12. Regression Discontinuity Designs

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Outline

1. Sharp Regression Discontinuity Designs

2. Estimation in the SRD

3. Fuzzy Regression Discontinuity Designs

4. R Codes for SRD and FRD

Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting, DML for selection on observables.
  - Instrumental variables for when this doesn't hold.
  - Panel methods (e.g., DiD, synthetic control, FE estimators) when units are observed repeatedly.

Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting, DML for selection on observables.
  - Instrumental variables for when this doesn't hold.
  - Panel methods (e.g., DiD, synthetic control, FE estimators) when units are observed repeatedly.
- Basic idea: find exogenous variation in the treatment assignment.
  - RCT: randomization provides exogenous variation.
  - Selection on observables: treatment as-if random conditional on $\mathbf{X}_i$.
  - IV: Instrument provides exogenous variation.
  - DiD: the $D_i \cdot \text{Post}_t$ interaction (the contrast between treated and control pre/post changes) provides exogenous variation under parallel trends.

Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting, DML for selection on observables.
  - Instrumental variables for when this doesn't hold.
  - Panel methods (e.g., DiD, synthetic control, FE estimators) when units are observed repeatedly.
- Basic idea: find exogenous variation in the treatment assignment.
  - RCT: randomization provides exogenous variation.
  - Selection on observables: treatment as-if random conditional on $\mathbf{X}_i$.
  - IV: Instrument provides exogenous variation.
  - DiD: the $D_i \cdot \text{Post}_t$ interaction (the contrast between treated and control pre/post changes) provides exogenous variation under parallel trends.
- Regression discontinuity: a discontinuity in treatment assignment.

Source: *Chapter 6* of Mostly Harmless Econometrics (Textbook 1) by J. Angrist & J. Pischke

1/ Sharp Regression Discontinuity Designs

The Setup

- The basic idea behind RDDs:
  - Treatment assignment is determined by a cutoff in some variable, X_i.
  - X_i is a **forcing/running variable**
- Treatment changes discontinuously at the cutoff,
  - but, unobserved confounders vary smoothly around the cutoff.

The Setup

- The basic idea behind RDDs:
  - Treatment assignment is determined by a cutoff in some variable, X_i.
  - X_i is a **forcing/running variable**
- Treatment changes discontinuously at the cutoff,
  - but, unobserved confounders vary smoothly around the cutoff.
- $\rightsquigarrow$ changes in the outcome at a threshold have a causal interpretation.

The Setup

- The basic idea behind RDDs:
  - Treatment assignment is determined by a cutoff in some variable, X_i.
  - X_i is a **forcing/running variable**
- Treatment changes discontinuously at the cutoff,
  - but, unobserved confounders vary smoothly around the cutoff.
- $\rightsquigarrow$ changes in the outcome at a threshold have a causal interpretation.
- Classic examples of the setup is in the educational context:
  - Merit scholarships that are allocated based on a test score threshold (Thistlethwaite & Campbell, 1960)
  - Class size on test scores using total student thresholds to create new classes (Angrist & Lavy, 1999)

Sharp RD

- Notations:
  - Treatment: $D_i = 1$ or $D_i = 0$
  - Potential outcomes: $Y_i(1)$ and $Y_i(0)$
  - Observed outcomes: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$
  - Continuous forcing variable: $X_i \in \mathbb{R}$ (discrete more complicated)

Sharp RD

- Notations:
  - Treatment: $D_i = 1$ or $D_i = 0$
  - Potential outcomes: $Y_i(1)$ and $Y_i(0)$
  - Observed outcomes: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$
  - Continuous forcing variable: $X_i \in \mathbb{R}$ (discrete more complicated)
- **Sharp RD**: $D_i = 1\{X_i \geq c\} \;\; \forall\, i$
  - Treatment is a **deterministic** function of the forcing variable and the threshold.
  - When test scores are above 1,500 $\longrightarrow$ offered scholarship
  - When test scores are below 1,500 $\longrightarrow$ not offered scholarship

Sharp RD

- Notations:
  - Treatment: $D_i = 1$ or $D_i = 0$
  - Potential outcomes: $Y_i(1)$ and $Y_i(0)$
  - Observed outcomes: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$
  - Continuous forcing variable: $X_i \in \mathbb{R}$ (discrete more complicated)
- **Sharp RD**: $D_i = 1\{X_i \geq c\} \quad \forall\, i$
  - Treatment is a **deterministic** function of the forcing variable and the threshold.
  - When test scores are above 1,500 $\longrightarrow$ offered scholarship
  - When test scores are below 1,500 $\longrightarrow$ not offered scholarship
- Note: positivity/overlap violated by design here
  - $\mathbb{P}[D_i = 1 | X_i = c - \varepsilon] = 0$
  - $\mathbb{P}[D_i = 1 | X_i = c + \varepsilon] = 1$
  - $\rightsquigarrow$ Can't use standard identification toolkit for ATE/ATT.

Plotting the RDD (Imbens and Lemieux, 2008)

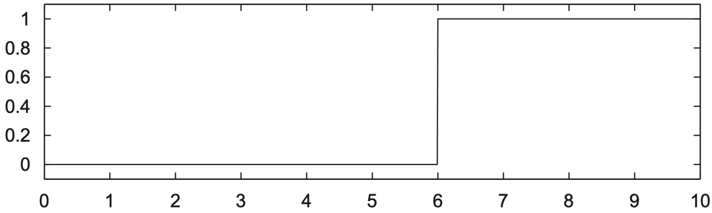


Fig. 1. Assignment probabilities (SRD).

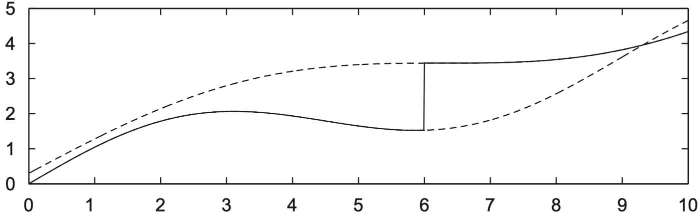


Fig. 2. Potential and observed outcome regression functions.

Source: *Figure 4* in Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142, no. 2 (2008): 615-635.

Plotting the RDD (Imbens and Lemieux, 2008)

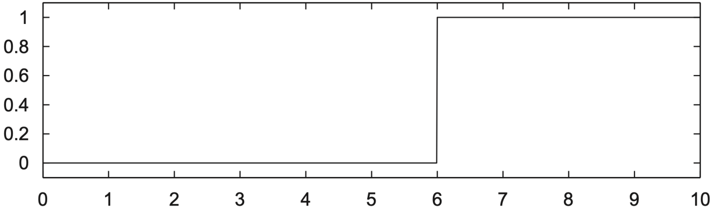


Fig. 1. Assignment probabilities (SRD).

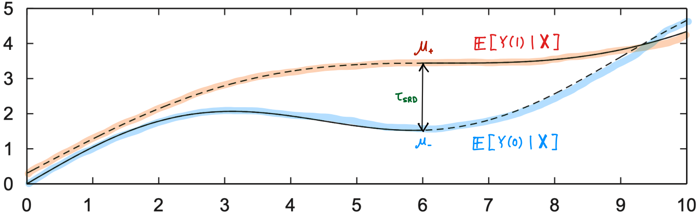


Fig. 2. Potential and observed outcome regression functions.

Source: *Figure 4* in Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142, no. 2 (2008): 615-635.

Quantity of Interest

- Estimand: **local** average treatment effect at the cutoff

$$\begin{aligned}\tau_{\text{SRD}} &= \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c] \\ &= \mathbb{E}[Y_i(1) | X_i = c] - \mathbb{E}[Y_i(0) | X_i = c]\end{aligned}$$

  - Very difficult to extrapolate beyond this.

Quantity of Interest

- Estimand: **local** average treatment effect at the cutoff

$$\begin{aligned}\tau_{\text{SRD}} &= \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c] \\ &= \mathbb{E}[Y_i(1) | X_i = c] - \mathbb{E}[Y_i(0) | X_i = c]\end{aligned}$$

  - Very difficult to extrapolate beyond this.
- Problem: X_i is continuous so we never observe $X_i = c$.
  - $\rightsquigarrow$ Identification comes from **extrapolation** around c
  - Extrapolation requires **smoothness**

Continuity of the CEFs

- **Assumption**: CEFs of potential outcomes are **continuous** in X_i
  - $\mu_1(x) = \mathbb{E}[Y_i(1)|X_i = x]$ is continuous
  - $\mu_0(x) = \mathbb{E}[Y_i(0)|X_i = x]$ is continuous

Continuity of the CEFs

- **Assumption**: CEFs of potential outcomes are **continuous** in X_i
  - $\mu_1(x) = \mathbb{E}[Y_i(1)|X_i = x]$ is continuous
  - $\mu_0(x) = \mathbb{E}[Y_i(0)|X_i = x]$ is continuous

- This continuity implies the following:

$$\mathbb{E}[Y_i(0)|X_i = c] = \lim_{x \uparrow c} \mathbb{E}[Y_i(0)|X_i = x] \quad \text{(continuity)}$$

Continuity of the CEFs

- **Assumption**: CEFs of potential outcomes are **continuous** in X_i
  - $\mu_1(x) = \mathbb{E}[Y_i(1)|X_i = x]$ is continuous
  - $\mu_0(x) = \mathbb{E}[Y_i(0)|X_i = x]$ is continuous

- This continuity implies the following:

$$\begin{aligned}
\mathbb{E}[Y_i(0)|X_i = c] &= \lim_{x \uparrow c} \mathbb{E}[Y_i(0)|X_i = x] \quad \text{(continuity)} \\
&= \lim_{x \uparrow c} \mathbb{E}[Y_i(0)|D_i = 0, X_i = x] \quad \text{(SRD)}
\end{aligned}$$

Continuity of the CEFs

- **Assumption**: CEFs of potential outcomes are **continuous** in X_i
  - $\mu_1(x) = \mathbb{E}[Y_i(1)|X_i = x]$ is continuous
  - $\mu_0(x) = \mathbb{E}[Y_i(0)|X_i = x]$ is continuous
- This continuity implies the following:

$$\begin{aligned}\mathbb{E}[Y_i(0)|X_i = c] &= \lim_{x \uparrow c} \mathbb{E}[Y_i(0)|X_i = x] \quad \text{(continuity)} \\ &= \lim_{x \uparrow c} \mathbb{E}[Y_i(0)|D_i = 0, X_i = x] \quad \text{(SRD)} \\ &= \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x] \quad \text{(consistency/SRD)}\end{aligned}$$

- Note that this is the same for the treated group:

$$\mathbb{E}[Y_i(1)|X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]$$

Identification Results

- Consistency + SRD + Continuity $\rightsquigarrow$ Identification:

$$\begin{aligned}\tau_{\text{SRD}} &= \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c] \\ &= \mathbb{E}[Y_i(1)|X_i = c] - \mathbb{E}[Y_i(0)|X_i = c]\end{aligned}$$

Identification Results

- Consistency + SRD + Continuity $\rightsquigarrow$ Identification:

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

Identification Results

- Consistency + SRD + Continuity $\rightsquigarrow$ Identification:

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

- Problem: Estimate two regression functions at a point.
    - Without parametric assumptions, can be challenging!
    - Nonparametric regression can be consistent, but convergence is slow and

Identification Results

- Consistency + SRD + Continuity $\rightsquigarrow$ Identification:

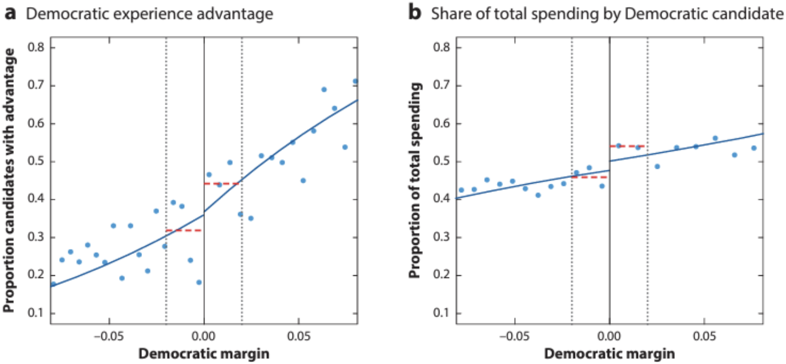
$$\begin{aligned}\tau_{\text{SRD}} &= \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c] \\ &= \mathbb{E}[Y_i(1) | X_i = c] - \mathbb{E}[Y_i(0) | X_i = c] \\ &= \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]\end{aligned}$$

- Problem: Estimate two regression functions at a point.
  - Without parametric assumptions, can be challenging!
  - Nonparametric regression can be consistent, but convergence is slow and
- NB/note well: Not equivalent to **local randomization**,

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp \mathbb{1}\{X_i > c\} \;\mid\; c_0 \leq X_i \leq c_1$$

  - LR is stronger than continuity b/c it rules out confounding around c.
  - Implies no slope in $\mathbb{E}[Y_i(d) \mid X_i = x]$ around c.

Issues with Local Randomization Assumptions



Source: Figure 1 in De la Cuesta, Brandon, and Kosuke Imai. "Misunderstandings about the regression discontinuity design in the study of close elections." Annual Review of Political Science 19 (2016): 375-396.

What Can Go Wrong?

- Key question: why is there a discontinuity in D_i but not $Y_i(d)$?
  - What else might change at the cutoff?
  - Using 16 age cutoff for RDD of Korea's game shutdown law?
- **Sorting around the threshold:** possible violation of smoothness.
  - Students retaking exams to pass some threshold for financial aid.
  - Students with more money $\rightsquigarrow$ more exam retaking $\rightsquigarrow$ sorting.

2/ Estimation in the SRD

Bin Plots

- Standard procedure: **binned means plot** for graphical analysis

$$\overline{Y}_k = \frac{1}{n_k} \sum_{i=1}^{N} Y_i \cdot \mathbf{1}(b_k < X_i \leq b_{k+1})$$

  - where b_k are the bin cutpoints.
  - n_k is the number of units within bin k.

Bin Plots

- Standard procedure: **binned means plot** for graphical analysis

$$\overline{Y}_k = \frac{1}{n_k} \sum_{i=1}^{N} Y_i \cdot \mathbf{1}(b_k < X_i \leq b_{k+1})$$

  - where b_k are the bin cutpoints.
  - n_k is the number of units within bin k.

- What to observe:
  - Obvious discontinuity at the threshold?
  - Also, are there other unexplained discontinuities?

Bin Plots

- Standard procedure: **binned means plot** for graphical analysis

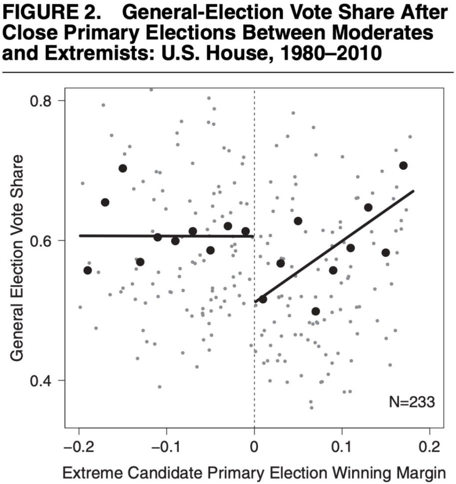
$$\overline{Y}_k = \frac{1}{n_k} \sum_{i=1}^{N} Y_i \cdot \mathbf{1}(b_k < X_i \leq b_{k+1})$$

  - where b_k are the bin cutpoints.
  - n_k is the number of units within bin k.

- What to observe:
  - Obvious discontinuity at the threshold?
  - Also, are there other unexplained discontinuities?

- Very difficult to sell an RDD without visually obvious results:
  - Imbens & Lemieux (2008): “statistical analysis are just fancy versions of this plot”
  - If it’s not in the binned means plot, unlikely to be a robust/credible effect.

Example of a Binned Means Plot



Source: *Figure 2* in Hall, Andrew B. "What happens when extremists win primaries?" American Political Science Review 109, no. 1 (2015): 18-42.

Other Graphs to Include

- Also good to include binned mean plots for pretreatment covariates.
- Intuition: key assumption in smoothness in the mean of $Y_i(d)$ in X_i.
- Discontinuities in mean of covariates $\rightsquigarrow$ problematic
    - Covariates unaffected by treatment so might indicate sorting.
    - May be an indication of discontinuities in the potential outcome means.
    - Similar to balance tests in matching

Other Graphs to Include

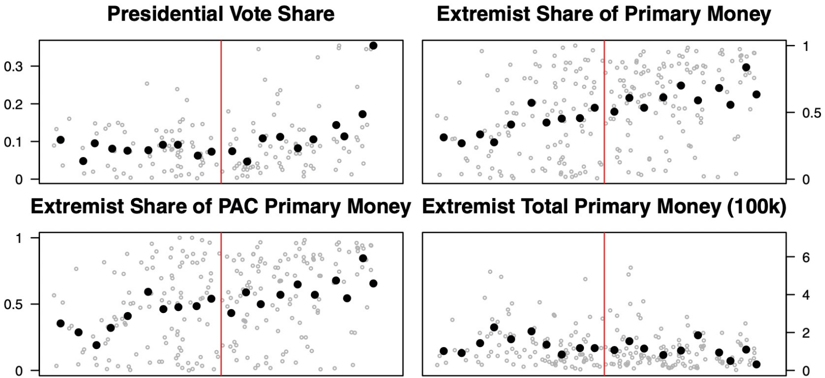
- Also good to include binned mean plots for pretreatment covariates.
- Intuition: key assumption in smoothness in the mean of $Y_i(d)$ in X_i.

- Discontinuities in mean of covariates $\rightsquigarrow$ problematic
  - Covariates unaffected by treatment so might indicate sorting.
  - May be an indication of discontinuities in the potential outcome means.
  - Similar to balance tests in matching

- **McCrary test:** plot density of the forcing variable.
  - Separate densities on either side of the cutoff.
  - If there's a discontinuity in the density, maybe a sign of sorting.

Checking Covariates at the Discontinuity

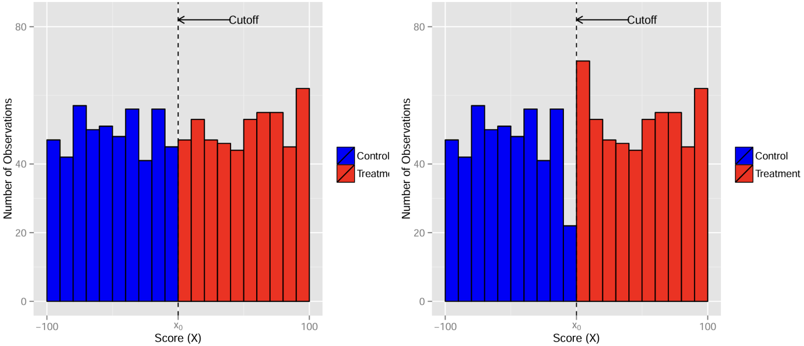
FIGURE A.2. Graphical Balance Tests



Source: *Figure A.2* of Hall, Andrew B. "What happens when extremists win primaries?" American Political Science Review 109, no. 1 (2015): 18-42.

McCrary Test

Figure 10: Histogram of Score



(a) No sorting (b) Sorting

Source: Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

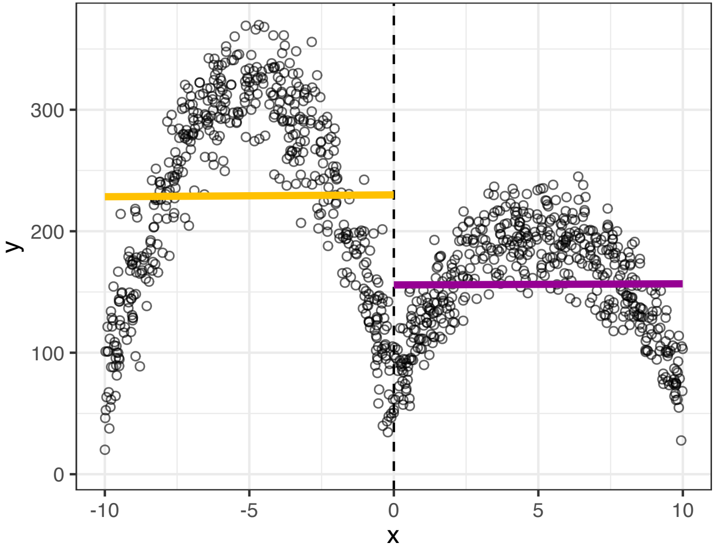
General Estimation Strategy

- The main goal of RD is to estimate the limits of CEFs such as:

$$\lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]$$

- Two features different from standard nonparametric regression:
  - We want to estimate this regression at a single point.
  - This point is a **boundary point**, making estimation challenging.
- Bias of nonparametric estimation at a boundary shrinks slowly.
  - Only getting data from one side of the boundary!
- Naive approach: difference in means
  - Problem: uses data too far from the boundary.

Example of Misleading Trends



Nonparametric and Semiparametric Approaches

- Upper and lower limit functions:

$$\mu_+(x) = \lim_{z \downarrow x} \mathbb{E}[Y_i(1) \mid X_i = z]$$

$$\mu_-(x) = \lim_{z \uparrow x} \mathbb{E}[Y_i(0) \mid X_i = z]$$

- For the SRD, we have $\tau_{\text{SRD}} = \mu_+(c) - \mu_-(c)$.

Nonparametric and Semiparametric Approaches

- Upper and lower limit functions:

$$\mu_+(x) = \lim_{z \downarrow x} \mathbb{E}[Y_i(1) \mid X_i = z]$$

$$\mu_-(x) = \lim_{z \uparrow x} \mathbb{E}[Y_i(0) \mid X_i = z]$$

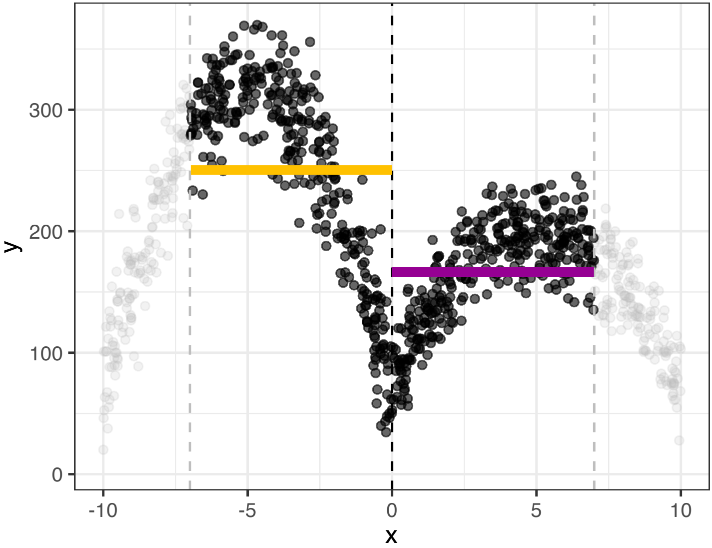
- For the SRD, we have $\tau_{\text{SRD}} = \mu_+(c) - \mu_-(c)$.

- Kernel regression with **uniform kernel**:

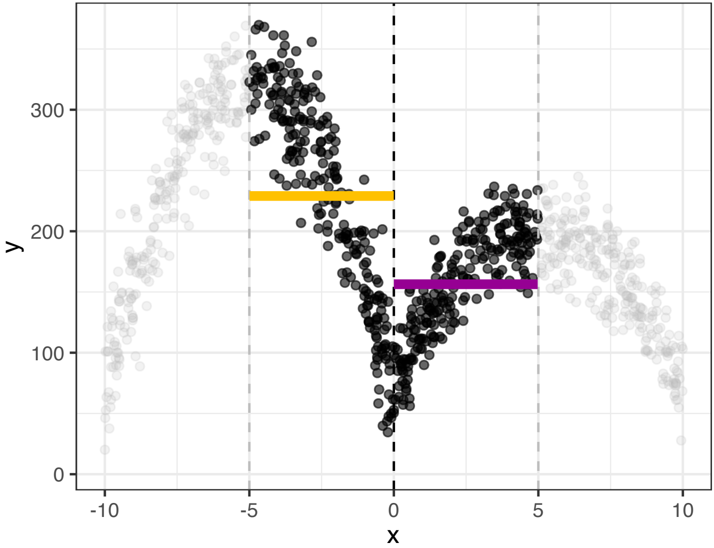
$$\widehat{\mu}_-(c) = \frac{\sum_{i=1}^N Y_i \cdot \mathbf{1}\{c - h \leq X_i < c\}}{\sum_{i=1}^N \mathbf{1}\{c - h \leq X_i < c\}}$$

  - h is a bandwidth/tuning parameter, selected by you.
- Basically means among units no more than h away from the threshold.

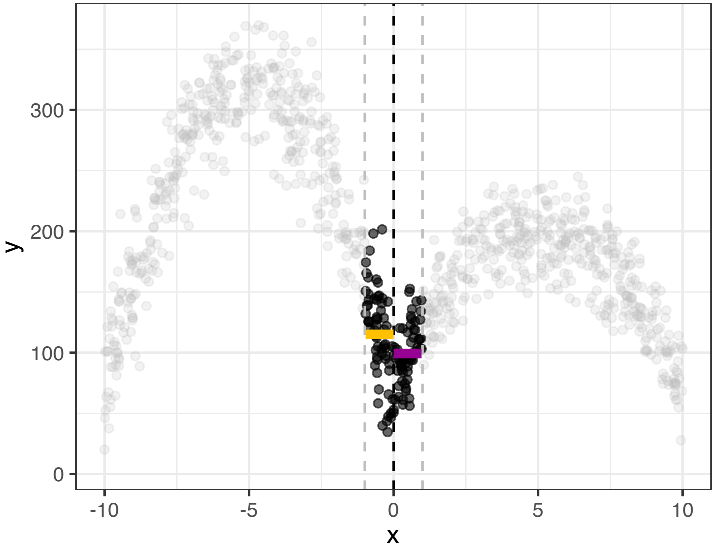
Bandwidth Equal to 7



Bandwidth Equal to 5



Bandwidth Equal to 1



Local Averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.
- Can also view as regression on those units less than h away from c:

$$(\widehat{\alpha}, \widehat{\tau}_{\text{SRD}}) = \arg\min_{\alpha,\tau} \sum_{i:X_i \in [c-h, c+h]} (Y_i - \alpha - \tau D_i)^2$$

  - Predictions about Y_i are locally constant on either side of the cutoff.

Local Averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.
- Can also view as regression on those units less than h away from c:

$$(\widehat{\alpha}, \widehat{\tau}_{\text{SRD}}) = \underset{\alpha,\tau}{\arg\min} \sum_{i:X_i \in [c-h, c+h]} (Y_i - \alpha - \tau D_i)^2$$

  - Predictions about Y_i are locally constant on either side of the cutoff.

- h is a **tuning parameter** that controls the **bias-variance tradeoff**:
  - High h: high bias, low variance (larger n, but farther from the cutoff)
  - Low h: low bias, high variance (smaller n, but closer to the cutoff)

Local Averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.
- Can also view as regression on those units less than h away from c:

$$(\widehat{\alpha}, \widehat{\tau}_{\text{SRD}}) = \arg\min_{\alpha,\tau} \sum_{i: X_i \in [c-h, c+h]} (Y_i - \alpha - \tau D_i)^2$$

  - Predictions about Y_i are locally constant on either side of the cutoff.

- h is a **tuning parameter** that controls the **bias-variance tradeoff**:
  - High h: high bias, low variance (larger n, but farther from the cutoff)
  - Low h: low bias, high variance (smaller n, but closer to the cutoff)

- Downside with averages: bias shrinks slowly as h shrinks.
  - Likely large finite sample bias, poor coverage of confidence intervals.

Local Linear Regression

- Instead of a local constant, we can use a **local linear regression**

Local Linear Regression

- Instead of a local constant, we can use a **local linear regression**
- Run a linear regression of Y_i on $X_i - c$ in the group $X_i \in [c-h, c)$:

$$(\widehat{\alpha}_-, \widehat{\beta}_-) = \arg\min_{\alpha,\beta} \sum_{i:X_i \in [c-h,c)} (Y_i - \alpha - \beta(X_i - c))^2$$

- Same regression for group with $X_i \in [c, c+h]$:

$$(\widehat{\alpha}_+, \widehat{\beta}_+) = \arg\min_{\alpha,\beta} \sum_{i:X_i \in [c,c+h]} (Y_i - \alpha - \beta(X_i - c))^2$$

Local Linear Regression

- Instead of a local constant, we can use a **local linear regression**
- Run a linear regression of Y_i on $X_i - c$ in the group $X_i \in [c - h, c)$:

$$(\widehat{\alpha}_-, \widehat{\beta}_-) = \arg\min_{\alpha,\beta} \sum_{i:X_i \in [c-h,c)} (Y_i - \alpha - \beta(X_i - c))^2$$

- Same regression for group with $X_i \in [c, c + h]$:

$$(\widehat{\alpha}_+, \widehat{\beta}_+) = \arg\min_{\alpha,\beta} \sum_{i:X_i \in [c,c+h]} (Y_i - \alpha - \beta(X_i - c))^2$$

- Our estimate is

$$\begin{aligned} \widehat{\tau}_{\text{SRD}} &= \widehat{\mu}_+(c) - \widehat{\mu}_-(c) \\ &= \widehat{\alpha}_+ + \widehat{\beta}_+(c - c) - \widehat{\alpha}_- - \widehat{\beta}_-(c - c) \\ &= \widehat{\alpha}_+ - \widehat{\alpha}_- \end{aligned}$$

More Practical Estimation

- Simplest to use one regression:

$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i:X_i \in [c-h, c+h]} \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

  - $\widehat{\tau}_{\text{SRD}} = \widehat{\tau}$ is the coefficient on the treatment.
  - Key: interaction between treatment and forcing variable.
- Yields numerically the same as the separate regressions.

More Practical Estimation

- Simplest to use one regression:

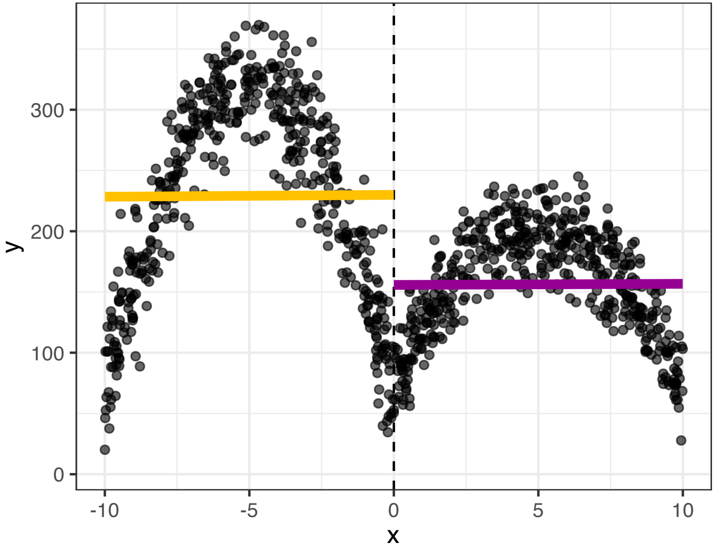
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i:X_i \in [c-h, c+h]} \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

  - $\widehat{\tau}_{\text{SRD}} = \widehat{\tau}$ is the coefficient on the treatment.
  - Key: interaction between treatment and forcing variable.
- Yields numerically the same as the separate regressions.
- Often better to use a **kernel** to weight points close to c more heavily.

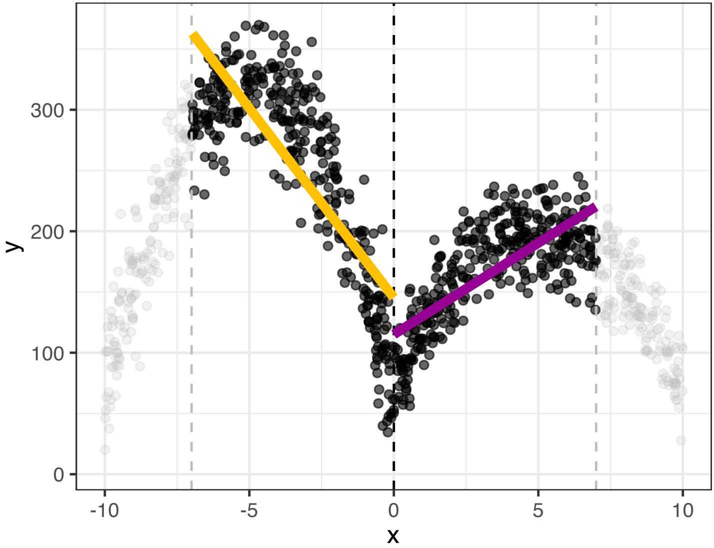
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

  - Popular choice is the **triangular kernel**: $K(u) = (1 - |u|) \cdot \mathbf{1}(|u| < 1)$
  - u is standardized distance from the cutoff

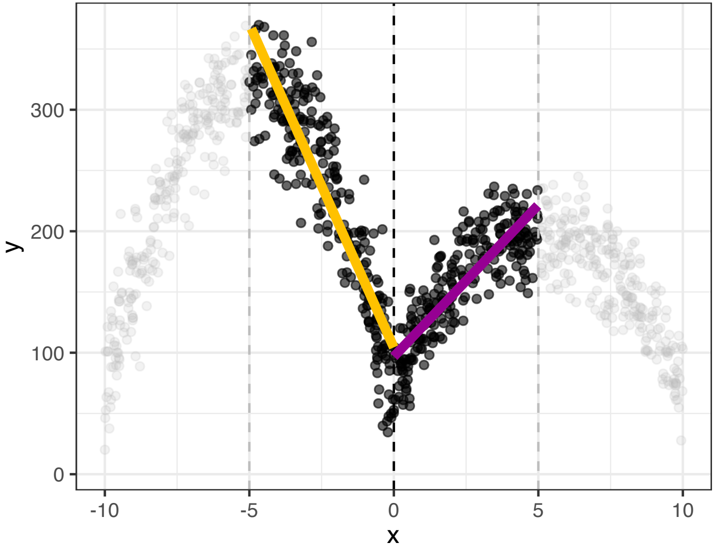
Bandwidth Equal to 10 (Global)



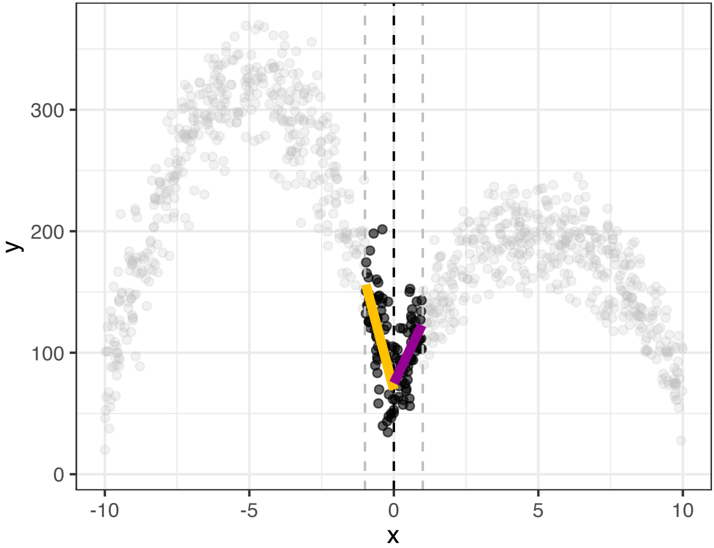
Bandwidth Equal to 7



Bandwidth Equal to 5



Bandwidth Equal to 1



Odds and ends for the SRD

- **Bandwidth selection**: bias $\propto h^2$ (extrapolation grows with h) vs. variance $\propto 1/(nh)$ (less data near the cutoff) $\rightsquigarrow$ MSE-optimal h balances the two.
  - $\text{MSE}(h) \approx h^4\mathcal{B}^2 + \mathcal{V}/(nh) \implies h_n^* = (\mathcal{V}/(4\mathcal{B}^2))^{1/5}n^{-1/5}$; $\mathcal{B}, \mathcal{V}$ estimated via a local quadratic plug-in (IK 2012; CCT 2014).

Odds and ends for the SRD

- **Bandwidth selection**: bias $\propto h^2$ (extrapolation grows with h) vs. variance $\propto 1/(nh)$ (less data near the cutoff) $\rightsquigarrow$ MSE-optimal h balances the two.
    - $\text{MSE}(h) \approx h^4\mathcal{B}^2 + \mathcal{V}/(nh) \implies h_n^* = (\mathcal{V}/(4\mathcal{B}^2))^{1/5} n^{-1/5}$; $\mathcal{B}, \mathcal{V}$ estimated via a local quadratic plug-in (IK 2012; CCT 2014).
- **Standard errors**: robust SE from local OLS is asymptotically valid; finite-sample CI coverage can be off, so CCT (2014) provides a better variance estimator. `rdrobust` handles this automatically.

Odds and ends for the SRD

- **Bandwidth selection**: bias $\propto h^2$ (extrapolation grows with h) vs. variance $\propto 1/(nh)$ (less data near the cutoff) $\rightsquigarrow$ MSE-optimal h balances the two.
    - $\text{MSE}(h) \approx h^4\mathcal{B}^2 + \mathcal{V}/(nh) \implies h_n^* = (\mathcal{V}/(4\mathcal{B}^2))^{1/5}n^{-1/5}$; $\mathcal{B}, \mathcal{V}$ estimated via a local quadratic plug-in (IK 2012; CCT 2014).
- **Standard errors**: robust SE from local OLS is asymptotically valid; finite-sample CI coverage can be off, so CCT (2014) provides a better variance estimator. `rdrobust` handles this automatically.
- **Covariates**: can add them to the local linear model, but be wary.
    - If covariates are continuous at the cutoff, won't affect estimates much.
    - If they aren't, raises suspicions about identification.
    - ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST

Odds and ends for the SRD

- **Bandwidth selection**: bias $\propto h^2$ (extrapolation grows with h) vs. variance $\propto 1/(nh)$ (less data near the cutoff) $\rightsquigarrow$ MSE-optimal h balances the two.
    - $\text{MSE}(h) \approx h^4\mathcal{B}^2 + \mathcal{V}/(nh) \implies h_n^* = (\mathcal{V}/(4\mathcal{B}^2))^{1/5}n^{-1/5}$; $\mathcal{B}, \mathcal{V}$ estimated via a local quadratic plug-in (IK 2012; CCT 2014).
- **Standard errors**: robust SE from local OLS is asymptotically valid; finite-sample CI coverage can be off, so CCT (2014) provides a better variance estimator. `rdrobust` handles this automatically.
- **Covariates**: can add them to the local linear model, but be wary.
    - If covariates are continuous at the cutoff, won't affect estimates much.
    - If they aren't, raises suspicions about identification.
    - ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST
- Possible to use local polynomial regression beyond linear, but performance is poor (very sensitive to end points).

Odds and ends for the SRD

- **Bandwidth selection**: bias $\propto h^2$ (extrapolation grows with h) vs. variance $\propto 1/(nh)$ (less data near the cutoff) $\rightsquigarrow$ MSE-optimal h balances the two.
  - $\text{MSE}(h) \approx h^4\mathcal{B}^2 + \mathcal{V}/(nh) \implies h_n^* = (\mathcal{V}/(4\mathcal{B}^2))^{1/5} n^{-1/5}$; $\mathcal{B}, \mathcal{V}$ estimated via a local quadratic plug-in (IK 2012; CCT 2014).
- **Standard errors**: robust SE from local OLS is asymptotically valid; finite-sample CI coverage can be off, so CCT (2014) provides a better variance estimator. `rdrobust` handles this automatically.
- **Covariates**: can add them to the local linear model, but be wary.
  - If covariates are continuous at the cutoff, won't affect estimates much.
  - If they aren't, raises suspicions about identification.
  - ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST
- Possible to use local polynomial regression beyond linear, but performance is poor (very sensitive to end points).
- Use {`rdrobust`} package for CCT bandwidths/estimation.

3/ Fuzzy Regression Discontinuity Designs

The Setup

- Recall treatment: $D_i = 1$ or $D_i = 0$ and forcing variable: X_i.
- **Fuzzy RD:** discontinuity in the probability of treatment.

$$\lim_{x \uparrow c} \mathbb{P}[D_i = 1 \mid X_i = x] \neq \lim_{x \downarrow c} \mathbb{P}[D_i = 1 \mid X_i = x]$$

  - No longer deterministic function of forcing variable.
  - SRD is a special case of the FRD.

The Setup

- Recall treatment: $D_i = 1$ or $D_i = 0$ and forcing variable: X_i.
- **Fuzzy RD:** discontinuity in the probability of treatment.

$$\lim_{x \uparrow c} \mathbb{P}[D_i = 1 \mid X_i = x] \neq \lim_{x \downarrow c} \mathbb{P}[D_i = 1 \mid X_i = x]$$

  - No longer deterministic function of forcing variable.
  - SRD is a special case of the FRD.
- Common use case: threshold allows participation in program.
  - Some might not participate even if allowed (noncompliance).

The Setup

- Recall treatment: $D_i = 1$ or $D_i = 0$ and forcing variable: X_i.
- **Fuzzy RD:** discontinuity in the probability of treatment.

$$\lim_{x \uparrow c} \mathbb{P}[D_i = 1 \mid X_i = x] \neq \lim_{x \downarrow c} \mathbb{P}[D_i = 1 \mid X_i = x]$$

  - No longer deterministic function of forcing variable.
  - SRD is a special case of the FRD.
- Common use case: threshold allows participation in program.
  - Some might not participate even if allowed (noncompliance).
- Forcing variable is an **instrument**:
  - Affects Y_i, but only through D_i (at the threshold).

Fuzzy RD in a Graph (Imbens and Lemieux, 2008)

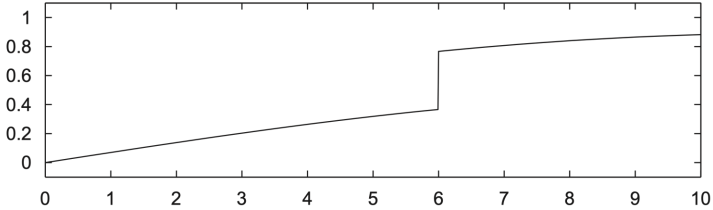


Fig. 3. Assignment probabilities (FRD).

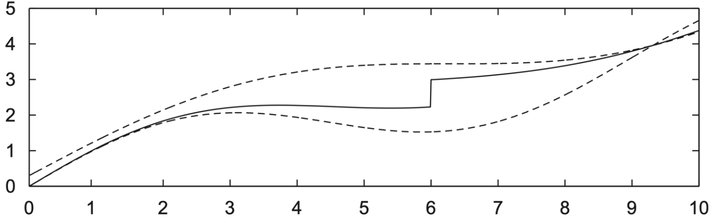


Fig. 4. Potential and observed outcome regression (FRD).

Source: *Figure 4* in Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142, no. 2 (2008): 615-635.

Fuzzy RD in a Graph (Imbens and Lemieux, 2008)

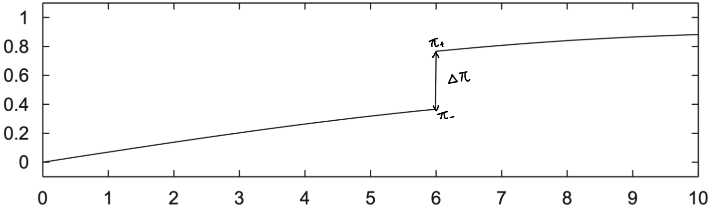


Fig. 3. Assignment probabilities (FRD).

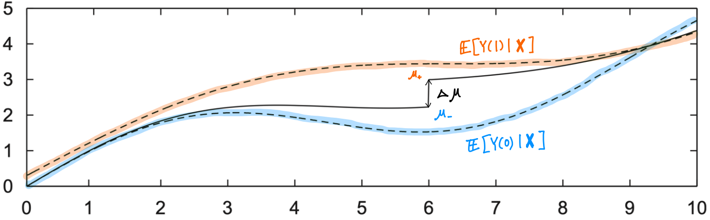


Fig. 4. Potential and observed outcome regression (FRD).

Source: *Figure 4* in Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142, no. 2 (2008): 615-635.

Fuzzy RD Assumptions

- $D_i(x)$ is potential value of treatment as cutoff changes around c.
  - $D_i(x) = 1$ if unit i would take treatment if cutoff were x.
  - $D_i(x) = 0$ if unit i would take control if cutoff were x.

Fuzzy RD Assumptions

- $D_i(x)$ is potential value of treatment as cutoff changes around c.
  - $D_i(x) = 1$ if unit i would take treatment if cutoff were x.
  - $D_i(x) = 0$ if unit i would take control if cutoff were x.
- **Monotonicity assumption:** $D_i(x)$ is non-increasing in x.
  - Lowering the cutoff can only increase participation.

Fuzzy RD Assumptions

- $D_i(x)$ is potential value of treatment as cutoff changes around c.
  - $D_i(x) = 1$ if unit i would take treatment if cutoff were x.
  - $D_i(x) = 0$ if unit i would take control if cutoff were x.
- **Monotonicity assumption:** $D_i(x)$ is non-increasing in x.
  - Lowering the cutoff can only increase participation.
- Compliers are those i such that for all $0 < e < \varepsilon$:

$$D_i(c - e) = 1 \quad \text{and} \quad D_i(c + e) = 0$$

  - Lowering or increasing the threshold would affect their treatment status.
  - Compliance status unobservable
    - $\rightsquigarrow$ Principal strata (Frangakis and Rubin, 2002. Biometrics).

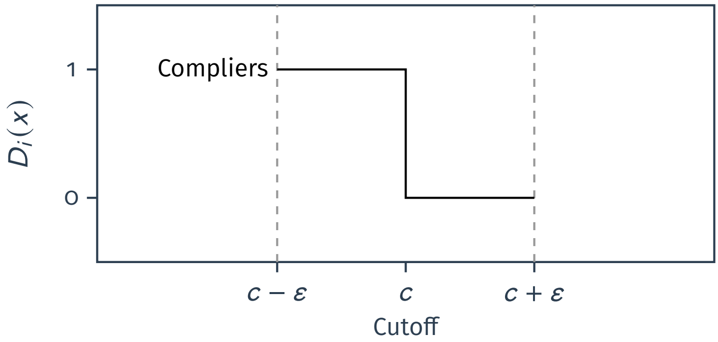
Fuzzy RD Assumptions

- $D_i(x)$ is potential value of treatment as cutoff changes around c.
  - $D_i(x) = 1$ if unit i would take treatment if cutoff were x.
  - $D_i(x) = 0$ if unit i would take control if cutoff were x.
- **Monotonicity assumption:** $D_i(x)$ is non-increasing in x.
  - Lowering the cutoff can only increase participation.
- Compliers are those i such that for all $0 < e < \varepsilon$:

$$D_i(c - e) = 1 \quad \text{and} \quad D_i(c + e) = 0$$

  - Lowering or increasing the threshold would affect their treatment status.
  - Compliance status unobservable
    - $\rightsquigarrow$ Principal strata (Frangakis and Rubin, 2002. Biometrics).
- Example: college students that get above a certain GPA are encouraged to apply to grad school.
  - Compliers wouldn't apply if threshold were slightly higher.
  - Compliers would apply if the threshold were slightly lower.

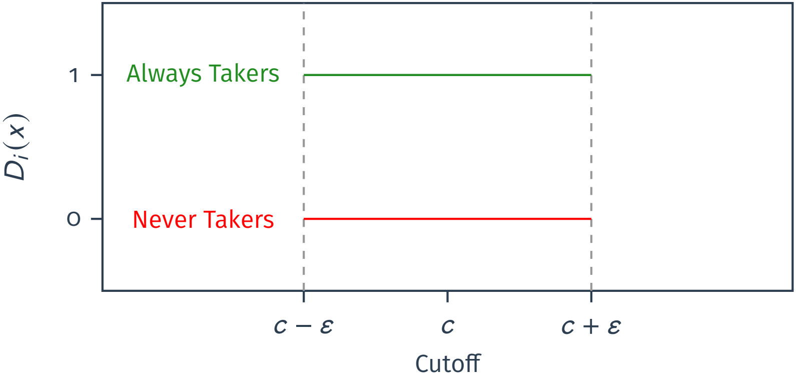
Compliance Graph



- Compliers would not take the treatment if they had $X_i = c$ and we increased the cutoff by some small amount.
- These are compliers at the threshold.

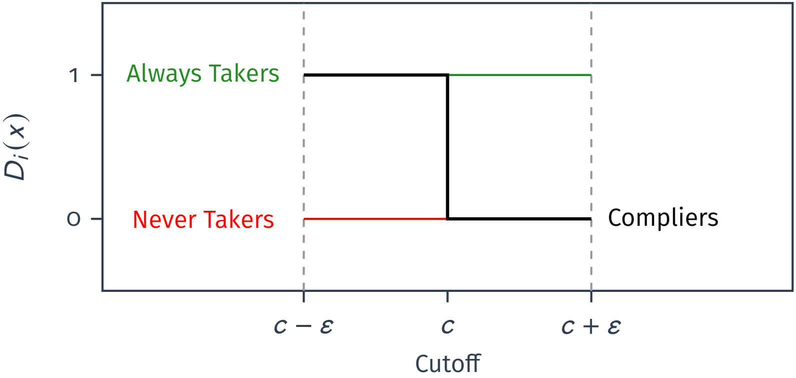
Compliance Groups

- Compliers: $D_i(c - e) = 1 \quad \text{and} \quad D_i(c + e) = 0$
- Always-takers: $D_i(c + e) = D_i(c - e) = 1$
- Never-takers: $D_i(c + e) = D_i(c - e) = 0$



Compliance Groups

- Compliers: $D_i(c-e) = 1$ and $D_i(c+e) = 0$
- Always-takers: $D_i(c+e) = D_i(c-e) = 1$
- Never-takers: $D_i(c+e) = D_i(c-e) = 0$



LATE in the Fuzzy RD

- We can define an estimator that is in the spirit of IV:

$$
\begin{aligned}
\tau_{\text{FRD}} &= \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i \mid X_i = x]} \\
&= \frac{\text{effect of threshold on } Y_i}{\text{effect of threshold on } D_i}
\end{aligned}
$$

LATE in the Fuzzy RD

- We can define an estimator that is in the spirit of IV:

$$\begin{aligned}\tau_{\text{FRD}} &= \frac{\lim_{x\downarrow c}\mathbb{E}[Y_i \mid X_i = x] - \lim_{x\uparrow c}\mathbb{E}[Y_i \mid X_i = x]}{\lim_{x\downarrow c}\mathbb{E}[D_i \mid X_i = x] - \lim_{x\uparrow c}\mathbb{E}[D_i \mid X_i = x]} \\ &= \frac{\text{effect of threshold on } Y_i}{\text{effect of threshold on } D_i}\end{aligned}$$

- Under the FRD assumptions (continuity, consistency, and monotonicity), we can write that the estimator is equal to the effect at the threshold for compliers.

$$\tau_{\text{FRD}} = \mathbb{E}[\tau_i \mid i \text{ is a complier},\ X_i = c]$$

  - Proof is very similar to the LATE proof.

LATE in the Fuzzy RD

- We can define an estimator that is in the spirit of IV:

$$\tau_{\text{FRD}} = \frac{\lim_{x\downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x\downarrow c} \mathbb{E}[D_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[D_i \mid X_i = x]}$$
$$= \frac{\text{effect of threshold on } Y_i}{\text{effect of threshold on } D_i}$$

- Under the FRD assumptions (continuity, consistency, and monotonicity), we can write that the estimator is equal to the effect at the threshold for compliers.

$$\tau_{\text{FRD}} = \mathbb{E}[\tau_i \mid i \text{ is a complier}, \; X_i = c]$$

  - Proof is very similar to the LATE proof.

- External validity? Doubly local $\rightsquigarrow$ careful about generalizing.

Estimation in FRD

- Remember that we had:

$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i \mid X_i = x]}$$

Estimation in FRD

- Remember that we had:

$$\tau_{\text{FRD}} = \frac{\lim_{x\downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x\downarrow c} \mathbb{E}[D_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[D_i \mid X_i = x]}$$

- Ratio of SRD estimands: use local linear regression for both.

$$\widehat{\tau}_{\text{FRD}} = \frac{\widehat{\tau}_{Y,\text{SRD}}}{\widehat{\tau}_{D,\text{SRD}}}$$

Estimation in FRD

- Remember that we had:

$$\tau_{\text{FRD}} = \frac{\lim_{x\downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x\downarrow c} \mathbb{E}[D_i \mid X_i = x] - \lim_{x\uparrow c} \mathbb{E}[D_i \mid X_i = x]}$$

- Ratio of SRD estimands: use local linear regression for both.

$$\widehat{\tau}_{\text{FRD}} = \frac{\widehat{\tau}_{Y,\text{SRD}}}{\widehat{\tau}_{D,\text{SRD}}}$$

- Again, CCT provides robust bias correction, bandwidths.
    - Calonico, Cattaneo, and Titiunik (CCT; Econometrica 2014)

More Practical FRD Estimation

- The ratio estimator above is equivalent to a TSLS approach.
- Use the same specification as above with the following covariates:

$$V_i = \begin{pmatrix} 1 \\ \mathbf{1}\{X_i < c\}(X_i - c) \\ \mathbf{1}\{X_i \geq c\}(X_i - c) \end{pmatrix}$$

More Practical FRD Estimation

- The ratio estimator above is equivalent to a TSLS approach.
- Use the same specification as above with the following covariates:

$$V_i = \begin{pmatrix} 1 \\ \mathbf{1}\{X_i < c\}(X_i - c) \\ \mathbf{1}\{X_i \geq c\}(X_i - c) \end{pmatrix}$$

- First stage:

$$D_i = \delta_1' V_i + \rho \mathbf{1}\{X_i \geq c\} + \nu_i$$

- Second stage:

$$Y_i = \delta_2' V_i + \tau D_i + \eta_i$$

More Practical FRD Estimation

- The ratio estimator above is equivalent to a TSLS approach.
- Use the same specification as above with the following covariates:

$$V_i = \begin{pmatrix} 1 \\ \mathbf{1}\{X_i < c\}(X_i - c) \\ \mathbf{1}\{X_i \geq c\}(X_i - c) \end{pmatrix}$$

- First stage:

$$D_i = \delta_1' V_i + \rho \mathbf{1}\{X_i \geq c\} + \nu_i$$

- Second stage:

$$Y_i = \delta_2' V_i + \tau D_i + \eta_i$$

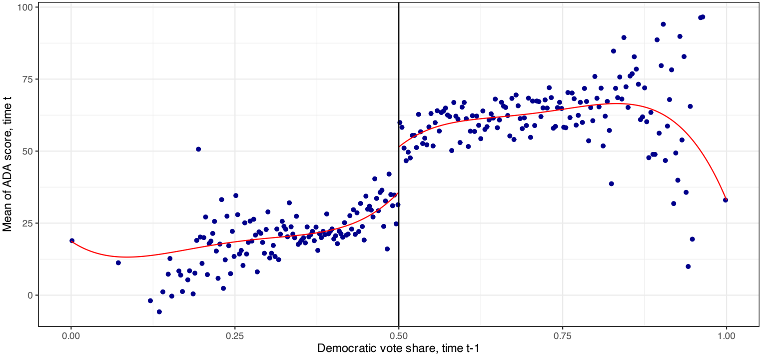
- Thus, being above the threshold is treated like an instrument, controlling for trends in X_i.

4/ R Codes for SRD and FRD

Visualization

- Visualization: showing discontinuity at the cut-off

```r
> pacman::p_load(tidyverse, rdrobust, rddensity, ggplot2)
> lmb_data <- read_csv("https://bit.ly/41LCnht")  ## Import data

> rdplot(y = lmb_data$score, x = lmb_data$lagdemvoteshare,  ## Visualize the discontinuity
        c = .5, title = "",
        x.label = "Democratic vote share, time t-1",
        y.label = "Mean of ADA score, time t")
```



Estimation

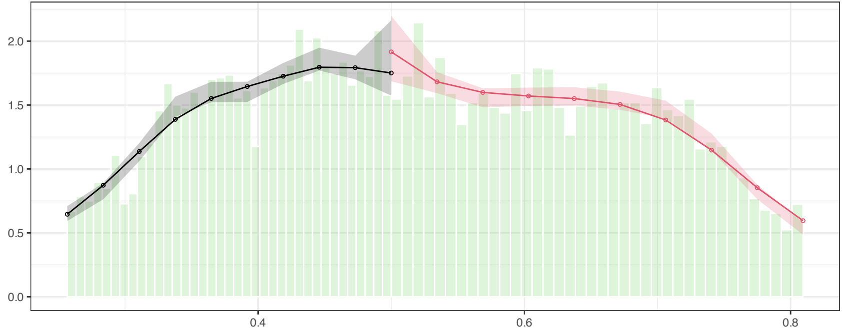
```
# Fit local linear regression
> fit <- rdrobust::rdrobust(y = lmb_data$score, x = lmb_data$lagdemvoteshare, c = .5)
> cbind(fit$coef, fit$se)

                   Coeff  Std. Err.
Conventional   18.66595  1.692466
Bias-Corrected 18.44932  1.692466
Robust         18.44932  2.037871
```

- Two types of point estimates:
  - The standard local linear estimator, $\widehat{\tau}_{\text{srd}}$
  - The local linear estimator with bias-correction, $\widehat{\tau}_{\text{BC}} = \widehat{\tau}_{\text{srd}} - \widehat{\text{bias}}$ (e.g., CCT 2014)
- Two standard errors:
  - Standard SE, $\widehat{\sigma}^2$
  - "Robust" SE, $\widehat{\sigma}^2_{\text{robust}}$, accounts for uncertainty in bias estimation.
- $\rightsquigarrow$ We report the "Robust" estimate: $\widehat{\tau}_{\text{BC}}$ with $\widehat{\sigma}^2_{\text{robust}}$.

Diagnostics: No sorting?

- McCrary density test:

```
test1 <- rddensity::rddensity(lmb_data$demvoteshare, c = .5)
rdplotdensity(rdd = test1,
              X = lmb_data$demvoteshare,
              type = "both") # lines? points? both?
```

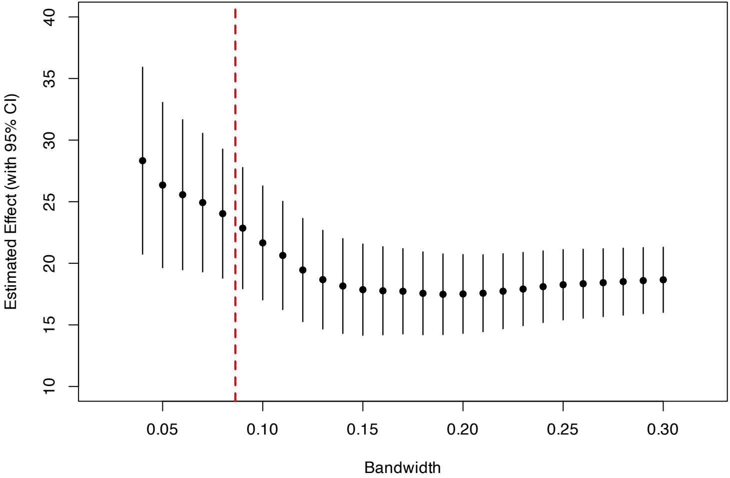


Estimated Effects Along Different Bandwidths

- Demonstrate how the results change w/ different bandwidths:

```
# fit local linear regression with bandwidth bws[b]
bws <- seq(.04, .3, by = .01); fits <- list()
b = 1
for (b in 1:length(bws)) {
  fits[[b]] <- rdrobust(y = lmb_data$score,
                        x = lmb_data$lagdemvoteshare,
                        c = .5, h = bws[b])
}; fits

# summarize result (use "robust")
plot(1, 1, type = 'n', xlim = c(0.02,.32), ylim = c(10,40),
     xlab = 'Bandwidth', ylab = 'Estimated Effect (with 95% CI)')

for (b in 1:length(bws)) {
  points(x = bws[b], y = fits[[b]]$coef[3], pch = 16)
  lines(c(bws[b], bws[b]), fits[[b]]$ci[3,], lwd = 1.2)
}
# abline(v = fit$bws[1,1], col = 'red', lwd = 1.5)
abline(v = fit$bws[1,1], col = 'red', lwd = 2, lty = "dashed")
```

Bandwidth Robustness Check



Free Tutoring Program and Academic Performance

- The Setting:
    - Students take an entrance exam at the beginning & end of a school year.
    - Those who scored below 70 are enrolled in a free tutoring program.

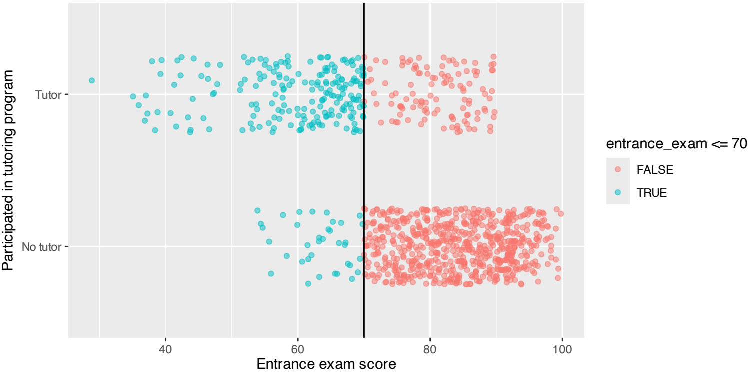
Name	Description
id	student id
entrance_exam	student's entrance exam score (out of 100)
exit_exam	student's exit exam score (out of 100)
tutoring	indicator showing if the student received tutoring

```
> pacman::p_load(tidyverse, broom, modelsummary, kableExtra, estimatr, rdrobust)
> tutoring <- read_csv("https://bit.ly/453ymbJ"); head(tutoring, 3)

# A tibble: 4 × 5
     id entrance_exam tutoring tutoring_text exit_exam
  <dbl>         <dbl> <lgl>    <chr>             <dbl>
1     1          92.4 FALSE    No tutor           78.1
2     2          72.8 FALSE    No tutor           58.2
3     3          53.7 TRUE     Tutor              62.0
```

Examine Compliance Around the Cutoff

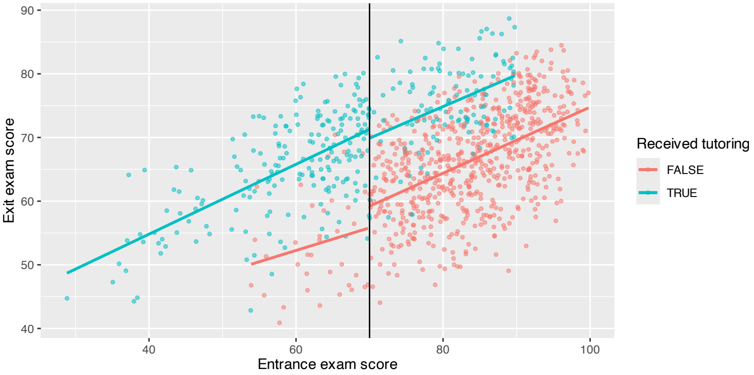
```
# Noncompliance around the cutoff
> tutoring |>
    group_by(tutoring, entrance_exam <= 70) |>
    summarize(count = n()) |>
    group_by(tutoring) |>
    mutate(prop = count / sum(count))

# A tibble: 4 × 4
# Groups:   tutoring [2]
  tutoring `entrance_exam <= 70` count   prop
  <lgl>    <lgl>                 <int>  <dbl>
1 FALSE    FALSE                   646 0.947
2 FALSE    TRUE                     36 0.0528
3 TRUE     FALSE                   116 0.365
4 TRUE     TRUE                    202 0.635

# Visualize Noncompliance
> ggplot(tutoring, aes(x = entrance_exam, y = tutoring_text, color = entrance_exam <= 70)) +
    geom_point(size = 1.5, alpha = 0.5,
               position = position_jitter(width = 0, height = 0.25, seed = 1234)) +
    geom_vline(xintercept = 70) +
    labs(x = "Entrance exam score", y = "Participated in tutoring program")
```

Visualizing Noncompliance around the Cutoff



Visualizing the Fuzzy Gap

```
> ggplot(tutoring, aes(x = entrance_exam, y = tutoring_text, color = entrance_exam <= 70)) +
    geom_point(size = 1.5, alpha = 0.5,
               position = position_jitter(width = 0, height = 0.25, seed = 1234)) +
    geom_vline(xintercept = 70) +
    labs(x = "Entrance exam score", y = "Participated in tutoring program")
```



Measuring the Fuzzy Gap

- Center the forcing variable, entrance_score and generate V_i:

```
> tutoring_centered <- tutoring %>%
    mutate(entrance_centered = entrance_exam - 70,
           entrance_cnt_b70 = if_else(entrance_exam < 70, entrance_centered, 0),
           entrance_cnt_a70 = if_else(entrance_exam >= 70, entrance_centered, 0),
           below_cutoff = entrance_exam <= 70)
```

- If this was a Sharp RD:

```
# Bandwidth ±10
> sharp_parametric <- lm(exit_exam ~ entrance_centered + tutoring,
                         data = filter(tutoring_centered,
                                       entrance_centered >= -10 &
                                         entrance_centered <= 10))
> tidy(sharp_parametric)
# A tibble: 3 × 5
  term              estimate std.error statistic   p.value
  <chr>                <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)         59.3      0.503     118.   9.75e-313
2 entrance_centered    0.511    0.0665      7.69 1.17e- 13
3 tutoringTRUE        11.5      0.744      15.4  1.77e- 42
```

  - tutoringTRUE estimate measures the size of the jump at the cutoff.
    - There would still be confounding as we have noncompliance.

Measuring the Fuzzy Gap Correctly: TSLS

- Use `estimatr::iv_robust()` for TSLS estimation:

```
> model_fuzzy <- iv_robust(
    exit_exam ~ entrance_cnt_b70 + entrance_cnt_a70 + tutoring |
      entrance_cnt_b70 + entrance_cnt_a70 + below_cutoff,
    data = filter(tutoring_centered, entrance_centered >= -10 & entrance_centered <= 10)
  )

> tidy(model_fuzzy, conf.int = F)
              term   estimate std.error statistic       p.value  df   outcome
1      (Intercept) 59.9731656 1.1165541 53.712726 9.984384e-185 399 exit_exam
2 entrance_cnt_b70  0.3773301 0.1855677  2.033383  4.267520e-02 399 exit_exam
3 entrance_cnt_a70  0.4715826 0.1344630  3.507156  5.044279e-04 399 exit_exam
4     tutoringTRUE  9.6265535 1.9424192  4.955961  1.065759e-06 399 exit_exam
```

Measuring the Fuzzy Gap Correctly: rdrobust

- Nonparametric estimation for FRD with `rdrobust()`:

```
> frd <- rdrobust::rdrobust(y = tutoring$exit_exam, x = tutoring$entrance_exam,
                     c = 70, fuzzy = tutoring$tutoring, kernel = "triangular", vce ="hc2")
> summary(frd)

Fuzzy RD estimates using local polynomial regression.

Number of Obs.                 1000
BW type                       mserd
Kernel                   Triangular
VCE method                      HC2

Number of Obs.                  238          762
Eff. Number of Obs.             170          347
Order est. (p)                    1            1
Order bias  (q)                   2            2
BW est. (h)                  12.961       12.961
BW bias (b)                  19.579       19.579
rho (h/b)                     0.662        0.662
Unique Obs.                     238          762

First-stage estimates.

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional    -0.708     0.073    -9.672     0.000    [-0.851 , -0.565]
        Robust         -         -    -8.350     0.000    [-0.909 , -0.563]
=============================================================================

Treatment effect estimates.

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     9.685     1.957     4.948     0.000     [5.849 , 13.522]
        Robust         -         -     4.127     0.000     [5.075 , 14.257]
=============================================================================
```

Measuring the Fuzzy Gap Correctly: Robust Estimates

- Report the "robust" estimates: $\widehat{\tau}_{\text{BC}}$ with $\widehat{\sigma}^2_{\text{robust}}$.

```
> cbind(frd$coef, frd$se)

                  Coeff Std. Err.
Conventional   9.685196  1.957433
Bias-Corrected 9.666085  1.957433
Robust         9.666085  2.342204
```

On to the Presentations & Discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/

Appendix

Bandwidths and Bias (Details)

- Bias of local linear estimator scales as h^2; SE scales as $1/\sqrt{nh}$.
- MSE-optimal rate *balances* the two: $h_n \propto n^{-1/5}$.
  - At this rate, Bias $\sim n^{-2/5}$ and SE $\sim n^{-2/5}$, i.e. *same asymptotic order*.
  - Standard normal CI is shifted by the residual bias $\Rightarrow$ undercoverage. Two fixes:
- **Undersmoothing**: shrink bandwidth faster, e.g. $h_n \propto n^{-1/4}$.
  - At this rate, Bias/SE $\sim n^{-1/8} \longrightarrow 0$, so bias is asymptotically negligible relative to SE.
  - Problem: most data-driven selectors target the MSE-optimal rate and return a bandwidth too big for valid undersmoothing inference. Bias strikes back.
- **Robust bias correction**: $\widehat{\tau}_{\text{SRD}}^{\text{rbc}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$.
  - Calonico, Cattaneo, & Titiunik (CCT, 2014, *Econometrica*) gives the form.
  - Allows the use of optimal bandwidths, but need to account for estimation of bias.
  - Bias estimation comes from using higher-order polynomial regression.
- Coverage of CIs can be very bad without RBC.

Selecting the Optimal Bandwidth (Details)

- Let $\mathcal{B}$ and $\mathcal{V}$ be approximations of the bias and variance of $\widehat{\tau}_{\text{SRD}}(h)$.
  - Based on quadratic approximation of $\mu_d(x)$ rather than linear.
- Idea: find the bandwidth that minimizes the estimation error.

$$\text{MSE}(h) = \mathbb{E}\left[(\widehat{\tau}(h) - \tau_{\text{SRD}})^2 \,\middle|\, X_1, \ldots, X_n\right] \approx h^4\mathcal{B}^2 + \frac{1}{nh}\mathcal{V}$$

  - Optimal bandwidth: $h_{\text{MSE}} = \left(\frac{\mathcal{V}}{4\mathcal{B}^2}\right)^{1/5} n^{-1/5}$.
  - These depend on unknown biases/variances.
- Plug-in procedure (CCT):
  1. Pick initial bandwidths to estimate $\mathcal{B}$ and $\mathcal{V}$ with local quadratic regression.
  2. Pick optimal bandwidth for the bias-correction term and estimate bias with local quadratic regression.
  3. Use both steps to pick optimal bandwidth for local linear regression (h_n).

Kink RD

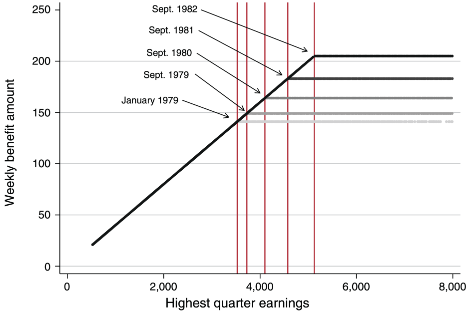


FIGURE 2. LOUISIANA: SCHEDULE OF UI WEEKLY BENEFIT AMOUNT, JAN. 1979–DEC. 1983

Source: https://www.aeaweb.org/articles?id=10.1257/pol.20130248

Also see: https://blogs.worldbank.org/en/impactevaluations/tools-trade-regression-kink-design

- **Sharp Kink RD**: discontinuities in the first derivatives rather than levels.
    - Unemployment benefits as a function of prior earnings.
    - If there is a cap on benefits, there's a kink in the assignment.
    - Look for changes in the slope of $\mathbb{E}[Y_i \mid X_i = x]$ at threshold.
    - Estimation similar, but better to use local quadratic regression.