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 covariate, x_i. Suppose, for example, that

$$D_i = \begin{cases} 1, & \text{if } x_i \geq x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

  where x_0 is a known threshold or cutoff.
- *Deterministic*: once we know x_i, we know D_i.
- *Discontinuous*: no matter how close x_i gets to x_0 (from the left), threatment is unchanged until $x_i = x_0$.
- 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,

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

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

Sharp RD

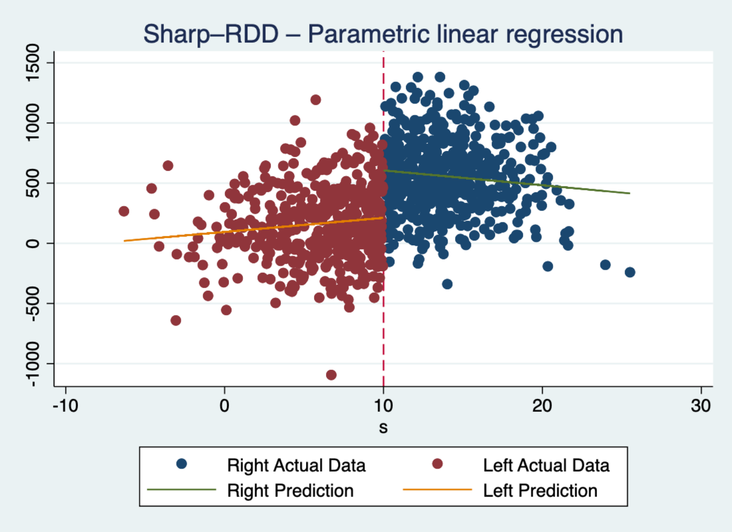
- For example, consider modeling $f(x_i)$ with a pth-order polynomial,

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

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

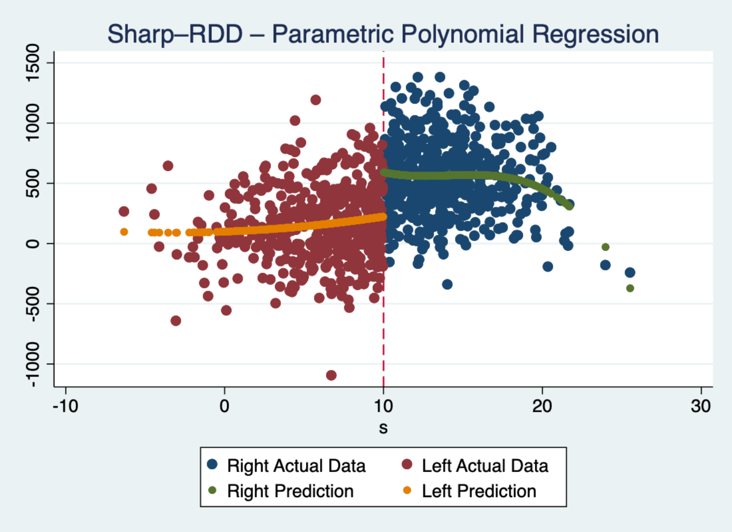
$$\alpha_{ATE,RD} = \underbrace{\lim_{x \to x_0^+} \mathbb{E}[y_i|x_i = x]}_{\simeq \mathbb{E}[y_{1i}|x_i]} - \underbrace{\lim_{x \to x_0^-} \mathbb{E}[y_i|x_i = x]}_{\simeq \mathbb{E}[y_{0i}|x_i]}$$

Sharp RD



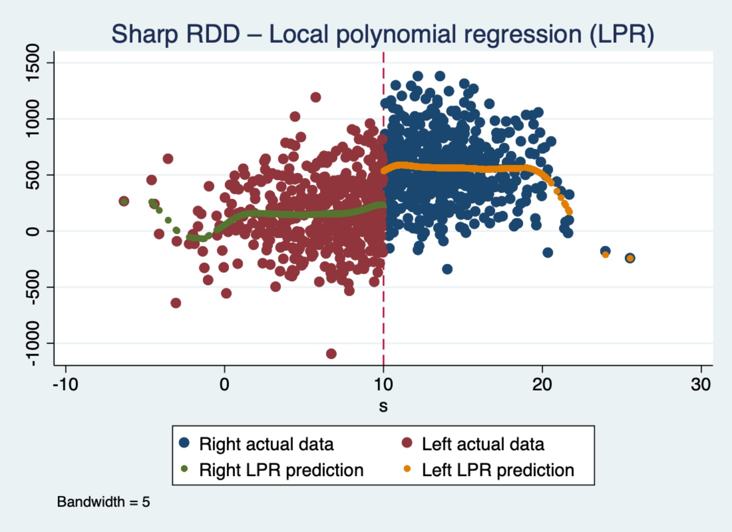
Linear $\mathbb{E}(y_{0i}|X_i)$

Sharp RD



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

Sharp RD

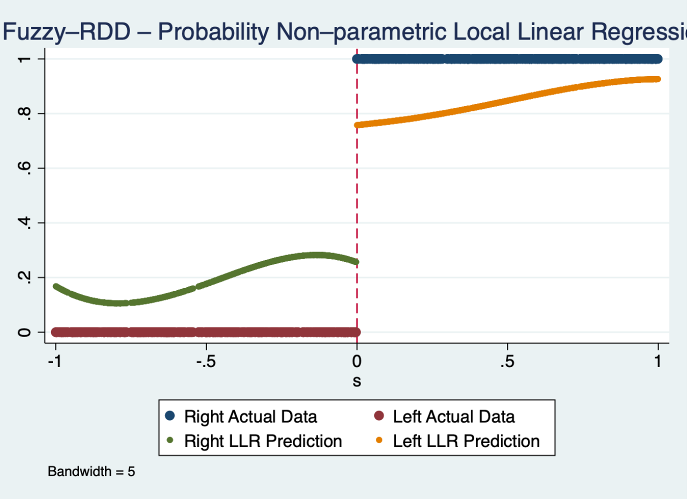


Local polynomials for f_0 and f_1

Identification

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

Sharp RD vs. Fuzzy RD



Discontinuity in D

Fuzzy RD

- Now, there is a jump in the probability in the probability of treatment at x_0, such that

$$\mathbb{P}(D_i = 1|x_i) = \begin{cases} g_1(x_i), & \text{if } x_i \geq x_0 \\ g_0(x_i), & \text{if } x_i < x_0 \end{cases}$$

  where $g_1(x_0) \neq g_0(x_0)$. We assume $g_1(x_0) > g_0(x_0)$ so that $x_i \geq x_0$ makes treatment more likely.
- Fuzzy RD is IV: the discontinuity becomes an instrumental variable for treatment status instead of deterministically switching treatment on or off.
- The ATE at the cutoff:

$$\alpha_{ATE,RD} = \frac{\mathbb{E}(y|x_0^+) - \mathbb{E}(y|x_0^+)}{\pi(x_0^+) - \pi(x_0^-)}$$

  In sharp RD, $\pi(x_0^+) - \pi(x_0^-) = 1 - 0 = 1$.

Fuzzy RD

The Oreopoulos (2006) Example

- The treatment variable D_i has conditional density:

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

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

$$Z_i = \begin{cases} 1, & x_i \geq 47 \\ 0, & x_i < 47 \end{cases}$$

- It follows that

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

Fuzzy RD is IV

The Oreopoulos (2006) Example

- We repeat the last line here

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

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

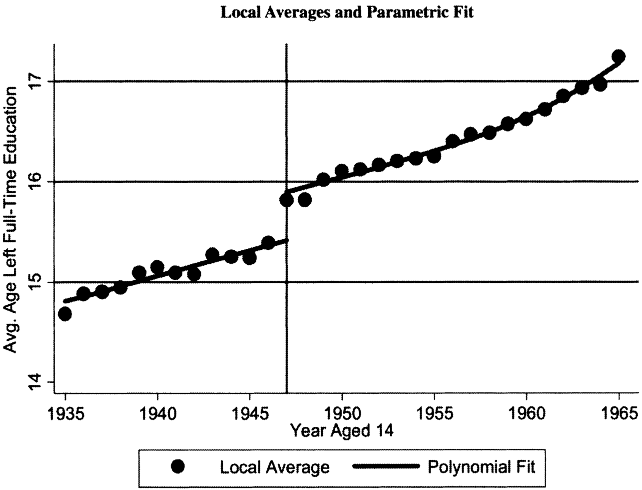
- To capture the non-linearity of the trend, we assume $g_1(x_i)$ and $g_0(x_i)$ each be some resonably smooth function, for example, a p-th order polynomial:

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

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

Fuzzy RD

The Oreopoulos (2006) Example



Discontinuity in D

Fuzzy RD is IV

The Oreopoulos (2006) Example

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

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

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

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

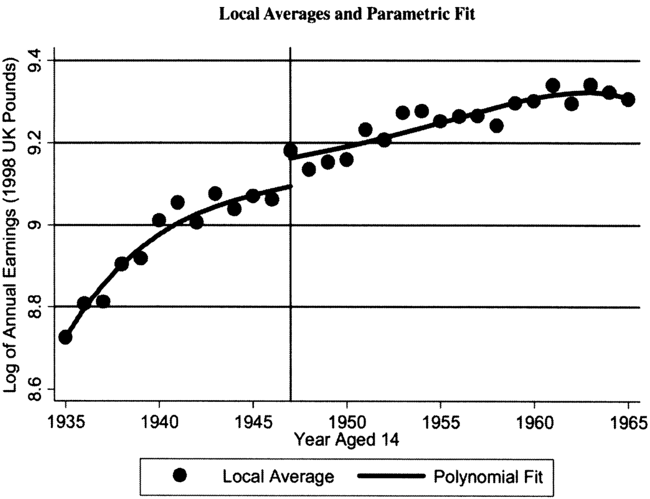
$$\begin{aligned} Y_i &= \alpha_i D_i + h(x_i) + \epsilon_i \\ &= \alpha_i D_i + \rho_0 + \rho_1 x_i + \rho_2 x_i^2 + \cdots + \rho_p x_i^p + \epsilon_i \end{aligned} \tag{2}$$

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

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

Fuzzy RD is IV

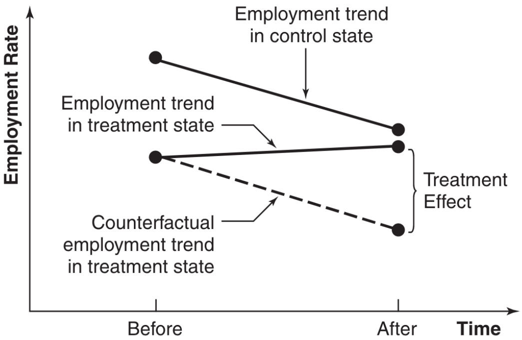
The Oreopoulos (2006) Example



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

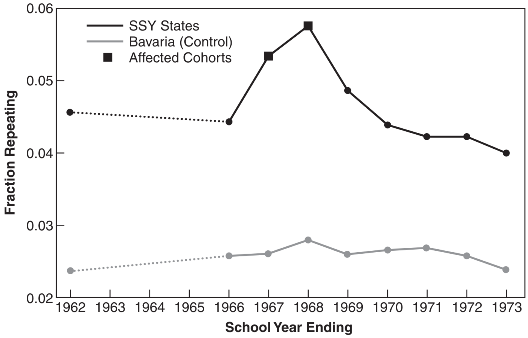


Figure 5.2.3 Average grade repetition rates in second grade for treatment and control schools in Germany (from Pischke, 2007). The data span a period before and after a change in term length for students outside Bavaria (SSY states).

- 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.