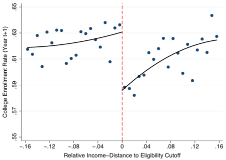
Session VII
TA Session: Using regression jumps and kinks for causal evaluation

Evaluating public policies

Arthur Heim (PSE & Cnaf)

(b) Enrolled in College in Following Year



Outline

Introduction

What we have seen so far

- When randomisation is not available we have seen how we can recover causal estimates at the cost of stronger identifying assumptions.
- With DID we may recover causal effects if we assume (conditional) parallel trends between treated and untreated units
- With IV, we may identify local average treatment effect (LATE) under independence and ignorability of the instrument, when there is a monotone relationship between the instrument and treatment.
- We have seen how regressions can estimate conditional expectation functions
- Here, we consider setting where treatment depends on the value of a continuous covariate R where there exist a cutoff c where those below are untreated and treated above (or the opposite)
- **Intuition** Estimate conditional expectation functions on each side of the cutoff ; the jump between the two is **a causal effect**.

Introduction

A jump in treatment probability

- In some situations, eligibility for a policy directly depends on continuous variables where there exist some sort of threshold that partition groups into treated (or eligible) and untreated (uneligible)
- **Examples:**

Introduction

A jump in treatment probability

- In some situations, eligibility for a policy directly depends on continuous variables where there exist some sort of threshold that partition groups into treated (or eligible) and untreated (uneligible)
- **Examples:**
  - You need to be 25 Y-O to receive RSA in France ;

Introduction

A jump in treatment probability

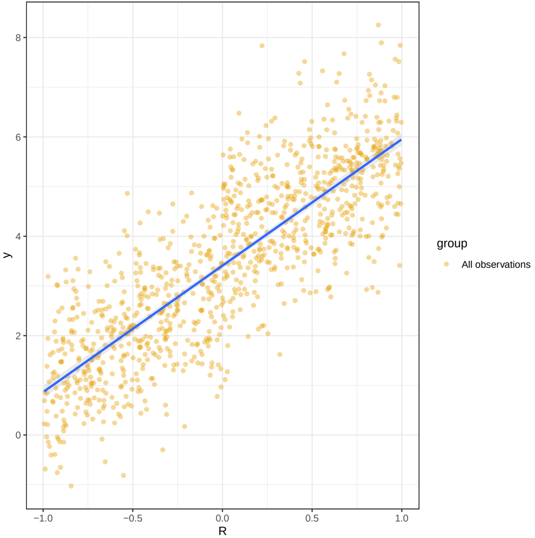
- In some situations, eligibility for a policy directly depends on continuous variables where there exist some sort of threshold that partition groups into treated (or eligible) and untreated (uneligible)
- **Examples:**
  - You need to be 25 Y-O to receive RSA in France ;
  - In some schooling systems, failing an exam or a mark makes you go to summer schools or repeat a grade
  - Some financial aids are mean-tested with thresholds defining the amount of aid.
  - Some firms may have tax exemptions based on gross income, number of employees,...
  - the .5 threshold in local election results assign left/right policies (?)
  - Many settings actually...
- Regression discontinuity design will take advantage of these thresholds to identify causal effects.

Visualisation

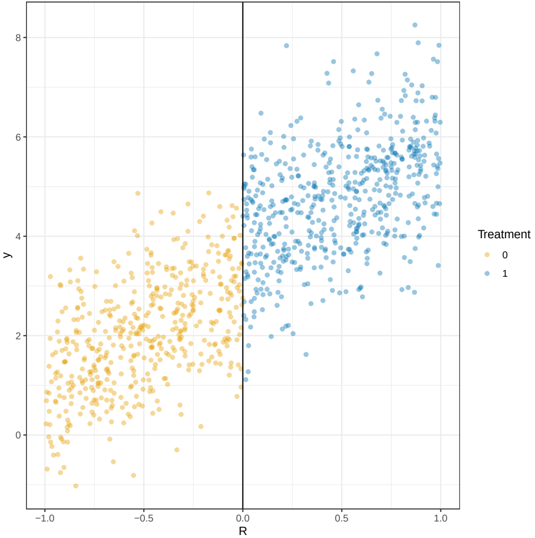
Generating data in R

```
# Generate data
set.seed(378)
# Generate running variable from uniform distribution
R <- runif(1000, -1, 1)
# Outcome, simple slope with a jump of .8 (not too obvious)
y <- 3 + 2 * R + 0.8 * (R >= 0) + rnorm(1000)
t <- ifelse(R > 0, 1, 0)
# quadratic function for wrong RDD
y2 <- 3 + -4 * R^2 + 1 + rnorm(1000)
t2 <- ifelse(R > -0.5, 1, 0)
datab <- data.frame(R, y, y2, t, t2)
```

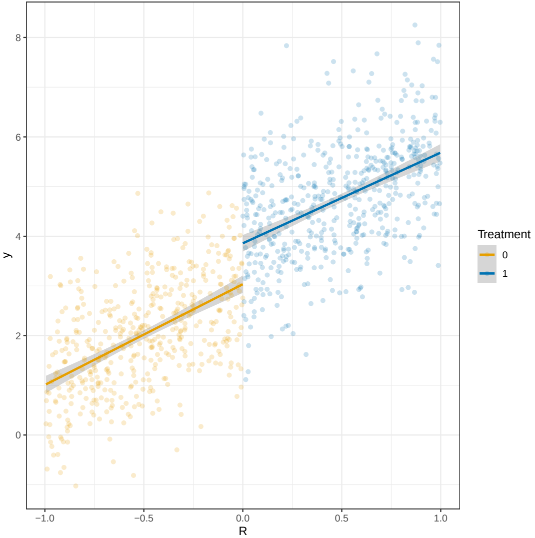
Visualisation



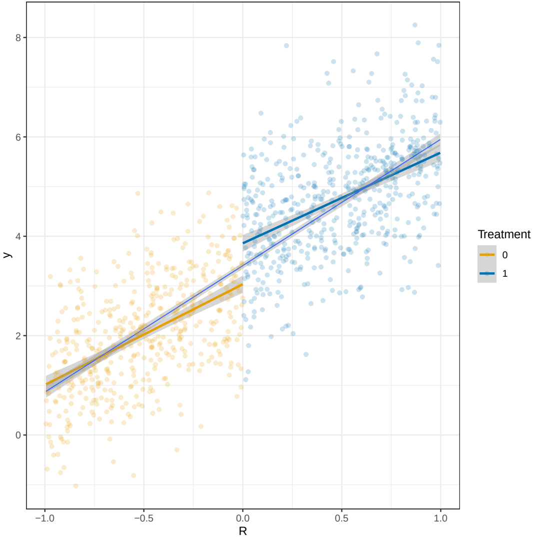
Visualisation



Visualisation



Visualisation



Introduction

Visualisation

- Regression discontinuity designs are well suited for such setting
- We need a "*forcing variable*" or "*running variable*" with a threshold or (*cutoff*) that induces a jump in treatment probability

Introduction

Visualisation

- Regression discontinuity designs are well suited for such setting
- We need a "*forcing variable*" or "*running variable*" with a threshold or (*cutoff*) that induces a jump in treatment probability
- We talk about (*Sharp design*) when probability jumps from 0 to 1, everybody is treated on one side, nobody is on the other.
    - Fack and Grenet (2015) use a sharp design to assess the effect of college grants in France

Introduction

Visualisation

- Regression discontinuity designs are well suited for such setting
- We need a "*forcing variable*" or "*running variable*" with a threshold or (*cutoff*) that induces a jump in treatment probability
- We talk about (*Sharp design*) when probability jumps from 0 to 1, everybody is treated on one side, nobody is on the other.
  - Fack and Grenet (2015) use a sharp design to assess the effect of college grants in France
- We talk about (*Fuzzy design*) when there's only a jump in probability (higher/lower share of treated on one side).
  - Abdulkadiroglu, Angrist, and Pathak (2014) use a fuzzy design to estimate the effect of charter schools in New-York and Boston on student achievement based on admission ranks from a centralized algorithm.

Introduction

Why would that work ?

Introduction

Why would that work ?

1. Simple idea: assignment mechanism is completely known ; we know that the probability of treatment jumps to 1 if forcing variable $> c$
   - Assumption is that individuals cannot manipulate with precision their assignment variable (think about the SAT test for instance)
   - Key word: **precision**. Consequence: very similar individuals near the cutoff

Introduction

Why would that work ?

1. Simple idea: assignment mechanism is completely known ; we know that the probability of treatment jumps to 1 if forcing variable $> c$
   - Assumption is that individuals cannot manipulate with precision their assignment variable (think about the SAT test for instance)
   - Key word: **precision**. Consequence: very similar individuals near the cutoff
2. The thresholds creates a local randomized experiment where people get treated or not for just a tiny change of the running variable
   - If treated and untreated individuals are similar near the cutoff point then data can be analyzed as if it were a (conditionally) randomized experiment
   - If this is true, then background characteristics should be similar near c (can be checked empirically)
   - The estimated treatment effect applies to those near the cutoff point (external validity)

Chaplin et al. (2018) compared the results of randomized experiments with regression discontinuity analysis of the same data and showed both yields very similar, unbiased, results

Why would that work ?

Figure 5: When we model the selection process (from Imbens and Lemieux (2008))

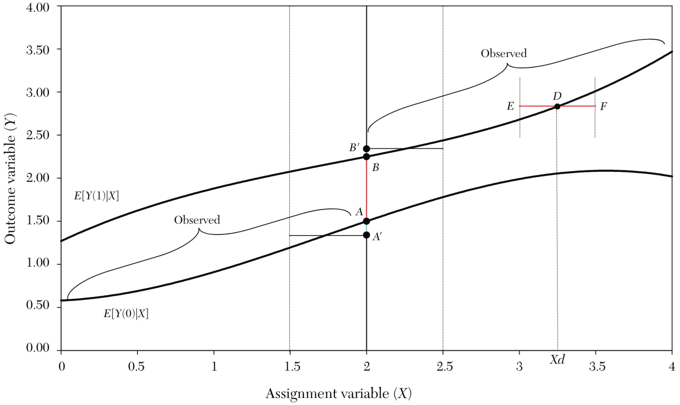


Figure 2. Nonlinear RD

Why would that work ?

Figure 6: RDD as local RCT (from Imbens and Lemieux (2008))

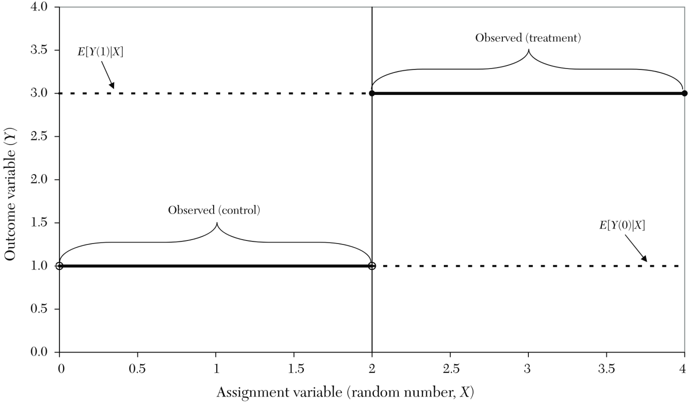


Figure 3. Randomized Experiment as a RD Design

Introduction

Very useful papers for RDD

- Imbens and Lemieux 2008
- David S. Lee and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48, no. 2 (June): 281–355
- These provide great overview but there's been a lot of development since. Latest review:
- Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations.* 1st ed. Cambridge University Press, November 30, 2019. ISBN: 978-1-108-68460-6 978-1-108-71020-6
- Their rdpackages.github webpage with all the packages and papers

Outline

Regression discontinuity: formal introduction

RCT vs sharp RDD

- Notation: $(Y_i(0), Y_i(1), X_i)\,, i = 1, 2, \ldots, n$
- Treatment: $T_i \in \{0, 1\}$, $\quad T_i$ independent of $(Y_i(0), Y_i(1), X_i)$.
- Data: $(Y_i, T_i, X_i)\,, i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- In an RCT, the Average Treatment Effect is:

$$\tau_{\text{ATE}} = \mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \mathbb{E}\left[Y_i \mid T = 1\right] - \mathbb{E}\left[Y_i \mid T = 0\right]$$

Regression discontinuity: formal introduction

RCT vs sharp RDD

- Notation: $(Y_i(0), Y_i(1), X_i)\,, i = 1, 2, \ldots, n$
- Treatment: $T_i \in \{0, 1\}$, $\quad T_i$ independent of $(Y_i(0), Y_i(1), X_i)$.
- Data: $(Y_i, T_i, X_i)\,, i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- With RDD, we estimate Average Treatment Effect at the cutoff (Continuity-based):

$$\begin{aligned} \tau_{\text{SRD}} &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c\right] \\ &= \lim_{x \downarrow c} \mathbb{E}\left[Y_i \mid X_i = x\right] - \lim_{x \uparrow c} \mathbb{E}\left[Y_i \mid X_i = x\right] \end{aligned}$$

- In words, we use the limit of the conditional mean around the cutoff to estimate the treatment effect there.

Regression discontinuity: formal introduction

The regressions for sharp RD

- Treatment status D_i is a deterministic function of the runing variable r_i with a discontinuity at r_0.

$$D_i = \begin{cases} 1 & if\ r_i \geq r_0 \\ 0 & if\ r_i < r_0 \end{cases}$$

- First model can be to assume a constant effect such that

$$\mathbb{E}[Y_i(0)|r_i] = \alpha + f(r_i)$$
$$Y_i(1) = Y_i(0) + \rho$$

- The corresponding regression is:

$$Y_i = \alpha + f(r_i) + \rho D_i + \eta_i$$

Regression discontinuity: formal introduction

The regressions for sharp RD

$$Y_i = \alpha + f(r_i) + \rho D_i + \eta_i$$

- We can model $f(r_i)$ flexibly using p-th order polynomials

$$f(r_i) = \mu + \beta_1 r_i + \beta_2 r_i^2 + \cdots + \beta_p r_i^p$$

- It remains restrictive ; we can instead have separate polynomial trends for each side of the cutoff:

$$E\left[\mathrm{Y}_{0i} \mid r_i\right] = f_0\left(r_i\right) = \alpha + \beta_{01}\tilde{r}_i + \beta_{02}\tilde{r}_i^2 + \ldots + \beta_{0p}\tilde{r}_i^p$$
$$E\left[\mathrm{Y}_{1i} \mid r_i\right] = f_1\left(r_i\right) = \alpha + \rho + \beta_{11}\tilde{r}_i + \beta_{12}\tilde{r}_i^2 + \ldots + \beta_{1p}\tilde{r}_i^p$$

with $\tilde{r}_i \equiv r_i - r_0$

Regression discontinuity: formal introduction

The regressions for sharp RD

- Use the fact that D_i is a deterministic function of r_i

$$E\left[\mathrm{Y}_i \mid r_i\right] = E\left[\mathrm{Y}_{0i} \mid r_i\right] + E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid r_i\right]\mathrm{D}_i$$

- Substituting polynomials for conditional expectations yields the regression

$$\begin{aligned}\mathrm{Y}_i =& \alpha + \beta_{01}\tilde{r}_i + \beta_{02}\tilde{r}_i^2 + \ldots + \beta_{0p}\tilde{r}_i^p \\ &+ \rho \mathrm{D}_i + \beta_1^* \mathrm{D}_i\tilde{r}_i + \beta_2^* \mathrm{D}_i\tilde{r}_i^2 + \ldots + \beta_p^* \mathrm{D}_i\tilde{r}_i^p + \eta_i\end{aligned}$$

- This regression makes **functional form** assumptions on the polynomial.
- the parameter of interest is at the threshold but the estimation with polynomial terms will be sensitive to observations far away
- Risk of **overfitting**.

Regression discontinuity: formal introduction

Polynomial regressions and overfitting

- Seeing a significant jump does not mean we have causal impact
- the "shapes" of the polynomial arms can induce discontinuity because of the parametric structure
- Overfitting is when your model interpret "noise" as "signal"
- Parametric assumptions "force" data to fit, it can end up badly

Regression discontinuity: formal introduction

Polynomial regressions and overfitting

Polynomial regressions and overfitting

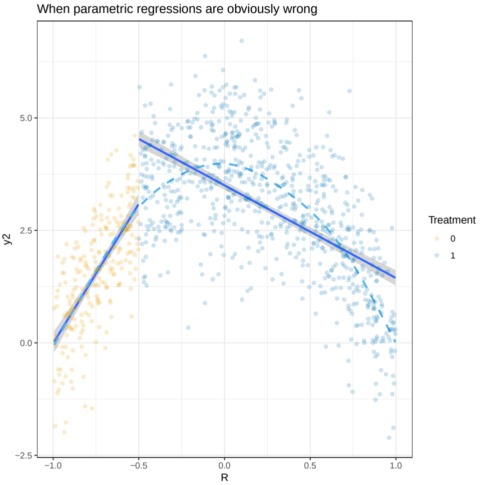
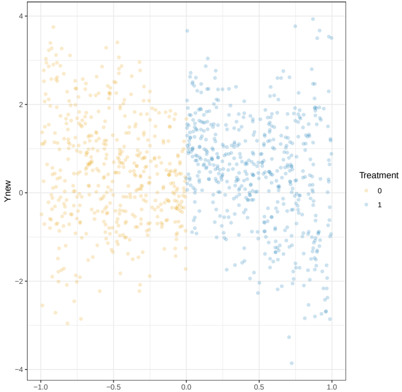


Figure 7: Obvious misspecification

Polynomial regressions and overfitting

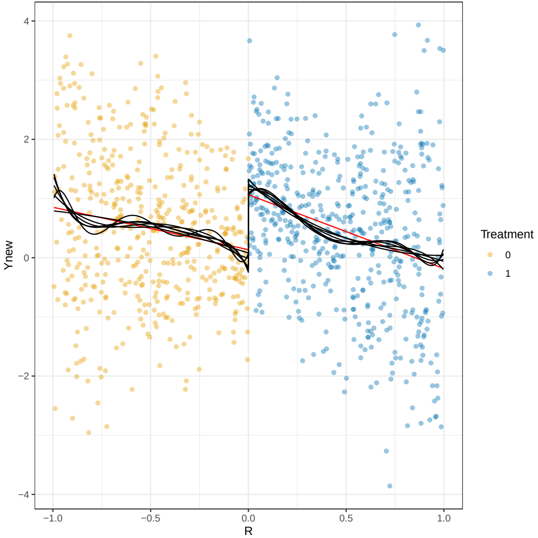
```
# New DGP more random, more noïsy, TE=1
otherx <- runif(1000, -3, 3)
Ynew <- otherx * R^2 - R + 1 * (R >= 0) + rnorm(1000)
datab <- data.frame(R, y, y2, Ynew, t, t2, otherx)

ggplot() + geom_point(aes(x = R, y = Ynew, color = as.factor(t)), alpha = 0.2) +
    scale_colour_manual("Treatment", values = cbbPalette[c(2, 6)])
```



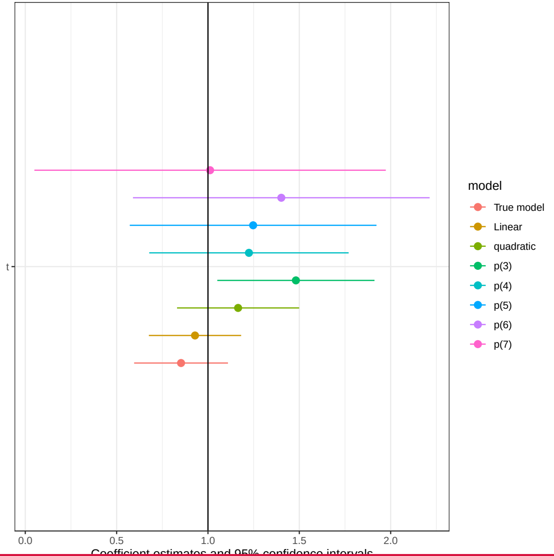
Polynomial regressions and overfitting

```
# estimating polynomial regressions
TRUTH <- lm_robust(Ynew ~ R * otherx + t + t * (R), data = datab, se_type = "HC3")

model1 <- lm_robust(Ynew ~ R + t + t * (R), data = datab, se_type = "HC3")
model2 <- lm_robust(Ynew ~ R + I(R^2) + t + t * (R + I(R^2)), data = datab, se_type = "HC3")
model3 <- lm_robust(Ynew ~ R + I(R^2) + I(R^3) + t + t * (R + I(R^2) + I(R^3)), data = datab,
    se_type = "HC3")
model4 <- lm_robust(Ynew ~ R + I(R^2) + I(R^3) + I(R^4) + t + t * (R + I(R^2) + I(R^3) +
    I(R^4)), data = datab, se_type = "HC3")
model5 <- lm_robust(Ynew ~ R + I(R^2) + I(R^3) + I(R^4) + I(R^5) + t + t * (R + I(R^2) +
    I(R^3) + I(R^4) + I(R^5)), data = datab, se_type = "HC3")
model6 <- lm_robust(Ynew ~ R + I(R^2) + I(R^3) + I(R^4) + I(R^5) + I(R^6) + t + t *
    (R + I(R^2) + I(R^3) + I(R^4) + I(R^5) + I(R^6)), data = datab, se_type = "HC3")
model7 <- lm_robust(Ynew ~ R + I(R^2) + I(R^3) + I(R^4) + I(R^5) + I(R^6) + I(R^7) +
    t + t * (R + I(R^2) + I(R^3) + I(R^4) + I(R^5) + I(R^6) + I(R^7)), data = datab,
    se_type = "HC3")

p1 <- predict(model1)
p2 <- predict(model2)
p3 <- predict(model3)
p4 <- predict(model4)
p5 <- predict(model5)
p6 <- predict(model6)
p7 <- predict(model7)
# data frame with the prediction
d <- data.frame(Ynew, R, t, p1, p2, p3, p4, p5, p6, p7)
```

Polynomial regressions and overfitting



Comparing estimates to the truth



Polynomial regressions and overfitting

So: polynomial regressions ? nope, according to Gelman and Imbens (2019)

- There is often no scientific reason to have high-order polynomials
- Over-fitting: parameter estimates rely on too few data points
- Large weights are given to observations far away from the discontinuity
- Genuine uncertainty from model dependence is not reflected in standard errors

In the end, you may use polynomial regressions as an entry point or robustness check but don't rely on it too much.

Outline

Estimation and inference: the local continuity approach

Roadmap in the local continuity approach

- Global polynomial regressions may introduce biais
- One method to reduce the likelihood of spurious effects is to narrow the bandwidth
- The bandwidth is the "window" below and above the cutoff
- Idea:
  - The closer we "zoom in" on the cutoff
  - ...the lower is the chance of picking up a spurious trend
  - ... but the lower the precision...
- Trade-off between **bias** and **precision** ; can we pick an "optimal" bandwidth ?
- Also, what model should we estimate around the cutoff ?

Estimation and inference: the local continuity approach

Roadmap in the local continuity approach

- If we want to restrict the sample to a bandwidth δ

$$E\left[\mathrm{Y}_i \mid x_0 - \delta < x_i < x_0\right] \simeq E\left[\mathrm{Y}_{0i} \mid x_i = x_0\right]$$
$$E\left[\mathrm{Y}_i \mid x_0 < x_i < x_0 + \delta\right] \simeq E\left[\mathrm{Y}_{1i} \mid x_i = x_0\right]$$

- the estimate becomes

$$\lim_{\delta \to 0} E\left[\mathrm{Y}_i \mid \mathrm{x}_0 < \mathrm{x}_i < \mathrm{x}_0 + \delta\right] - E\left[\mathrm{Y}_i \mid \mathrm{x}_0 - \delta < \mathrm{x}_i < \mathrm{x}_0\right] =$$
$$E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{x}_i = \mathrm{x}_0\right]$$

Estimation and inference: the local continuity approach

Roadmap in the local continuity approach

- Local polynomial regression: captures idea of "localization".
  1. Choose low poly order (p) and weighting scheme $(K(\cdot))$
  2. Choose bandwidth h: Mean squared error-optimal (MSE-optimal) or coverage-error optimal (CE-optimal)
  3. Construct point estimator $\hat{\tau}$
  4. (MSE-optimal h $\Rightarrow$ optimal estimator)
  5. Conduct robust bias-corrected inference
  6. (CE-optimal h $\Rightarrow$ optimal distributional approximation)
- Let's start with bandwidth to understand the issue

Estimation and inference: the local continuity approach

Bandwidth selection

- **Idea:** approximate regression functions for control and treatment units locally.
- **"Local-linear"** $(p = 1)$ estimator (w/ weights $K(\cdot)$) :

$$\begin{array}{c|c} -h \leq X_i < c : & c \leq X_i \leq h : \\ Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i} & Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i} \end{array}$$

- Treatment effect (at the cutoff): $\hat{\tau}_{\text{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$
- Can be estimated using linear models (w/ weights $K(\cdot)$) :

$$Y_i = \alpha + \tau_{\text{SRD}} \cdot T_i + (X_i - c) \cdot \beta_1 + T_i \cdot (X_i - c) \cdot \gamma_1 + \varepsilon_i, \quad |X_i - c| \leq h$$

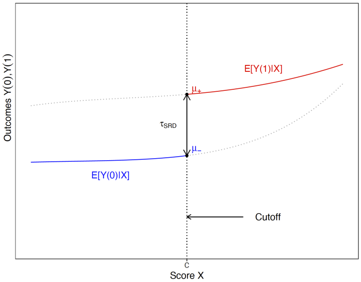
- Given p, K, h chosen $\Longrightarrow$ weighted least squares estimation.

Estimation and inference: the local continuity approach

Bandwidth selection

Figure 11: Continuity of the CEF around the cutoff (from (Cattaneo, Idrobo, and Titiunik 2019))

$$\tau_{\text{SRD}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}}$$



Estimation and inference: the local continuity approach

Bandwidth selection

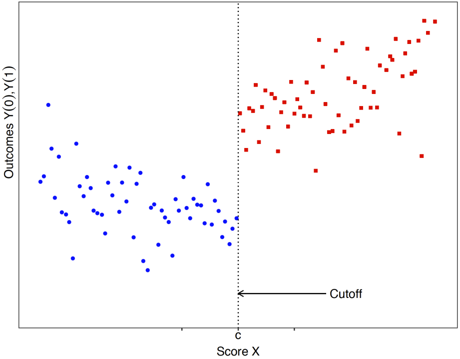
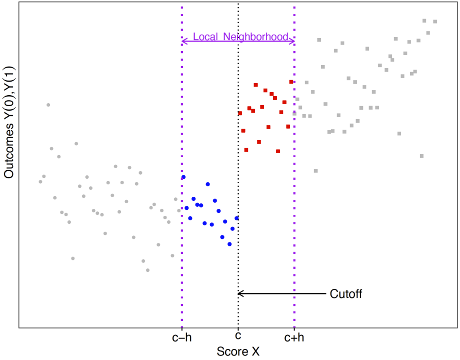


Figure 12: Raw data or binned means (from (Cattaneo, Idrobo, and Titiunik 2019))

Estimation and inference: the local continuity approach

Bandwidth selection

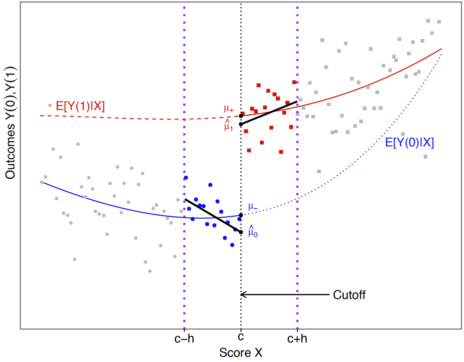
Figure 13: Define bandwidth around the cutoff (from (Cattaneo, Idrobo, and Titiunik 2019))



Estimation and inference: the local continuity approach

Bandwidth selection

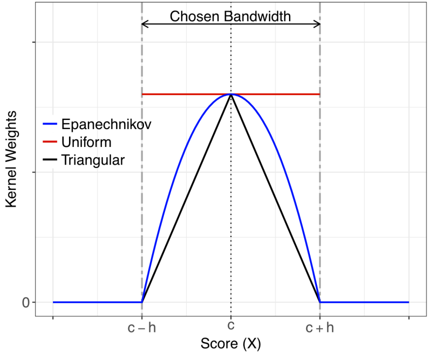
Figure 14: Local linear regression as local estimate of the CEF (from (Cattaneo, Idrobo, and Titiunik 2019))



Estimation and inference: the local continuity approach

Choosing weights

Figure 15: Weighting functions: Kernel, triangular, Epanechnikov, uniform (from (Cattaneo, Idrobo, and Titiunik 2019))



Estimation and inference: the local continuity approach

Choosing weights

- The kernel function $K(\cdot)$ assigns non-negative weights to each transformed observation $\frac{X_i - c}{h}$, based on the distance between the observation's score Xi and the cutoff c.
- The recommended choice is the triangular kernel function, $K(u) = (1 - |u|)\mathbb{1}(|u| \leq 1)$.
- That's because Triangular Kernel in conjunction with a bandwidth that optimizes the mean squared error (MSE) leads to a point estimator with optimal properties
- The weight is maximized at $Xi = c$, and declines symmetrically and linearly as the value of the score gets farther from the cutoff and 0 weights further away from $c \pm h$
- Employing a local linear estimation with bandwidth h and uniform kernel is therefore equivalent to estimating a simple linear regression without weights using only observations whose distance from the cutoff is at most h.

Estimation and inference: the local continuity approach

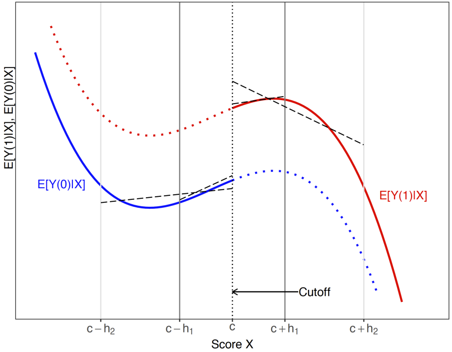
Preference for local linear polynomial

- A more consequential decision is the choice of the local polynomial order, which must consider various factors
  1. A polynomial of order zero—a constant fit—has undesirable theoretical properties at boundary points, which is precisely where RD estimation must occur.
  2. for a given bandwidth, increasing the order of the polynomial generally improves the accuracy of the approximation but also increases the variability of the treatment effect estimator.
  3. higher-order polynomials tend to produce overfitting
- In general the local linear estimator seems to deliver a good trade-off between simplicity, precision, and stability in RD settings.
- An appropriately chosen bandwidth will adjust to the chosen polynomial order so that the linear approximation to the unknown regression functions is reliable.

Estimation and inference: the local continuity approach

Preference for local linear polynomial

Figure 16: The more local, the less biais (from (Cattaneo, Idrobo, and Titiunik 2019))



Estimation and inference: the local continuity approach

Preference for local linear polynomial

- Mean Square Error Optimal (MSE-optimal).

$$h_{\text{MSE}} = C_{\text{MSE}}^{1/5} \cdot n^{-1/5} \quad C_{\text{MSE}} = C(K) \cdot \frac{\text{Var}\left(\hat{\tau}_{\text{SRD}}\right)}{\text{Bias}\left(\hat{\tau}_{\text{SRD}}\right)^2}$$

- Coverage Error Optimal (CE-optimal).

$$h_{\text{CE}} = C_{\text{CE}}^{1/4} \cdot n^{-1/4} \quad C_{\text{CE}} = C(K) \cdot \frac{\text{Var}\left(\hat{\tau}_{\text{SRD}}\right)}{\left|\text{Bias}\left(\hat{\tau}_{\text{SRD}}\right)\right|}$$

- **Key idea:** Trade-off bias and variance of $\hat{\tau}_{\text{SRD}}(h)$. Heuristically:

$$\uparrow \text{Bias}\left(\hat{\tau}_{\text{SRD}}\right) \implies \downarrow \hat{h} \quad \text{and} \quad \uparrow \text{Var}\left(\hat{\tau}_{\text{SRD}}\right) \implies \uparrow \hat{h}$$

- **Implementations:** RDrobust chooses optimal bandwidth for us based on these rules. More info in Calonico, Cattaneo, and Titiunik (2014)

Estimation and inference: the local continuity approach

Conventional inference

- "Local-linear" $(p = 1)$ estimator (w/ weights $K(\cdot)$) :

$$\begin{array}{c|c} -h \leq X_i < c : & c \leq X_i \leq h : \\ Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i} & Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i} \end{array}$$

- Treatment effect (at the cutoff): $\hat{\tau}_{\text{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$
- Construct usual t-test. For $\text{H}_0 : \tau_{\text{SRD}} = 0$,

$$T(h) = \frac{\hat{\tau}_{\text{SRD}}}{\sqrt{\hat{\text{V}}}} = \frac{\hat{\alpha}_+ - \hat{\alpha}_-}{\sqrt{\hat{\text{V}}_+ + \hat{V}_-}} \approx_d \mathcal{N}(0, 1)$$

- Naïve 95% Confidence interval:

$$I(h) = \left[\hat{\tau}_{\text{SRD}} \pm 1.96 \cdot \sqrt{\hat{\text{V}}}\right]$$

Estimation and inference: the local continuity approach

Bias robust correction approach

- Key Problem:

$$T\left(h_{\text{MSE}}\right)=\frac{\hat{\tau}_{\text{SRD}}}{\sqrt{\hat{V}}}\approx_d \mathcal{N}(\text{B},1)\neq\mathcal{N}(0,1)$$

  B captures bias due to misspecification error.

- Robust bias correction: distributional approximation:

$$T^{\text{bc}}(h)=\frac{\hat{\tau}_{\text{SRD}}-\hat{\text{B}}_n}{\sqrt{\hat{\text{V}}}}=\underbrace{\frac{\hat{\tau}_{\text{SRD}}-\text{B}_n}{\sqrt{\hat{\text{V}}}}}_{\approx_d\mathcal{N}(0,1)}+\underbrace{\frac{\text{B}-\hat{\text{B}}}{\sqrt{\hat{\text{V}}}}}_{\approx_d\mathcal{N}(0,\gamma)}$$

- $\hat{B}$ is constructed to estimate leading bias B, that is, misspecification error.
- RBC 95% Confidence Interval:

$$I_{\text{RBC}}=\left[\left(\hat{\tau}_{\text{SRD}}-\hat{\text{B}}\right)\pm 1.96\cdot\sqrt{\hat{\text{V}}+\hat{\text{W}}}\right]$$

Outline

Estimation and inference: local randomization approach

RDD as randomisation around the cutoff

- **Key assumption:** exists window $\mathcal{W} = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff:
  1. Joint probability distribution of scores for units in the $\mathcal{W}$ is known: $\mathbb{P}[\mathbf{X}_{\mathcal{W}} \leq \mathbf{x}] = F(\mathbf{x})$, for some known joint c.d.f. $F(\mathbf{x})$ where $\mathbf{X}_{\mathcal{W}}$ denotes the vector of scores for all i such that $X_i \in \mathcal{W}$.
  2. Potential outcomes not affected by value of the score:

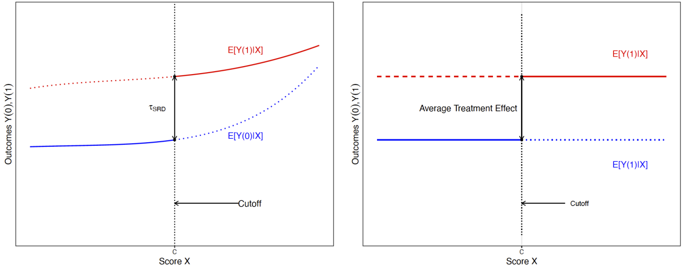
$$\begin{aligned} Y_i(0,x) &= Y_i(0), \\ Y_i(1,x) &= Y_i(1), \quad \text{for all } X_i \in \mathcal{W} \end{aligned}$$

- **Note:** stronger assumption than continuity-based approach.
- Potential outcomes are a constant function of the score (can be relaxed).
- Regression functions are not only continuous at c, but also completely unaffected by the running variable in $\mathcal{W}$.

Estimation and inference: local randomization approach

RDD as randomisation around the cutoff

Figure 17: Continuity with RDD vs RCTs (from (Cattaneo, Idrobo, and Titiunik 2019))

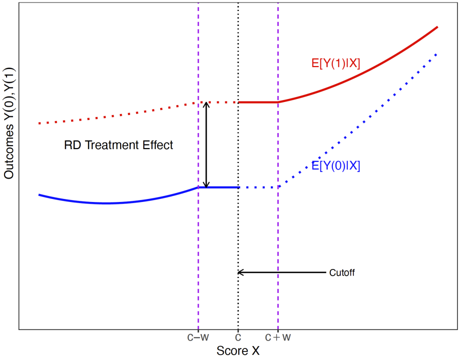


(a) RD Design (b) Randomized Experiment

Estimation and inference: local randomization approach

RDD as randomisation around the cutoff

Figure 18: A neighborhood of local randomization (from (Cattaneo, Idrobo, and Titiunik 2019))



Estimation and inference: local randomization approach

RDD as randomisation around the cutoff

- **Key idea:** exists window $\mathcal{W} = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff.
- Two Steps (analogous to local polynomial methods):
  1. Select window $\mathcal{W}$.
  2. Given window $\mathcal{W}$, perform estimation and inference.
- Challenges
  - Window (neighborhood) selection.
  - As-if random assumption good approximation only very near cutoff
  - Small sample.

Estimation and inference: local randomization approach

Select window $\mathcal{W}$.

- Find neighborhood where (pre-intervention) covariate-balance holds.
- Find neighborhood where outcome and score independent.
- Domain-specific or application-specific choice.

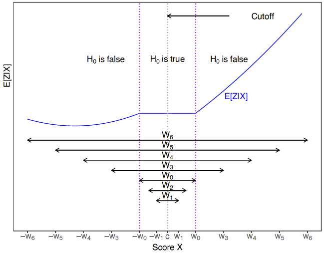


Figure 19: Choosing bandwidth of local randomisation (from (Cattaneo, Idrobo, and Titiunik 2019))

Estimation and inference: local randomization approach

Select window $\mathcal{W}$.

- Given $\mathcal{W}$ where local randomization holds:
  - **Randomization inference** (Fisher): sharp null, finite-sample exact.
  - **Design-based** (Neyman): large-sample valid, conservative.
  - **Large-sample standard**: random potential outcomes, large-sample valid.
- All methods require window ($\mathcal{W}$) selection, and choice of statistic.
- First two also require choice/assumptions assignment mechanism.
- Covariate-adjustments (score or otherwise) possible.

Outline

R practice: replicating Ludwig and Miller (2007)

Abstract

This paper exploits a new source of variation in Head Start funding to identify the program's effects on health and schooling. In 1965 the Office of Economic Opportunity (OEO) provided technical assistance to the 300 poorest counties to develop Head Start proposals. The result was a large and lasting discontinuity in Head Start funding rates at the OEO cutoff for grant-writing assistance. We find evidence of a large drop at the OEO cutoff in mortality rates for children from causes that could be affected by Head Start, as well as suggestive evidence for a positive effect on educational attainment.

R practice: replicating Ludwig and Miller (2007)

Data and notation

- **Problem:** impact of Head Start on Infant Mortality
- Data can be downloaded from HERE in csv
- Once saved, load them

```
data <- read.csv("headstart.csv")
# data <- read.csv('TA Sessions/RDD-HEIM/headstart.csv')
attach(data)
```

- Y_i = child mortality 5 to 9 years old
- T_i = whether county received Head Start assistance
- Running variable $X_i = 1960$ poverty index $\quad (c = 59.1984)$
- Covariates Z_i = see database .
- Potential outcomes:
  - $Y_i(0)$ = child mortality if had not received Head Start
  - $Y_i(1)$ = child mortality if had received Head Start
- Causal Inference:

$$Y_i(0) \neq Y_i \mid T_i = 0 \text{ and } Y_i(1) \neq Y_i \mid T_i = 1$$

R practice: replicating Ludwig and Miller (2007)

Preparing data for RDD analysis

- Packages you need to load for RDD:

```
library(rdrobust)

library(rdlocrand)

library(rddensity)
```

- prepare the data

```
Y <- mort_age59_related_postHS
X <- povrate60
Z <- cbind(census1960_pop, census1960_pctsch1417, census1960_pctsch
    census1960_pop1417, census1960_pop534, census1960_pop25plus, ce
    census1960_pctblack)
C <- 59.1984

R <- X - C
T <- (X > C)
```

R practice: replicating Ludwig and Miller (2007)

Preparing data for RDD analysis

We use *rdrobust::rdplot* function to plot the discontinuity, with Epanechnikov weights, default polynomial=4.

```
rdplot(Y, X, C, binselect = "esmv")
```

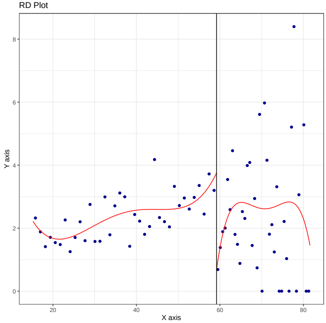


Figure 20: Discontinuity in child mortality around the threshold of headstart eligibility

R practice: replicating Ludwig and Miller (2007)

Preparing data for RDD analysis

Linear arms instead of polynomial

```
rdplot(Y, X, C, p = 1, binselect = "esmv")
```

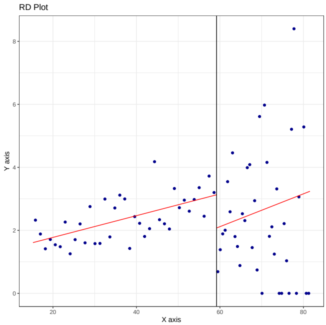


Figure 21: Discontinuity in child mortality around the threshold of headstart eligibility

R practice: replicating Ludwig and Miller (2007)

Preparing data for RDD analysis

More bins, more sensitivity

```r
rdplot(Y, X, C, binselect = "esmv", nbins = 300)
```

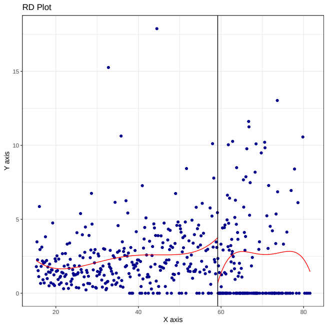


Figure 22: Discontinuity in child mortality around the threshold of headstart eligibility

R practice: replicating Ludwig and Miller (2007)

Replication: OLS regression

In the paper, they run $Y_i = \alpha + \beta T_i + \gamma R_i + \delta T_i \times R_i + \varepsilon_i$ using OLS

	(1)
(Intercept)	3.466***
	(0.846)
TTRUE	−1.895+
	(0.985)
R	0.104
	(0.141)
TTRUE × R	0.080
	(0.194)
Num.Obs.	524
R2	0.008
R2 Adj.	0.003
RMSE	5.77

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

R practice: replicating Ludwig and Miller (2007)

Replication: OLS regression

Using rdrobust, without covariates: formula: *rdrobust(Y, R, h=9, kernel="uni", vce="hc0")*

	(1)
Conventional	−1.895+
	(0.980)
Bias-Corrected	−2.623**
	(0.980)
Robust	−2.623*
	(1.307)
nobs.left	2489
nobs.right	294
nobs.effective.left	309
nobs.effective.right	215
cutoff	0
order.regression	2
order.bias	2
kernel	Uniform
bwselect	Manual

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

R practice: replicating Ludwig and Miller (2007)

Replication: OLS regression

Using rdrobust, with covariates: formula: *rdrobust(Y, X, C, covs=Z, h=9, kernel="uni", vce="hc0")*

	No covariates	Covariates
Conventional	$-1.895+$	-1.943*
	(0.980)	(0.928)
Bias-Corrected	-2.623**	-2.782**
	(0.980)	(0.928)
Robust	-2.623*	-2.782*
	(1.307)	(1.226)
nobs.left	2489	2485
nobs.right	294	294
nobs.effective.left	309	309
nobs.effective.right	215	215
cutoff	0	59.1984
order.regression	2	2
order.bias	2	2
kernel	Uniform	Uniform
bwselect	Manual	Manual

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

R practice: replicating Ludwig and Miller (2007)

Local polynomial

Using rdrobust we can also run local polynomial regression with opitmal bandwidth selection. Comparing weighting functions

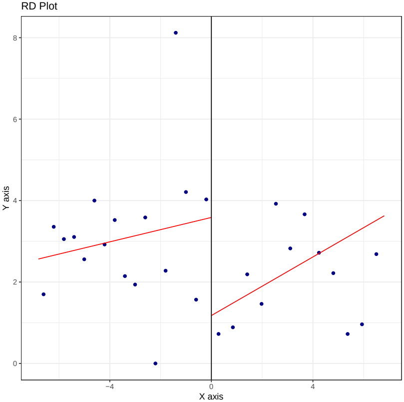
	BW by MSERD	Uniform Kernel	Triangular kernel
Conventional	−2.409*	−1.895+	−2.182*
	(1.206)	(0.980)	(1.036)
Bias-Corrected	−2.781*	−2.623**	−3.036**
	(1.206)	(0.980)	(1.036)
Robust	−2.781*	−2.623*	−3.036*
	(1.368)	(1.307)	(1.283)
nobs.left	2489	2489	2489
nobs.right	294	294	294
nobs.effective.left	234	309	309
nobs.effective.right	180	215	215
cutoff	59.1984	59.1984	59.1984
order.regression	2	2	2
order.bias	2	2	2
kernel	Triangular	Uniform	Triangular
bwselect	mserd	Manual	Manual

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

R practice: replicating Ludwig and Miller (2007)

Local polynomial

We plot the local polynomial RDD

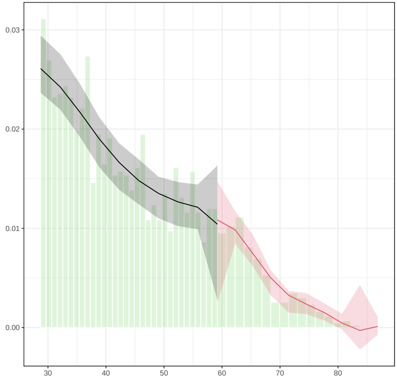


Outline

Falsification tests

Manipulation of the running variable

rddensity implements manipulation testing procedures using the local polynomial density. You shouldn't see bunching or jumps around the cutoff.

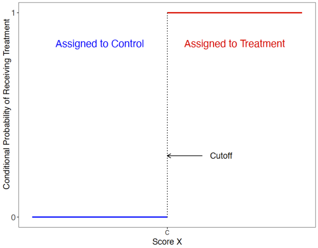


Outline

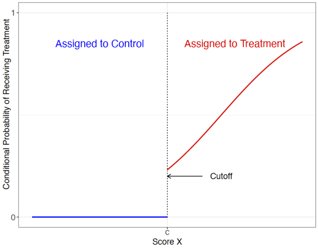
Fuzzy regression discontinuity design

Fuzzy RDD is IV

Figure 23: Only a jump in probability of treatment (from (Cattaneo, Idrobo, and Titiunik 2019))



(a) Sharp RD



(b) Fuzzy RD (one-sided compliance)

Fuzzy regression discontinuity design

Fuzzy RDD is IV

- Imperfect compliance.
- probability of receiving treatment changes at c, but not necessarily from 0 to 1.
- Canonical Parameter:

$$\begin{aligned}\tau_{\mathrm{FRD}} &= \frac{\mathbb{E}\left[(Y_i(1) - Y_i(0)\left(D_i(1) - D_i(0)\right) \mid X_i = c\right]}{\mathbb{E}\left[D_i(1) \mid X_i = c\right] - \mathbb{E}\left[D_i(0) \mid X_i = c\right]} \\ &= \frac{\lim_{x \downarrow c} \mathbb{E}\left[Y_i \mid X_i = x\right] - \lim_{x \uparrow c} \mathbb{E}\left[Y_i \mid X_i = x\right]}{\lim_{x \downarrow c} \mathbb{E}\left[D_i \mid X_i = x\right] - \lim_{x \uparrow c} \mathbb{E}\left[D_i \mid X_i = x\right]}\end{aligned}$$

- Similarly for Local Randomization framework.
- Different interpretations under different assumptions.

Outline

Validity

How to assess the validity of the dessign

- **RD plots and related graphical methods:**
  - Always plot data: main advantage of RD designs. (Check if RD design!)
  - Plot histogram of X_i (score) and its density. Careful: boundary bias.
  - RD plot $\mathbb{E}[Y_i \mid X_i = x]$ (outcome) and $\mathbb{E}[Z_i \mid X_i = x]$ (pre-intervention covariates).
  - Be careful not to oversmooth data/plots.
- **Sensitivity and related methods:**
  - Score density continuity: binomial test and continuity test.
  - Pre-intervention covariate no-effect (covariate balance).
  - Placebo outcomes no-effect.
  - Placebo cutoffs no-effect: informal continuity test away from c.
  - Donut hole: testing for outliers/leverage near c.
  - Different bandwidths: testing for misspecification error.
  - Many other setting-specific (fuzzy, geographic, etc.).

Bibliography I

- Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–196.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.
- Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations.* 1st ed. Cambridge University Press, November 30, 2019. ISBN: 978-1-108-68460-6 978-1-108-71020-6.
- Chaplin, Duncan D., Thomas D. Cook, Jelena Zurovac, Jared S. Coopersmith, Mariel M. Finucane, Lauren N. Vollmer, and Rebecca E. Morris. 2018. "The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis of 15 Within-Study Comparisons." *Journal of Policy Analysis and Management* 37 (2): 403–429.
- Fack, Gabrielle, and Julien Grenet. 2015. "Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program." *American Economic Journal: Applied Economics* 7, no. 2 (April): 1–34.
- Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37, no. 3 (July 3, 2019): 447–456.
- Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142, no. 2 (February): 615–635.
- Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48, no. 2 (June): 281–355.

Bibliography II

- Ludwig, Jens, and Douglas Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *The Quarterly Journal of Economics* 122 (1): 159–208.