Regression Discontinuity

Francis J. DiTraglia

University of Oxford

Treatment Effects: The Basics

Regression Discontinuity (RD) Overview - Cutoff Determines Treatment

Selection Bias

- Problem: people can choose their own treatment.
- Sometimes legal/administrative *cutoffs* completely/partially remove choice.

Sharp RD

- Cutoff *completely* determines treatment: treatment "jumps" from 0 to 1
- Everyone below is untreated; everyone above is treated
- Estimation: compare mean outcomes on either side of cutoff

Fuzzy RD

- Cutoff *partially* determines treatment: probability of treatment jumps at cutoff
- Estimation: Wald estimator on either side of cutoff

Sharp RD Example

American Economic Journal: Applied Economics 2020, 12(3): 207–225
https://doi.org/10.1257/app.20140105

Prestige Matters: Wage Premium and Value Addition in Elite Colleges†

By Sheetal Sekhri*

This paper provides evidence that graduates of elite public institutions in India have an earnings advantage in the labor market even though attending these colleges has no discernible effect on academic outcomes. Admission to the elite public colleges is based on the scores obtained in the Senior Secondary School Examinations. I exploit this feature in a regression discontinuity design. Using administrative data on admission and college test scores and an in-depth survey, I find that the salaries of elite public college graduates are higher at the admission cutoff although the exit test scores are no different. (*JEL* I23, I26, J24, J31, O15)

Sharp RD Example: Sekhri (2020; AEJ Applied)

Research Question

- What is the causal effect of attending an elite college on later-life outcomes?

Background

- The most prestigious colleges in India are public.
- Admissions to public college determined by threshold rule.
- Senior secondary exam score above threshold $\implies$ admitted.

Causal Identification

- Students with scores *just below* the cutoff effectively identical to those *just above*
- Elite college “as if” randomly assigned for students with scores *near* the cutoff.
- Do later-life outcomes “jump” at the admissions cutoff?

Sharp RD Design

Notation

$\mathbf{1}(A)$ is the *indicator* of the event A: equals one if A occurs, zero otherwise.

Assumption: Sharp RD Design

- Binary treatment $D = \mathbf{1}(X \geq c)$
- X is observed: **running variable**
- c is a *known* threshold.

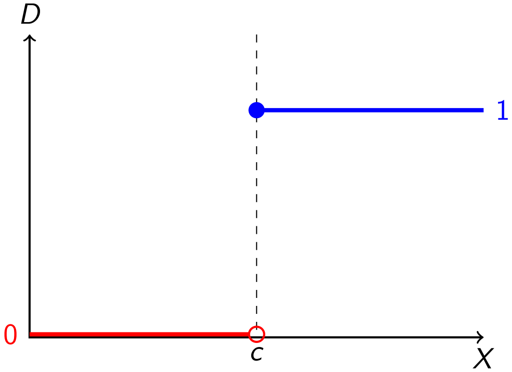


Figure 1: Everyone with $X \geq c$ is treated: $D = 1$. No one with $X < c$ is treated: $D = 0$.

Sharp RD Versus Selection on Observables

Selection on Observables *Holds*

Since $D = \mathbf{1}(X \geq c)$, a function of X

$$\mathbb{E}[Y_1|D, X] = \mathbb{E}[Y_1|\mathbf{1}(X \geq c), X] = \mathbb{E}[Y_1|X]$$

$$\mathbb{E}[Y_0|D, X] = \mathbb{E}[Y_0|\mathbf{1}(X \geq c), X] = \mathbb{E}[Y_0|X]$$

Overlap *Fails*

- For a given value of X either *everyone* is treated or *nobody* is treated!
- Can't compare people with the same X but different D as in propensity score weighting / regression adjustment

What do we need for causal identification?

Intuition

Students whose test scores fall in a *sufficiently small* neighborhood of the admissions cutoff must be "effectively identical."

Identical how?

- Need treatment to be the *only thing that jumps* at the cutoff.
- Potential wages Y_0, Y_1 probably *do depend* on secondary school test scores!
- D jumps at the cutoff so Y jumps from Y_0 to Y_1.
- Rule out a *jump* in the *potential outcomes* Y_0 and Y_1

Formal Assumption: Continuity of Conditional Means

$\mathbb{E}[Y_0|X = x]$ and $\mathbb{E}[Y_1|X = x]$ are both continuous functions of x at the point $x = c$.

How could continuity fail?

Continuity of Conditional Means

$\mathbb{E}[Y_0|X = x]$ and $\mathbb{E}[Y_1|X = x]$ are both continuous functions of x at the point $x = c$.

Sekhri Example

- Assumption fails if students can "precisely manipulate" secondary test scores, e.g.
  - Students can predict whether they will score *just below* c
  - Very diligent students in this situation put in extra effort.
  - This causes diligence to "jump" discontinuously at c.
  - Diligence is related to later-life wage (Y_0, Y_1).
- Only a problem if this generates *another discontinuity* (besides D).
- Unlikely to affect Sekhri (2020): cutoff changes year-to-year/subject-to-subject

Sharp RD Identifies Conditional ATE at $X = c$

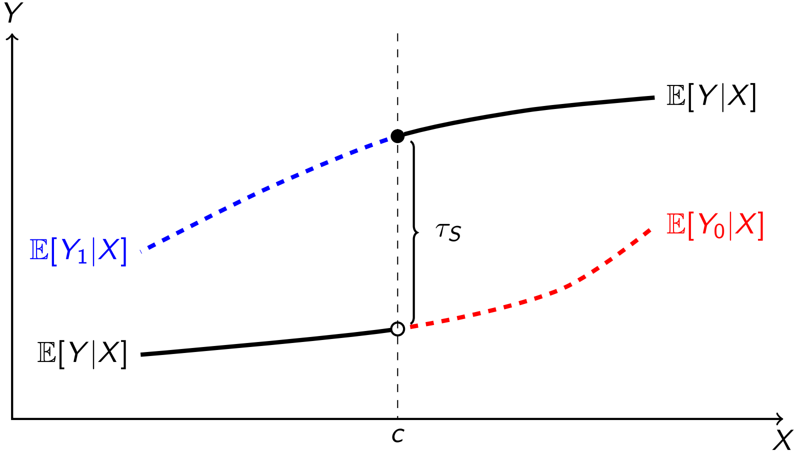


Figure 2: For $X < c$ we observe $\mathbb{E}[Y|X] = \mathbb{E}[Y_0|X]$; for $X \geq c$ we observe $\mathbb{E}[Y|X] = \mathbb{E}[Y_1|X]$. The "jump" in $\mathbb{E}[Y|X]$ at c is the conditional ATE when $X = c$, namely $\tau_S \equiv \mathbb{E}[Y_1 - Y_0|X = c]$.

Proof of Sharp RD Result - It's all in the picture!

Since $D = \mathbf{1}(X \geq c)$:

$$Y = (1 - D)Y_0 + DY_1 = \mathbf{1}(X < c)Y_0 + \mathbf{1}(X \geq c)Y_1.$$

Take conditional expectations:

$$\mathbb{E}[Y|X = x] = \mathbf{1}(x < c)\mathbb{E}[Y_0|X = x] + \mathbf{1}(x \geq c)\mathbb{E}[Y_1|X = x]$$

Take limits as $x \downarrow c$ and $x \uparrow c$:

$$\lim_{x \downarrow c} \mathbb{E}[Y|X = x] = \lim_{x \downarrow c} \{\mathbf{1}(x < c)\mathbb{E}[Y_0|X = x] + \mathbf{1}(x \geq c)\mathbb{E}[Y_1|X = x]\} = \lim_{x \downarrow c} \mathbb{E}[Y_1|X = x]$$

$$\lim_{x \uparrow c} \mathbb{E}[Y|X = x] = \lim_{x \uparrow c} \{\mathbf{1}(x < c)\mathbb{E}[Y_0|X = x] + \mathbf{1}(x \geq c)\mathbb{E}[Y_1|X = x]\} = \lim_{x \uparrow c} \mathbb{E}[Y_0|X = x]$$

Proof of Sharp RD Result - Continued

From Previous Slide:

$$\lim_{x \downarrow c} \mathbb{E}[Y|X = x] = \lim_{x \downarrow c} \mathbb{E}[Y_1|X = x], \quad \lim_{x \uparrow c} \mathbb{E}[Y|X = x] = \lim_{x \uparrow c} \mathbb{E}[Y_0|X = x]$$

Apply Continuity Assumption:

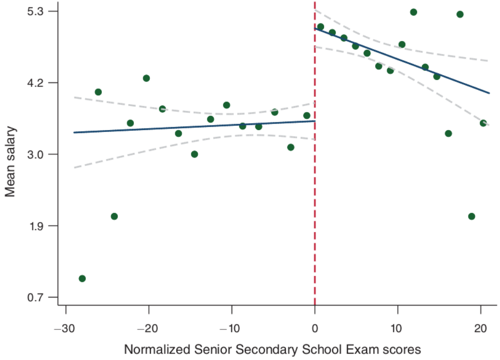
$$\lim_{x \downarrow c} \mathbb{E}[Y_1|X = x] = \mathbb{E}[Y_1|X = c], \quad \lim_{x \uparrow c} \mathbb{E}[Y_0|X = x] = \mathbb{E}[Y_0|X = c]$$

Therefore:

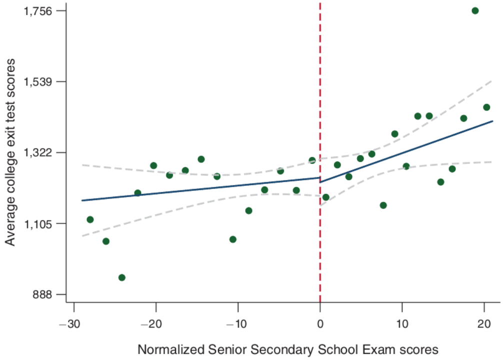
$$\lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x] = \mathbb{E}[Y_1 - Y_0|X = c]$$

Sekhri (2020) - CATE for Students Near the Cutoff

Salary



Exit Test Scores



Estimation and Inference in a Sharp RD Design

Basic Idea

- Only use data from *close* to the cutoff.
- Fit one regression model for $\mathbb{E}[Y|X = x]$ when $x < c$ and another when $x \geq c$
- Implement using treatment dummy $D = \mathbf{1}(X \geq c)$ *interacted* with X.

Some Subtleties

- Bias variance tradeoff: how close is close?
  - Ideally use only observations *really close* to cutoff $\implies$ low bias
  - But there are very few such observations $\implies$ high variance
- Allow $\mathbb{E}[Y|X = x]$ to be nonlinear to avoid confusing a "jump" with a "bend"
- Don't use high-order polynomials: Imbens & Gelman (2019; JBES)
- Implementation Details, Non-parametrics: Cattaneo, Idrobo & Titiunik (2019)
- Classic review article on RD: Lee & Lemieux (2010; JEL)

Simple Sharp RD Example

Separate Regressions

$$Y_i = \begin{cases} \alpha_0 + \alpha_1 X_i + \epsilon_i, & \text{for } X_i < c \\ \beta_0 + \beta_1 X_i + \epsilon_i, & \text{for } X_i \geq c \end{cases}$$

$$\implies \mathbb{E}[Y_1 - Y_0 | X = c] = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)c$$

Single Regression with Interaction

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \gamma_3 D_i X_i + \epsilon_i$$

$$D_i = \mathbf{1}(X_i \geq c)$$

$$\gamma_0 = \alpha_0, \quad \gamma_1 = (\beta_0 - \alpha_0)$$

$$\gamma_2 = \alpha_1, \quad \gamma_3 = (\beta_1 - \alpha_1)$$

$$\implies \mathbb{E}[Y_1 - Y_0 | X = c] = \gamma_1 + \gamma_3 c.$$

A Simpler Approach: Re-define the Running Variable

Let $\widetilde{X} \equiv (X_i - c)$

$$Y_i = \delta_0 + \delta_1 D_i + \delta_2 \widetilde{X}_i + \delta_3 D_i \widetilde{X}_i + \epsilon_i \implies \mathbb{E}[Y_1 - Y_0 | X = c] = \delta_1$$

Why does this work?

- Modified running variable $\widetilde{X}_i$ has a cutoff $\widetilde{c} = 0$ regardless of the value of c
- $D_i = \mathbf{1}(X_i \geq c) = \mathbf{1}(X_i - c \geq 0) = \mathbf{1}(\widetilde{X}_i \geq 0)$
- Read off result from previous slide, setting $c = 0$.

Sharp RD Example in R - Parameters

```r
# Parameters of the "separate" regressions
a0 <- 0.3
a1 <- 0.2
b0 <- 0.8
b1 <- -0.3

# Implied parameters of the "joint" regression
g0 <- a0
g1 <- b0 - a0
g2 <- a1
g3 <- b1 - a1
```

Sharp RD Example in R - Simulation Draws

```r
# Simulation draws
set.seed(1234)
n <- 500
x <- runif(n)
cutoff <- 0.5
D <- 1 * (x > cutoff)
epsilon <- rnorm(n, sd = 0.1)
y <- g0 + g1 * D + g2 * x + g3 * D * x + epsilon
```

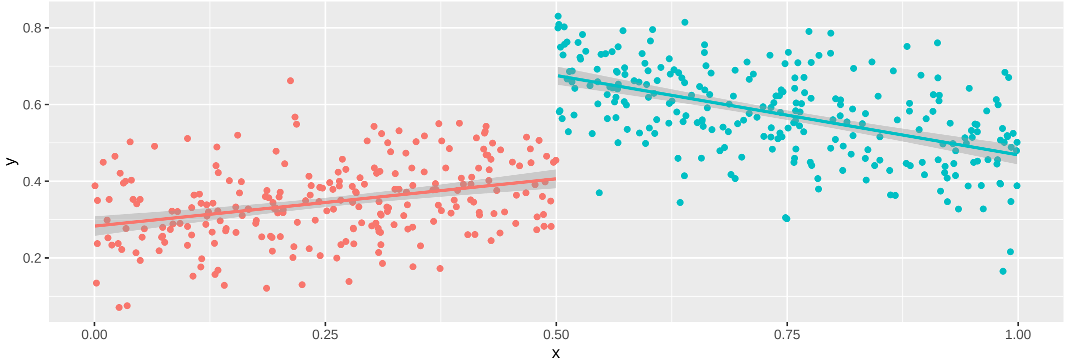
Sharp RD Example in R - Estimation and Inference

```r
# Fit linear regression model, centering X around the cutoff
xtilde <- x - cutoff
rd <- lm(y ~ D * xtilde)
library(broom)
tidy(rd)
```

```
## # A tibble: 4 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)    0.406    0.0129     31.5  1.33e-120
## 2 D              0.269    0.0174     15.5  2.35e- 44
## 3 xtilde         0.247    0.0458      5.39 1.12e-  7
## 4 D:xtilde      -0.660    0.0615    -10.7  2.61e- 24
```

Sharp RD Example in R - Plotting the Results

```
library(tidyverse)
ggplot(data.frame(x = x, y = y), aes(x, y, color = factor(D))) +
  geom_point() +
  geom_smooth(method = 'lm', formula = y ~ x) +
  theme(legend.position = 'none') # Get rid of the legend!
```



Confusing a "Bend" with a "Jump" - Simulation Design

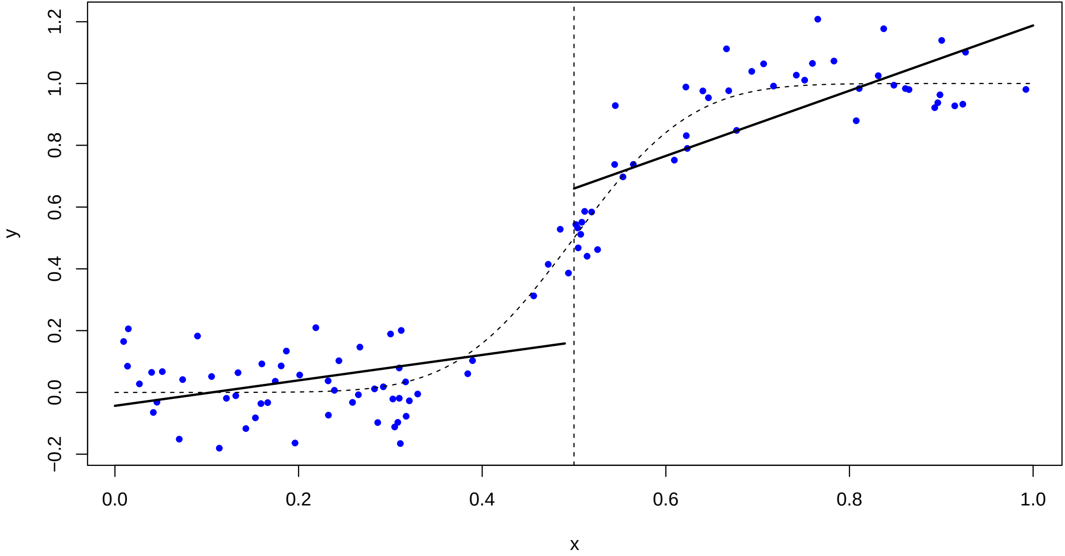
```r
# Nonlinear simulation design: no discontinuity!
set.seed(1234)
n <- 100
x <- runif(n)
y <- pnorm(x, 0.5, 0.1) + rnorm(n, sd = 0.1)
D <- 1 * (x >= 0.5)
```

Confusing a “Bend” with a “Jump” - Linear RD Estimates

```r
# Linear RD results
xtilde <- x - 0.5
rd1 <- lm(y ~ D + xtilde + xtilde:D)
tidy(rd1)
```

```
## # A tibble: 4 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    0.162    0.0455      3.57 5.63e- 4
## 2 D              0.498    0.0575      8.66 1.11e-13
## 3 xtilde         0.412    0.148       2.78 6.48e- 3
## 4 D:xtilde       0.644    0.203       3.17 2.02e- 3
```

Confusing a "Bend" with a "Jump" - Plot of Linear RD



Confusing a “Bend” with a “Jump” - Quadratic RD Isn’t Fooled!

```r
# Quadratic RD results
rd2 <- lm(y ~ (xtilde + I(xtilde^2)) * D)
tidy(rd2)
```

```
## # A tibble: 6 x 5
##   term           estimate std.error statistic  p.value
##   <chr>             <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)      0.423     0.0525      8.06 2.37e-12
## 2 xtilde           2.89      0.405       7.15 1.87e-10
## 3 I(xtilde^2)      4.60      0.725       6.35 7.45e- 9
## 4 D                0.0984    0.0614      1.60 1.12e- 1
## 5 xtilde:D         0.658     0.543       1.21 2.29e- 1
## 6 I(xtilde^2):D  -10.4       1.09       -9.57 1.48e-15
```

Fuzzy RD Example: Jacob & Lefgren (2004; ReStat)[1]

Research Question

Causal effect of remedial education: *summer school* and *repeating a grade*.

Background

- Chicago Public Schools (CPS) used to practice *social promotion*: students advance to next school grade regardless of academic performance.
- Policy change in 1996–1997: achievement thresholds in 3rd, 6th & 8th grades.
- End to social promotion: have to pass math and reading tests to advance.
- Pass in June $\implies$ automatically advance.
- Fail in June $\implies$ attend summer school and re-test in August.
- Fail in August $\implies$ repeat a grade.

[1] Remedial education and student achievement: A regression-discontinuity analysis

Fuzzy RD Example: Remedial Education

Why is this an RD design?

Administrative threshold determines treatment: summer school / repeating a grade.

Why isn't is a *sharp* RD?

Administrative threshold only *imperfectly* determined treatment:

- 3% of students who failed in June were exempted from Summer School.
- 14% of students who failed August re-test were exempted from repeating a grade
- Some students were required to repeat a grade despite passing in June.

Fuzzy RD Design

Probability of treatment "jumps" at the cutoff c, but not from zero to one:

$$\lim_{x \downarrow c} \mathbb{P}(D = 1|X = x) \neq \lim_{x \uparrow c} \mathbb{P}(D = 1|X = x)$$

Fuzzy RD Estimand

$$\tau_F \equiv \frac{\lim_{x\downarrow c} \mathbb{E}[Y|X=x] - \lim_{x\uparrow c} \mathbb{E}[Y|X=x]}{\lim_{x\downarrow c} \mathbb{E}[D|X=x] - \lim_{x\uparrow c} \mathbb{E}[D|X=x]}$$

Fuzzy RD Design

Since D is binary, re-write as: $\lim_{x\downarrow c} \mathbb{E}[D|X=x] \neq \lim_{x\uparrow c} \mathbb{E}[D|X=x]$

Intuition

- Numerator of τ_F is the sharp RD estimand.
- Fuzzy RD: some people below threshold are untreated; some above are treated.[2]
- $\lim_{x\downarrow c} \mathbb{E}[Y|X=x]$ and $\lim_{x\uparrow c} \mathbb{E}[Y|X=x]$ both contain a "mix" of Y_1 and Y_0.
- Similar to "reduced form" regression in IV, so divide by "first stage."

[2] In Jacob & Lefgren everything is reversed: just flip the running variable.