tilde{\beta}_0 + \tilde{\beta}_1 z_i + \tilde{\beta}_2 z_i^2 + \cdots + \tilde{\beta}_p z_i^p + \gamma S_i + u_i \tag{1}$$

2. The second stage of the 2SLS: discontinuity in y

Assume $E(Y_{0i}|z_i) = h(z_i)$, where $h(z_i)$ is also a p-th order polynomial of z_i.

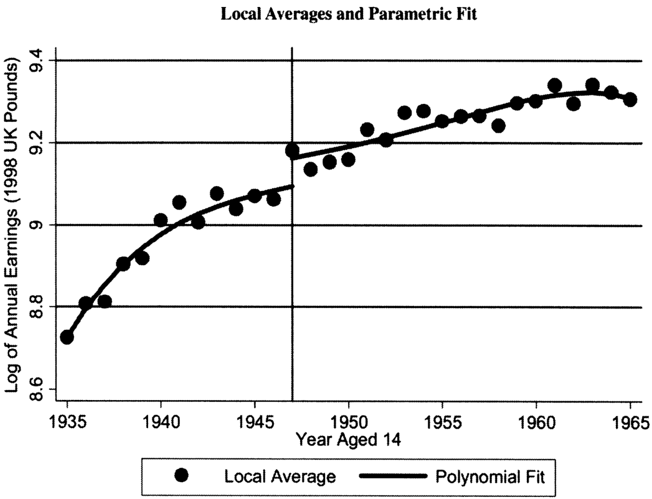
$$\begin{aligned} Y_i &= \alpha_i D_i + h(z_i) + \epsilon_i \\ &= \alpha_i D_i + \rho_0 + \rho_1 z_i + \rho_2 z_i^2 + \cdots + \rho_p z_i^p + \epsilon_i \end{aligned} \tag{2}$$

The fuzzy RD reduced form is obtained by substituting (1) into (2):

$$\begin{aligned} Y_i =& \alpha_i \left[\tilde{\beta}_0 + \tilde{\beta}_1 z_i + \tilde{\beta}_2 z_i^2 + \cdots + \tilde{\beta}_p z_i^p + \gamma S_i + u_i \right] \\ &+ \rho_0 + \rho_1 z_i + \rho_2 z_i^2 + \cdots + \rho_p z_i^p + \epsilon_i \\ =& \theta_0 + \theta_1 z_i + \theta_2 z_i^2 + \cdots + \theta_p z_i^p + \tilde{\epsilon} \end{aligned}$$

Fuzzy RD is IV

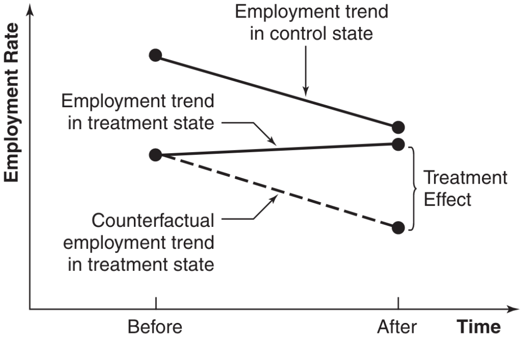
Oreopoulos (2006)



Discontinuity in y

Differences-in-Differences

The Simplest Case: 2×2



- The basic DiD model is a two-way fixed effects model:

$$y_{it} = \alpha D_{it} + X_{it}'\beta + \nu_i + \gamma_t + \varepsilon_{it}$$

Trend specification

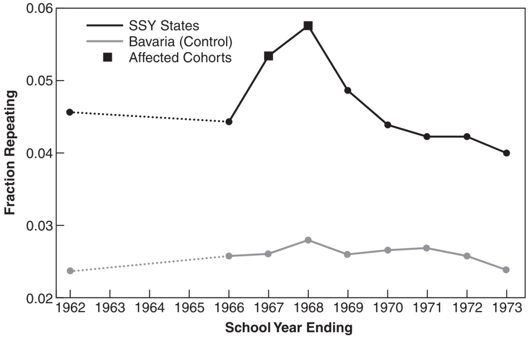


Figure 5.2.3 Average grade repetition rates in second grade for treatment and control schools in Germany (from Pischke, 2007). The data span a period before and after a change in term length for students outside Bavaria (SSY states).

- Message from the graph:
  1. strong visual evidence of treatment and control states with a common underlying trend, and
  2. a treatment effect that induces a sharp but transitory deviation from this trend.

References

- Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge university press. Chapter 25.
- Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press. Chapter 21.
- Cameron, A. C., & Trivedi, P. K. (2022). Microeconometrics using stata (Second Edition). Stata press. Chapters 24, 25.
- Hansen, B. E. (2022). Econometrics. Chapters 18, 21.