# 12. Regression Discontinuity Designs

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

# Where are we? Where are we going?

- So far:
    - Randomized experiments identify causal effects.
    - Regression, matching, weighting for selection on observables.
    - Instrumental variables for when this doesn't hold.

# Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting for selection on observables.
  - Instrumental variables for when this doesn't hold.
- 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.

# Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting for selection on observables.
  - Instrumental variables for when this doesn't hold.
- 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.
- 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**: $D_i = 1\{X_i \geq c\} \ \ \forall\ i$
  - 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)

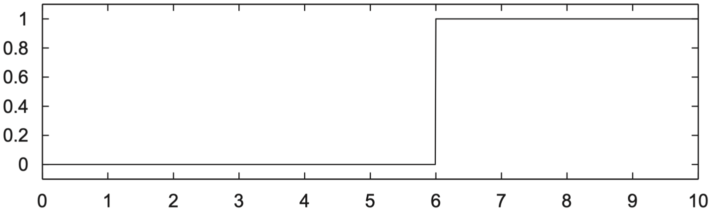



Fig. 1. Assignment probabilities (SRD).

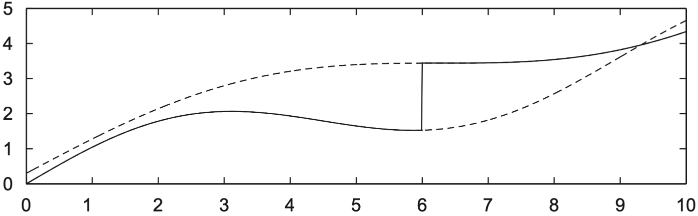



Fig. 2. Potential and observed outcome regression functions.

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)

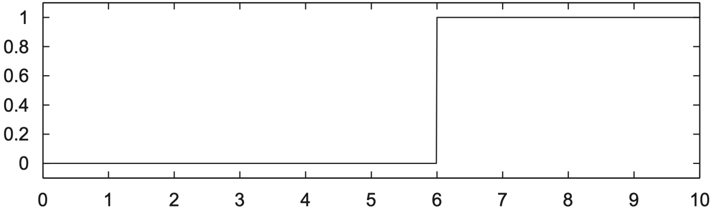



Fig. 1. Assignment probabilities (SRD).

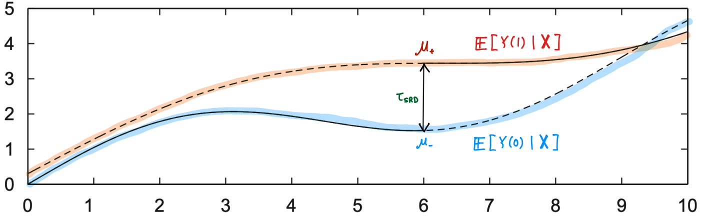



Fig. 2. Potential and observed outcome regression functions.

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:

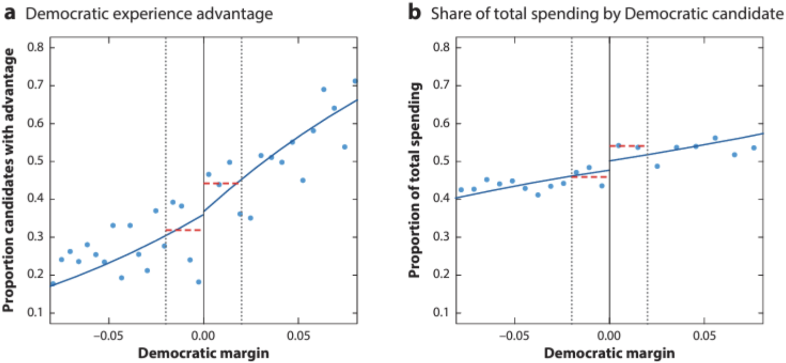
$$\begin{aligned} \tau_{\text{SRD}} &= \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c] \\ &= \mathbb{E}[Y_i(1) | X_i = c] - \mathbb{E}[Y_i(0) | X_i = c] \\ &= \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x] \end{aligned}$$

- 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

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

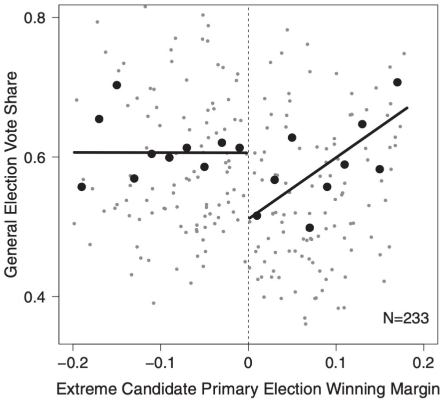
    - where $b_k$ are the bin cutpoints.
    - $n_k$ is the number of units within bin $k$.

- 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

**FIGURE 2. General-Election Vote Share After Close Primary Elections Between Moderates and Extremists: U.S. House, 1980–2010**




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

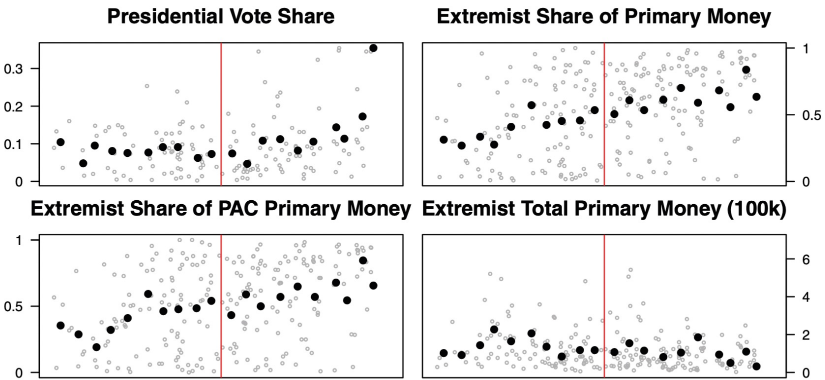
- Also good to include binned mean plots for pretreatment covariates.
- Intuition: key assumption in smoothness in the mean of $Y_i(d)$ in $X_i$.

- 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

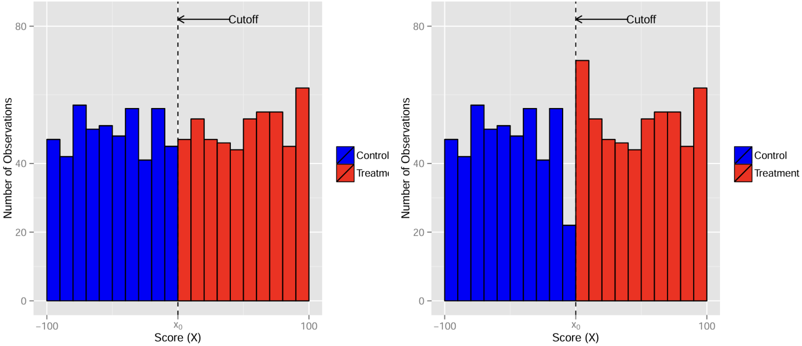
**FIGURE A.2. Graphical Balance Tests**




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: Skovron, Christopher, and Rocıo Titiunik. ”A practical guide to regression discontinuity designs in political science.” American Journal of Political Science 2015 (2015): 1-36.

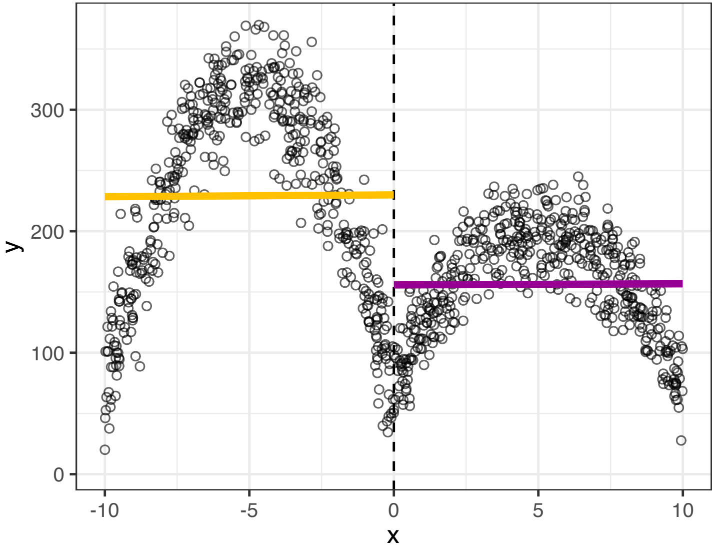
# General Estimation Strategy

- 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:

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

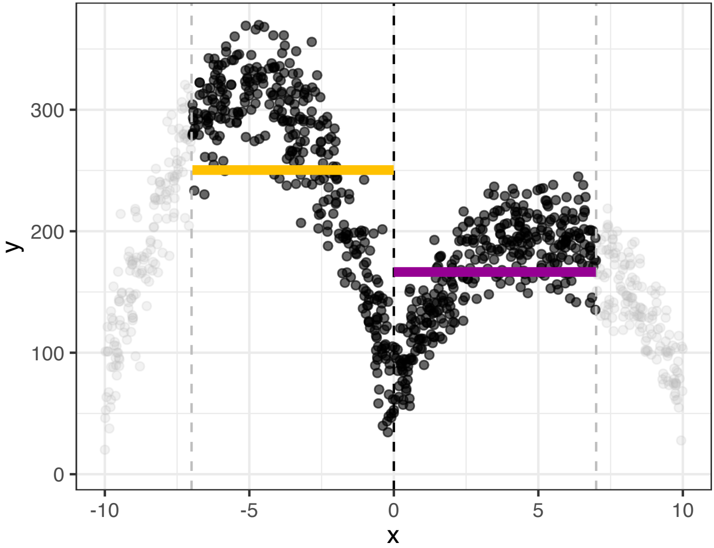
$$\mu_-(x) = \lim_{z \uparrow x} \mathbb{E}[Y_i(0) \mid X_i = z]$$

- For the SRD, we have $\tau_{\text{SRD}} = \mu_+(c) - \mu_-(c)$.
- Kernel regression with **uniform kernel**:

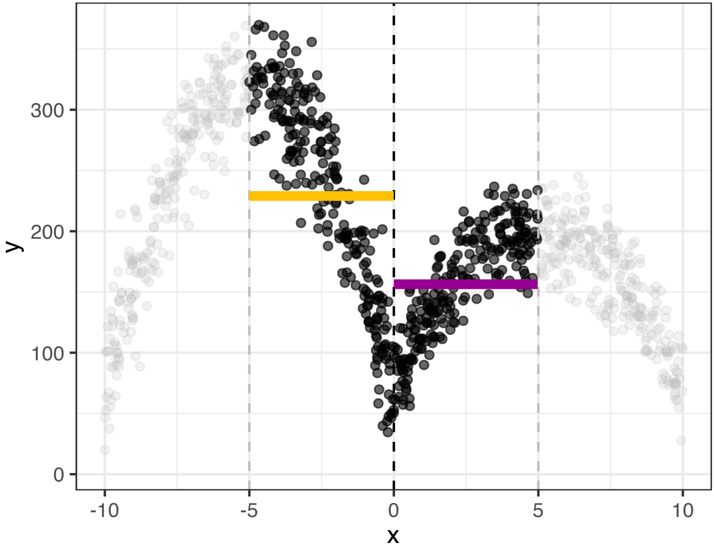
$$\widehat{\mu}_-(c) = \frac{\sum_{i=1}^N Y_i \cdot \mathbf{1}\{c - h \leq X_i < c\}}{\sum_{i=1}^N \mathbf{1}\{c - h \leq X_i < c\}}$$

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

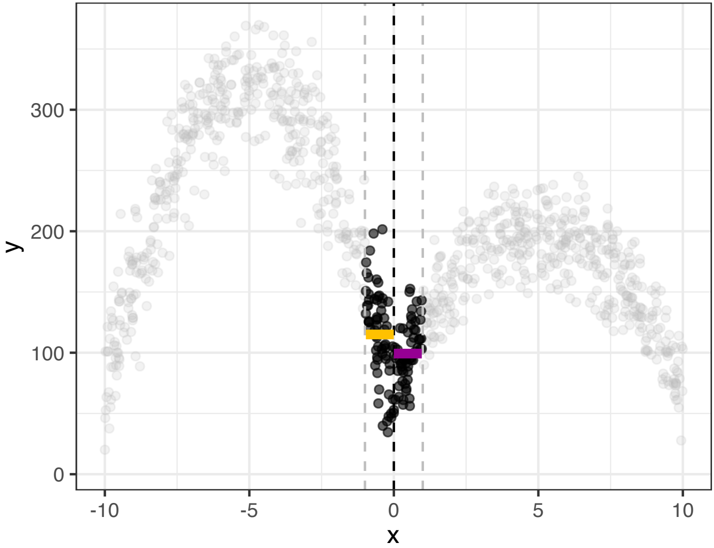
# Bandwidth Equal to 7



# 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 $c$:

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

  - 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)

# 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:

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

  - $\widehat{\tau}_{\text{SRD}} = \widehat{\tau}$ 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:

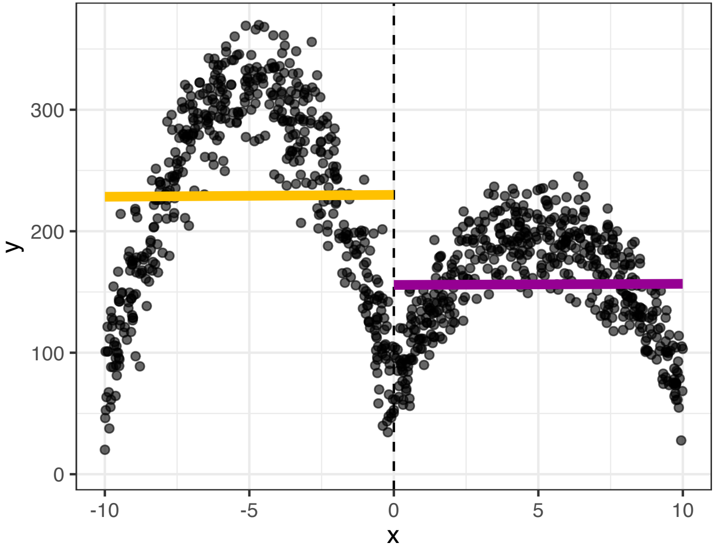
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i:X_i \in [c-h,c+h]} \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

  - $\widehat{\tau}_{\text{SRD}} = \widehat{\tau}$ 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.

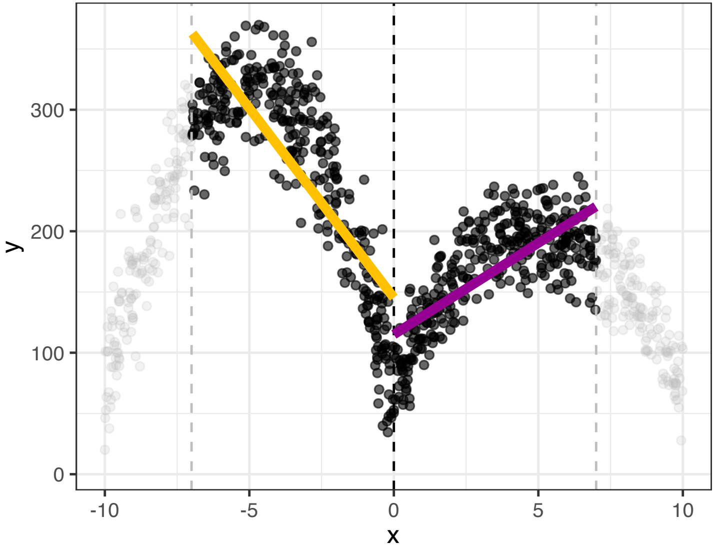
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

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

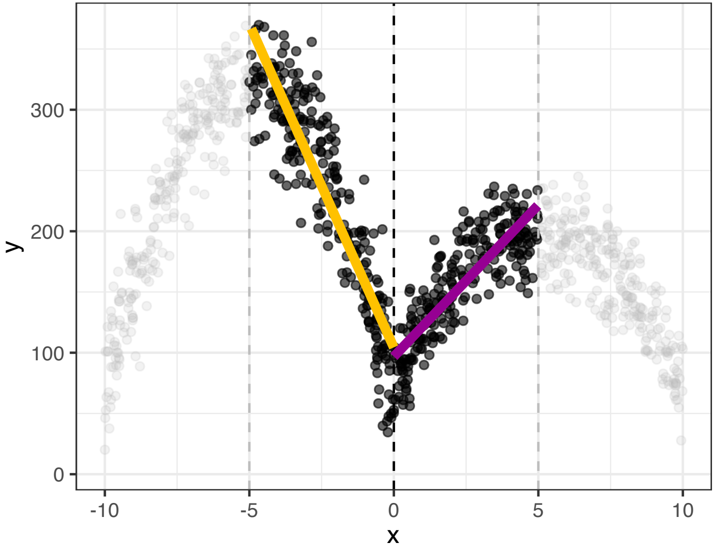
# Bandwidth Equal to 10 (Global)



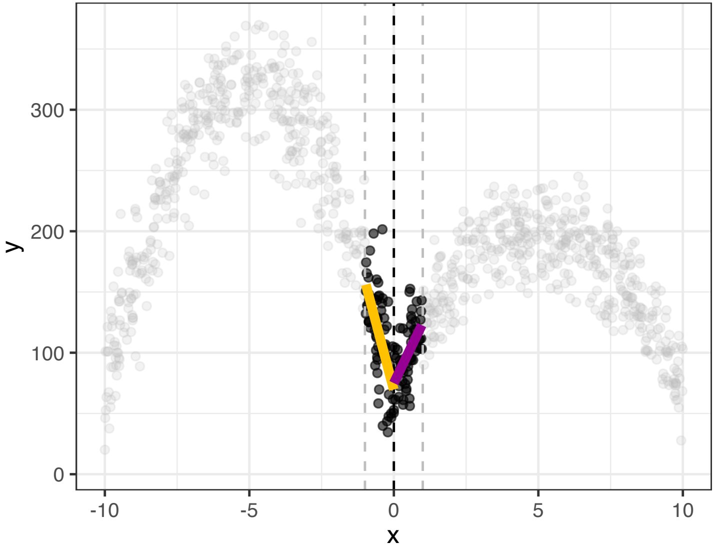
# Bandwidth Equal to 7



# Bandwidth Equal to 5



# Bandwidth Equal to 1



# Odds and Ends for the SRD

Practical considerations:

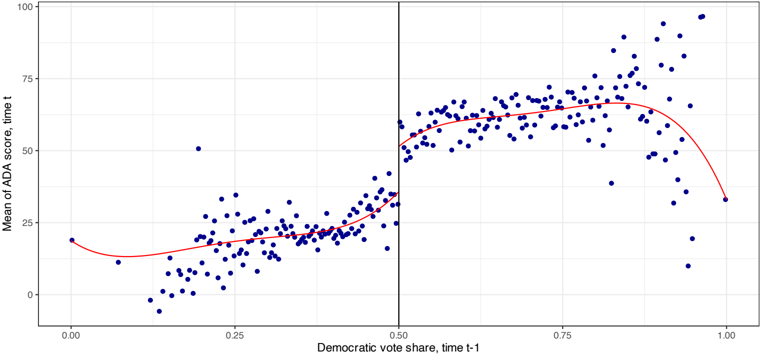
- **Standard errors:** robust standard errors from local OLS are valid.
  - Not great in finite samples b/c the bandwidth isn't designed for this purpose.
  - CCT derives a variance estimator that has better coverage.
  - Calonico, Cattaneo, and Titiunik (CCT; Econometrica 2014)
- **Covariates:** can 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/ Example R Codes

# Visualization

- Visualization: showing discontinuity at the cut-off

```
> 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

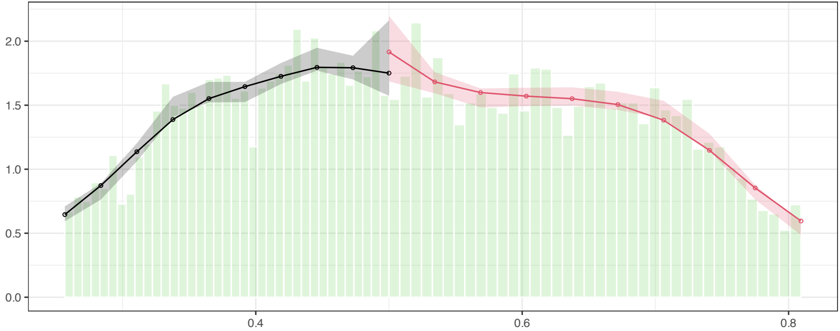
```
# Fit local linear regression
> fit <- rdrobust::rdrobust(y = lmb_data$score, x = lmb_data$lagdemvoteshare, c = .5)
> cbind(fit$coef, fit$se)

                    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:

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

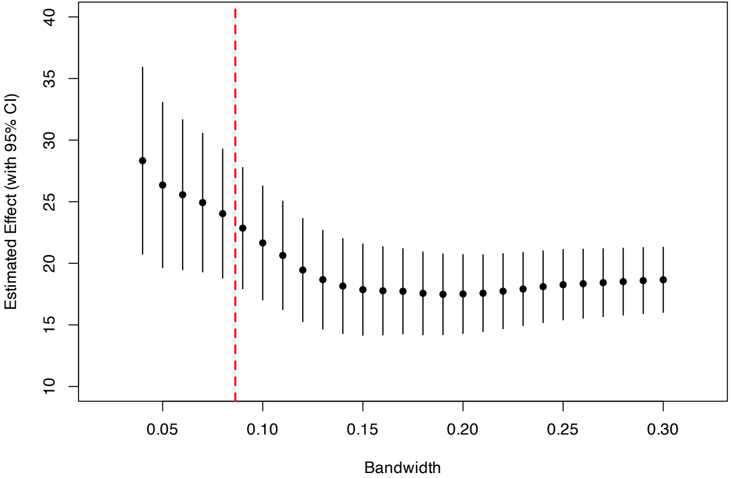


# Estimated Effects Along Different Bandwidths

- 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



## Have a great weekend!

*Contact Information:*
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

# Appendix

# Bandwidths and bias

- Optimal bandwidth shrinks fast enough so $h_n \propto n^{-1/5}$.
  - But this results in asymptotic bias, two possible solutions.
- **Undersmoothing**: have bandwidth shrink more quickly e.g. $h_n \propto n^{-1/4}$
  - Smaller bandwidths $\rightsquigarrow$ less bias.
  - Problem: most ways of actually selecting the optimal bandwidth will be too big. Bias strikes back.
- **Robust bias correction**: $\widehat{\tau}_{\text{SRD}}^{\text{rbc}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$
  - Calonico, Cattaneo, and 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 polynomials regression.
- Coverage of CIs can be very bad without RBC!

# Selecting the optimal bandwidth

- 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.

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

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