Module 8: Regression Discontinuity Designs

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

Where are we? Where are we going?

- So far:

Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.

Where are we? Where are we going?

- So far:
  - Randomized experiments identify causal effects.
  - Regression, matching, weighting for selection on observables.

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 exogeneous variation in the treatment assignment

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 exogeneous variation in the treatment assignment
  - RCT: randomization provides exogeneous 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 exogeneous variation in the treatment assignment
  - RCT: randomization provides exogeneous variation.
  - Selection on observables: treatment as-if random conditional on X_i

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 exogeneous variation in the treatment assignment
    - RCT: randomization provides exogeneous variation.
    - Selection on observables: treatment as-if random conditional on X_i
    - IV: instrument provides exogeneous 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 exogeneous variation in the treatment assignment
  - RCT: randomization provides exogeneous variation.
  - Selection on observables: treatment as-if random conditional on X_i
  - IV: instrument provides exogeneous variation
- Regression discontinuity: a discontinuity in treatment assignment

Plan of attack

1. Sharp Regression Discontinuity Designs

2. Estimation in the SRD

3. Bandwidth selection

4. Fuzzy Regression Discontinuity Designs

1/ Sharp Regression Discontinuity Designs

Setup

- The basic idea behind RDDs:

Setup

- The basic idea behind RDDs:
  - Treatment assignment is determined by a cutoff in some variable X_i.

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**

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,

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.

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 threshold have a causal interpretation

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 threshold have a causal interpretation
- The classic example of this is in the educational context:

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 threshold have a causal interpretation
- The classic example of this is in the educational context:
  - Scholarships allocated based on a test score threshold (Thistlethwaite and Campbell, 1960)

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 threshold have a causal interpretation
- The classic example of this is in the educational context:
  - Scholarships allocated based on a test score threshold (Thistlethwaite and Campbell, 1960)
  - Class size on test scores using total student thresholds to create new classes (Angrist and Lavy, 1999)

Sharp RD

- Notation

Sharp RD

- Notation
  - Treatment: $D_i = 1$ or $D_i = 0$

Sharp RD

- Notation
  - Treatment: $D_i = 1$ or $D_i = 0$
  - Potential outcomes, $Y_i(1)$ and $Y_i(0)$

Sharp RD

- Notation
  - Treatment: $D_i = 1$ or $D_i = 0$
  - Potential outcomes, $Y_i(1)$ and $Y_i(0)$
  - Observed outcomes: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$

Sharp RD

- Notation
  - 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

- Notation
  - 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**: $D_i = 1\{X_i \geq c\} \ \forall i$

Sharp RD

- Notation
  - 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.

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship
  - When test scores are below 1500 $\rightarrow$ not offered scholarship

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship
  - When test scores are below 1500 $\rightarrow$ not offered scholarship
- Note: positivity violated by assumption here

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship
  - When test scores are below 1500 $\rightarrow$ not offered scholarship
- Note: positivity violated by assumption here
  - $\mathbb{P}[D_i = 1 \mid X_i = c - \varepsilon] = 0$

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship
  - When test scores are below 1500 $\rightarrow$ not offered scholarship
- Note: positivity violated by assumption here
  - $\mathbb{P}[D_i = 1 \mid X_i = c - \varepsilon] = 0$
  - $\mathbb{P}[D_i = 1 \mid X_i = c + \varepsilon] = 1$

Sharp RD

- Notation
  - 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 1500 $\rightarrow$ offered scholarship
  - When test scores are below 1500 $\rightarrow$ not offered scholarship
- Note: positivity violated by assumption here
  - $\mathbb{P}[D_i = 1 \mid X_i = c - \varepsilon] = 0$
  - $\mathbb{P}[D_i = 1 \mid X_i = c + \varepsilon] = 1$
  - Can't use standard identification toolkit for ATE/ATT.

Plotting the RDD (Imbens and Lemieux, 2008)

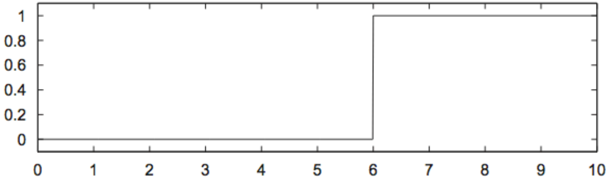


Fig. 1. Assignment probabilities (SRD).

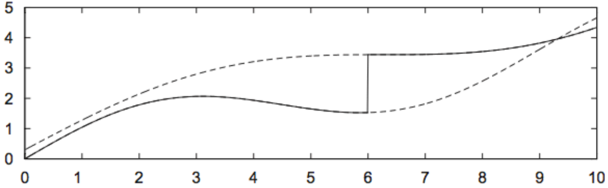


Fig. 2. Potential and observed outcome regression functions.

Quantity of interest

- Quantity of interest: **local** average treatment effect at the cutoff

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

Quantity of interest

- Quantity of interest: **local** average treatment effect at the cutoff

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

  - Very difficult to extrapolate beyond this.

Quantity of interest

- Quantity of interest: **local** average treatment effect at the cutoff

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

    - Very difficult to extrapolate beyond this.
- Problem: X_i is continuous so we never observe any $X_i = c$.

Quantity of interest

- Quantity of interest: **local** average treatment effect at the cutoff

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

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

Quantity of interest

- Quantity of interest: **local** average treatment effect at the cutoff

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

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

Continuity of the CEFs

- **Assumption**: CEFs of potential outcomes are **continuous** in X_i

Continuity of the CEFs

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

Continuity of the CEFs

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

Continuity of the CEFs

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

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

Continuity of the CEFs

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

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

Continuity of the CEFs

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

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

Continuity of the CEFs

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

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

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

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

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c]$$

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

- Problem: estimate two regression functions at a point.

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

- Problem: estimate two regression functions at a point.
    - Without parametric assumptions, very hard!

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

- Problem: estimate two regression functions at a point.
    - Without parametric assumptions, very hard!
    - Nonparametric regression can be consistent, but converge is slow and

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

- Problem: estimate two regression functions at a point.
    - Without parametric assumptions, very hard!
    - Nonparametric regression can be consistent, but converge is slow and
- NB: not equivalent to **local randomization**,

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

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

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

- Problem: estimate two regression functions at a point.
  - Without parametric assumptions, very hard!
  - Nonparametric regression can be consistent, but converge is slow and
- NB: not equivalent to **local randomization**,

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

  - LR stronger than continuity because it rules out confounding around c

Identification results

- Consistency + SRD Assumption + Continuity $\rightsquigarrow$ identification:

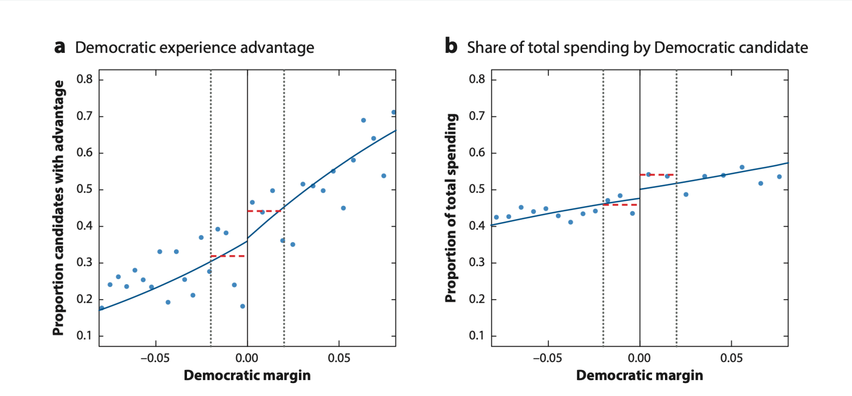
$$\begin{aligned}\tau_{SRD} &= E[Y_i(1) - Y_i(0) | X_i = c] \\ &= E[Y_i(1) | X_i = c] - E[Y_i(0) | X_i = c] \\ &= \lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]\end{aligned}$$

- Problem: estimate two regression functions at a point.
  - Without parametric assumptions, very hard!
  - Nonparametric regression can be consistent, but converge is slow and
- NB: not equivalent to **local randomization**,

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

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

Problems with local randomization assumptions



What can go wrong?

- Key question: why is there a discontinuity in D_i but not $Y_i(d)$?

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?

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 65 age cutoff for RDD of AARP membership?

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 65 age cutoff for RDD of AARP membership?
- **Sorting around the threshold**: possible violation of smoothness.

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 65 age cutoff for RDD of AARP membership?
- **Sorting around the threshold**: possible violation of smoothness.
    - Students retaking exams to pass some threshold for financial aid.

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 65 age cutoff for RDD of AARP membership?
- **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

- Binned means plot is very standard:

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

Bin plots

- Binned means plot is very standard:

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

  - b_k are the bin cutpoints.

Bin plots

- Binned means plot is very standard:

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

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

Bin plots

- Binned means plot is very standard:

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

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

- What to observe:

Bin plots

- Binned means plot is very standard:

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

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

- What to observe:

  - Obvious discontinuity at the threshold?

Bin plots

- Binned means plot is very standard:

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

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

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

Bin plots

- Binned means plot is very standard:

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

  - b_k are the bin cutpoints.
  - n_k is the number of units within bin k.
- What to observe:
  - Obvious discontinuity at the threshold?
  - Are there other, unexplained discontinuities?
- Very difficult to sell an RDD without visually obvious result.

Bin plots

- Binned means plot is very standard:

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

  - b_k are the bin cutpoints.
  - n_k is the number of units within bin k.
- What to observe:
  - Obvious discontinuity at the threshold?
  - Are there other, unexplained discontinuities?
- Very difficult to sell an RDD without visually obvious result.
  - Imbens & Lemieux: Statistical analysis are just fancy versions of this plot

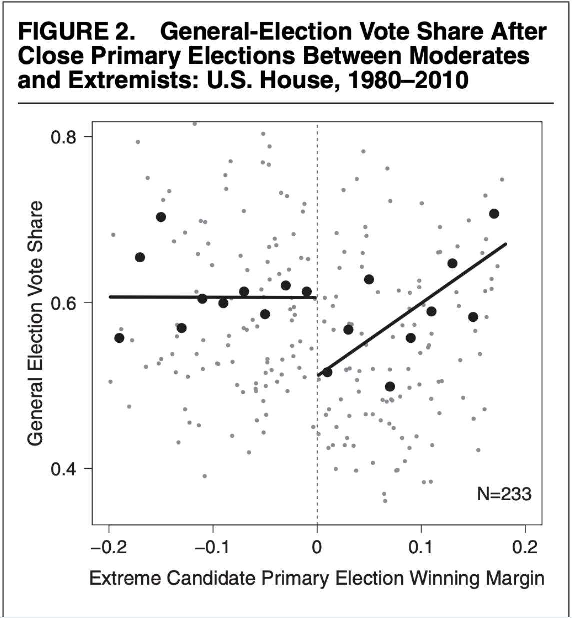
Bin plots

- Binned means plot is very standard:

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

  - b_k are the bin cutpoints.
  - n_k is the number of units within bin k.
- What to observe:
  - Obvious discontinuity at the threshold?
  - Are there other, unexplained discontinuities?
- Very difficult to sell an RDD without visually obvious result.
  - Imbens & Lemieux: Statistical analysis are just fancy versions of this plot
  - If it's not in the binned mean plot, unlikely to be a robust/credible effect.

Example from close election RD



Other graphs to include

- Also good to include binned mean plots for pretreatment covariates.

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.

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

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.

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.
    - Might be an indication of discontinuities in the potential outcome means.

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.
  - Might 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.
    - Might be an indication of discontinuities in the potential outcome means.
    - Similar to balance tests in matching
- **McCrary test**: plot density of the forcing variable.

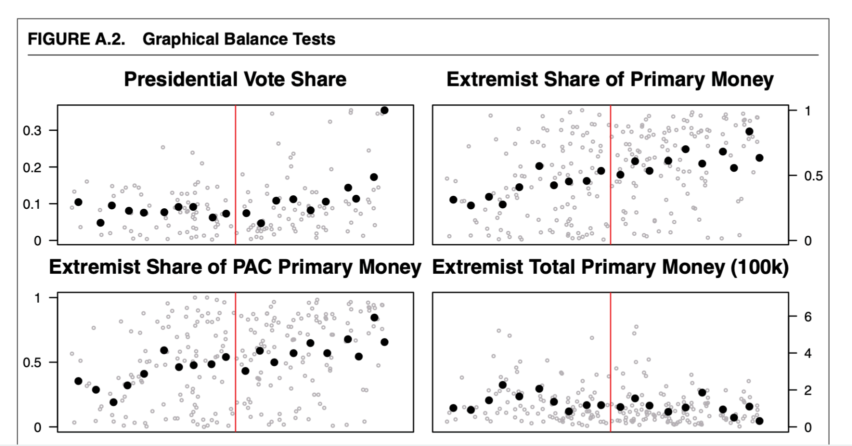
Other graphs to include

- 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.
    - Might 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.

Other graphs to include

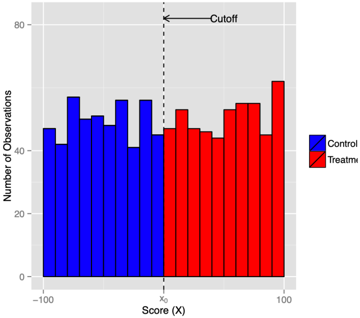
- Also good to include binned mean plots for pretreatment covariates.
- Intuition: key assumption in smoothness in the mean of $Y_i(d)$ in X_i.
- Discontinuities in mean of covariates $\rightsquigarrow$ problematic
    - Covariates unaffected by treatment so might indicate sorting.
    - Might be an 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

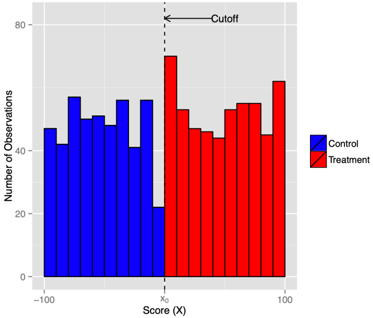


McCrary Test

Figure 10: Histogram of Score



(a) No sorting



(b) Sorting

General estimation strategy

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

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

General estimation strategy

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

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

- Two features different from standard nonparametric regression:

General estimation strategy

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

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

- Two features different from standard nonparametric regression:
  - We want to estimate this regression at a single point.

General estimation strategy

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

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

- 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.

General estimation strategy

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

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

- 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.

General estimation strategy

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

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

- 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!

General estimation strategy

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

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

- 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

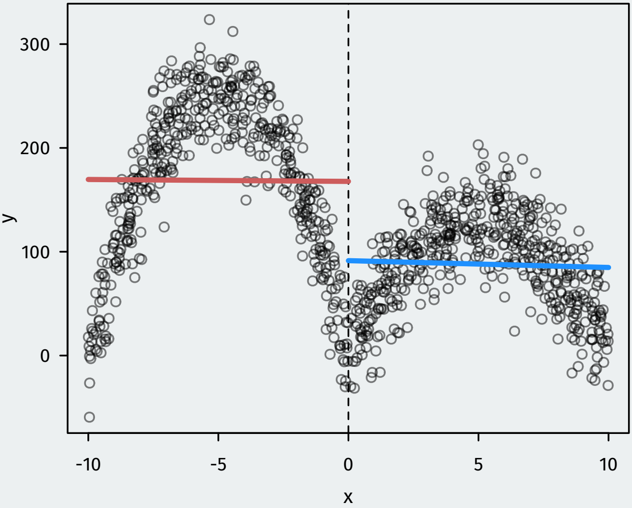
General estimation strategy

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

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

- 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:

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

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

Nonparametric and semiparametric approaches

- Upper and lower limit functions:

$$\mu_+(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_-(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$

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

Nonparametric and semiparametric approaches

- Upper and lower limit functions:

$$\mu_+(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_-(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$

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

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

Nonparametric and semiparametric approaches

- Upper and lower limit functions:

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

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

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

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

  - h is a bandwidth parameter, selected by you.

Nonparametric and semiparametric approaches

- Upper and lower limit functions:

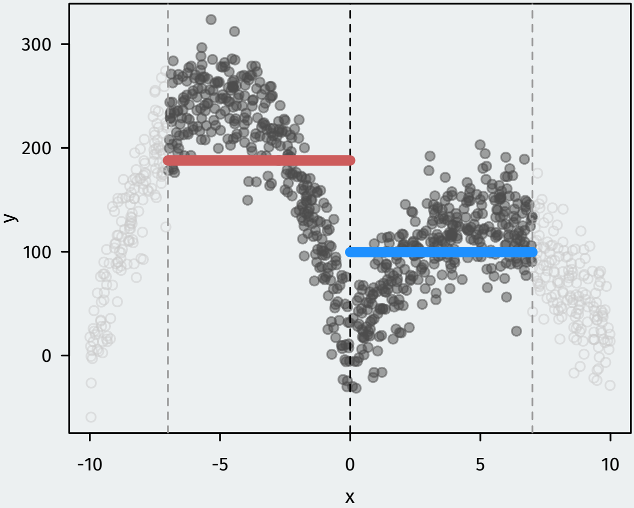
$$\mu_+(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_-(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$

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

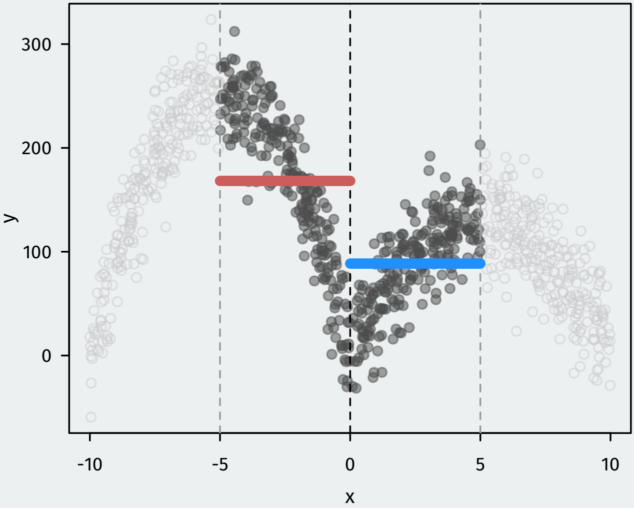
$$\hat{\mu}_-(c) = \frac{\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{c - h \leq X_i < c\}}{\sum_{i=1}^{N} \mathbf{1}\{c - h \leq X_i < c\}}$$

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

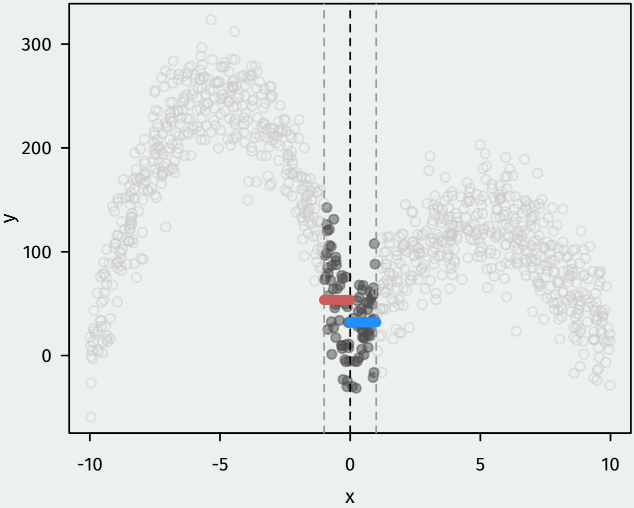
Bandwidth equal to 7



Bandwidth equal to 5



Bandwidth equal to 1



Local averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.

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:

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

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:

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

- 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:

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

- Predictions about Y_i are locally constant on either side of the cutoff.
- h is a **tuning parameter** that controls the **bias-variance tradeoff**:

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:

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

- 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 (more data points, farther from 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 (more data points, farther from the cutoff)
    - Low h: low bias, high variance (fewer data points, 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:

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

- 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 (more data points, farther from the cutoff)
  - Low h: low bias, high variance (fewer data points, closer to the cutoff)
- Downside with averages: bias shrinks slowly as h shrinks.

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:

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

- 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 (more data points, farther from the cutoff)
    - Low h: low bias, high variance (fewer data points, 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 $X_i \in [c - h, c)$:

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

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 $X_i \in [c-h, c)$:

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

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

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

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 $X_i \in [c-h, c)$:

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

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

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

- Our estimate is

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

More practical estimation

- Simplest to use one regression:

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

More practical estimation

- Simplest to use one regression:

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

  - $\widehat{\tau}_{SRD} = \widehat{\tau}$ is the coefficient on the treatment.

More practical estimation

- Simplest to use one regression:

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

  - $\widehat{\tau}_{SRD} = \widehat{\tau}$ is the coefficient on the treatment.
  - Key: interaction between treatment and forcing variable.

More practical estimation

- Simplest to use one regression:

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

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

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

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

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

More practical estimation

- Simplest to use one regression:

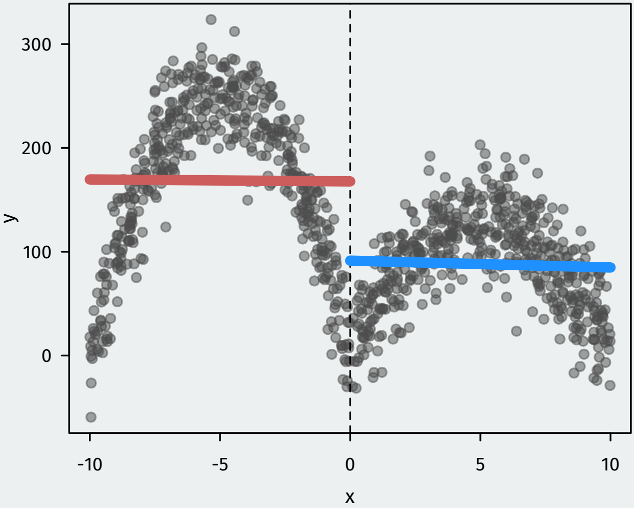
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{i:X_i \in [c-h,c+h]} \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

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

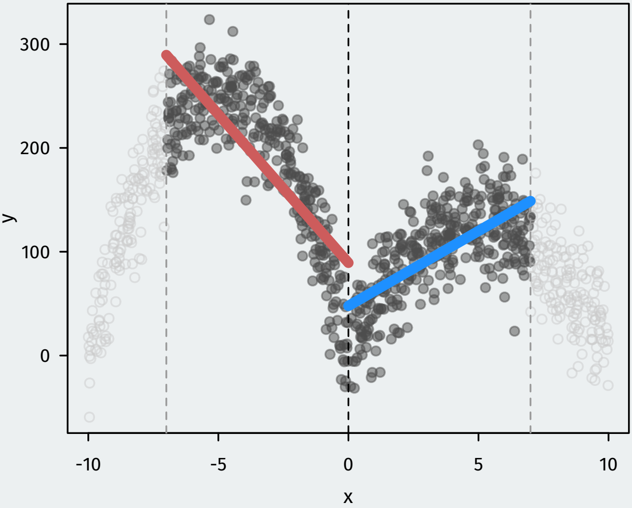
$$\underset{(\alpha,\beta,\tau,\gamma)}{\arg\min} \sum_{1}^{n} K\left(\frac{X_i - c}{h}\right) \{Y_i - \alpha - \beta(X_i - c) - \tau D_i - \gamma(X_i - c)D_i\}^2$$

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

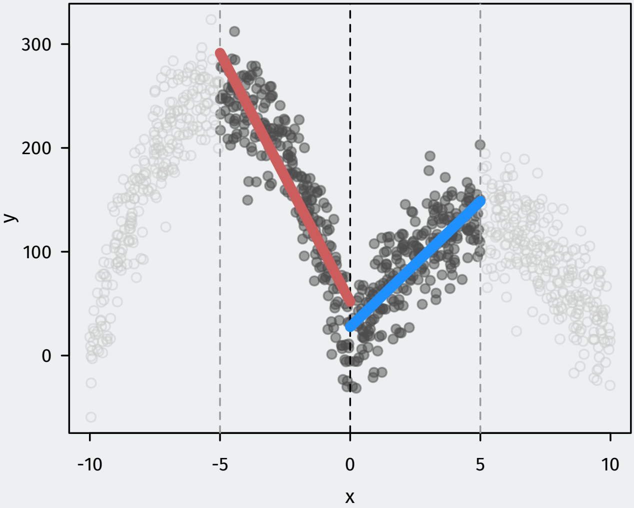
Bandwidth equal to 10 (Global)



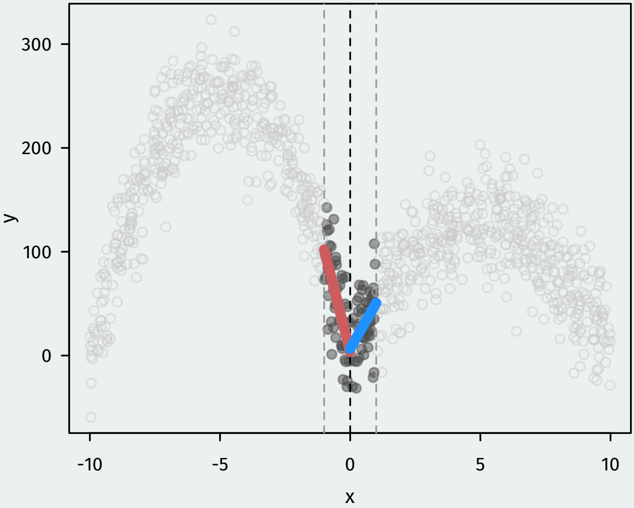
Bandwidth equal to 7



Bandwidth equal to 5



Bandwidth equal to 1



3/ Bandwidth selection

Bandwidths and bias

- Optimal bandwidth shrinks fast enough so $h_n \propto n^{-1/5}$.

Bandwidths and bias

- Optimal bandwidth shrinks fast enough so $h_n \propto n^{-1/5}$.
    - But this results in asymptotic bias, two possible solutions.

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. $h_n \propto n^{-1/4}$

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.

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.

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**: $\widehat{\tau}_{\text{SRD}}^{\text{rbc}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$

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**: $\widehat{\tau}^{\text{rbc}}_{\text{SRD}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$
  - Calonico, Cattaneo, and Titiunik (CCT, 2014, Econometrica) gives the form.

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**: $\widehat{\tau}^{\text{rbc}}_{\text{SRD}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$
  - Calonico, Cattaneo, and Titiunik (CCT, 2014, Econometrica) gives the form.
  - Allows the use of optimal bandwidths, but need to account for estimation of bias.

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.

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**: $\widehat{\tau}^{\text{rbc}}_{\text{SRD}} = \widehat{\tau}_{\text{SRD}} - \widehat{\text{bias}}$
    - 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 $\widehat{\tau}_{\mathsf{SRD}}(h)$

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.

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.

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

Selecting the optimal bandwidth

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

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

- Optimal bandwidth: $h_{\mathsf{MSE}} = \left(\frac{\mathcal{V}}{4\mathcal{B}^2}\right)^{1/5} n^{-1/5}$

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.

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

- 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.

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.

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

  - 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:

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.

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

  - 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.

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.

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

  - 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.

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.

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

    - 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)

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
    - Not great in finite samples because the bandwidth isn't designed for this purpose.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
    - Not great in finite samples because the bandwidth isn't designed for this purpose.
    - CCT derives nearest neighbors variance estimator that has better coverage.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
  - Not great in finite samples because the bandwidth isn't designed for this purpose.
  - CCT derives nearest neighbors variance estimator that has better coverage.
  - If using RBC, you need to account for that in variance.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
  - Not great in finite samples because the bandwidth isn't designed for this purpose.
  - CCT derives nearest neighbors variance estimator that has better coverage.
  - If using RBC, you need to account for that in variance.
- **Covariates**: can add them to the local linear model, but be wary.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
  - Not great in finite samples because the bandwidth isn't designed for this purpose.
  - CCT derives nearest neighbors variance estimator that has better coverage.
  - If using RBC, you need to account for that in variance.
- **Covariates**: can add them to the local linear model, but be wary.
  - If covariates are continuous at the cutoff, won't affect estimates much.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
  - Not great in finite samples because the bandwidth isn't designed for this purpose.
  - CCT derives nearest neighbors variance estimator that has better coverage.
  - If using RBC, you need to account for that in variance.
- **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.

Odds and ends for the SRD

- **Standard errors**: robust standard errors from local OLS are valid.
  - Not great in finite samples because the bandwidth isn't designed for this purpose.
  - CCT derives nearest neighbors variance estimator that has better coverage.
  - If using RBC, you need to account for that in variance.
- **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

- **Standard errors**: robust standard errors from local OLS are valid.
    - Not great in finite samples because the bandwidth isn't designed for this purpose.
    - CCT derives nearest neighbors variance estimator that has better coverage.
    - If using RBC, you need to account for that in variance.
- **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

- **Standard errors**: robust standard errors from local OLS are valid.
    - Not great in finite samples because the bandwidth isn't designed for this purpose.
    - CCT derives nearest neighbors variance estimator that has better coverage.
    - If using RBC, you need to account for that in variance.
- **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.

4/ Fuzzy Regression Discontinuity Designs

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

  - No longer deterministic function of forcing variable.

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

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

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

  - No longer deterministic function of forcing variable.
  - SRD is a special case of the FRD.
- Common use case: threshold allows participation in program.

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

  - 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)

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

  - 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**:

Setup

- **Fuzzy RD**: discontinuity in the probability of treatment.

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

  - 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

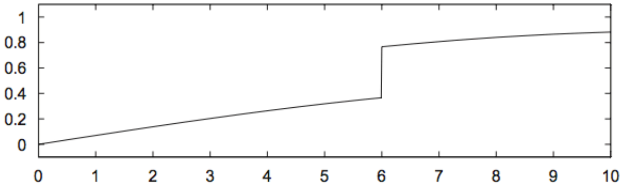


Fig. 3. Assignment probabilities (FRD).

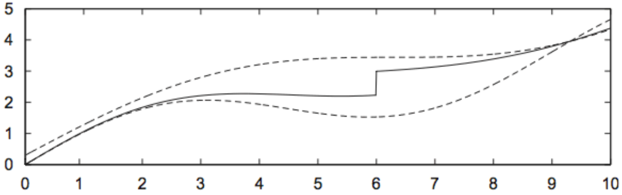


Fig. 4. Potential and observed outcome regression (FRD).

Fuzzy RD assumptions

- $D_i(x)$ potential value of treatment as cutoff changes around c.

Fuzzy RD assumptions

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

Fuzzy RD assumptions

- $D_i(x)$ 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

- $D_i(x)$ 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.

Fuzzy RD assumptions

- $D_i(x)$ 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

- $D_i(x)$ 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 $0 < e < \varepsilon$:

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

Fuzzy RD assumptions

- $D_i(x)$ 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.

Fuzzy RD assumptions

- $D_i(x)$ 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 unobserveable.

Fuzzy RD assumptions

- $D_i(x)$ 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 unobserveable.
- Example: college students that get above a certain GPA are encouraged to apply to grad school.

Fuzzy RD assumptions

- $D_i(x)$ 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 unobserveable.
- 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.

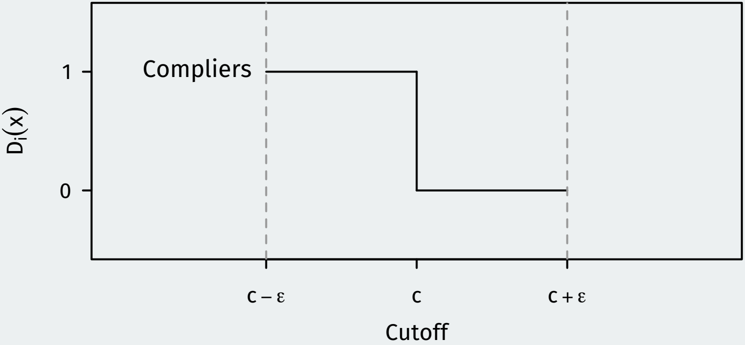
Fuzzy RD assumptions

- $D_i(x)$ 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 $0 < e < \varepsilon$:

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

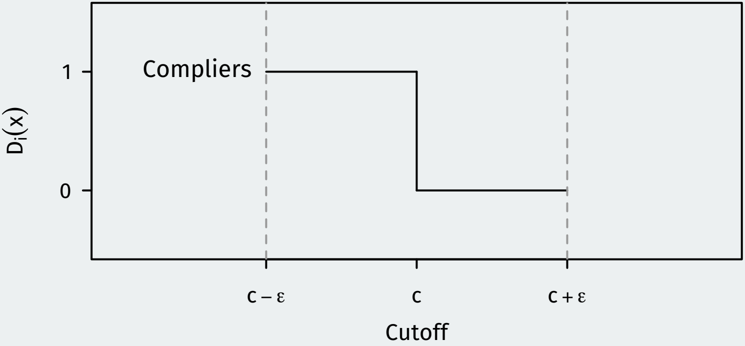
    - Lowering or increasing the threshold would affect their treatment status.
    - Compliance status unobserveable.
- Example: college students that get above a certain GPA are encouraged to apply to grad school.
    - Compliers wouldn't apply if threshold were slightly higher.
    - Compliers would apply 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

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

Compliance groups

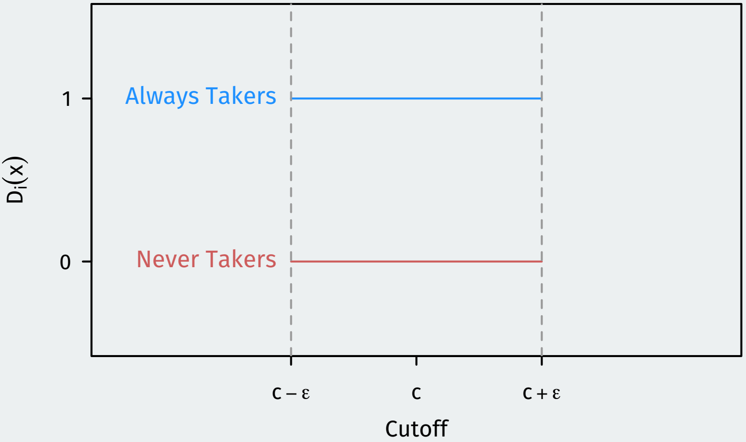
- Compliers: $D_i(c - e) = 1 \quad \text{and} \quad D_i(c + e) = 0$
- Always-takers: $D_i(c + e) = D_i(c - e) = 1$

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: $D_i(c + e) = D_i(c - e) = 0$

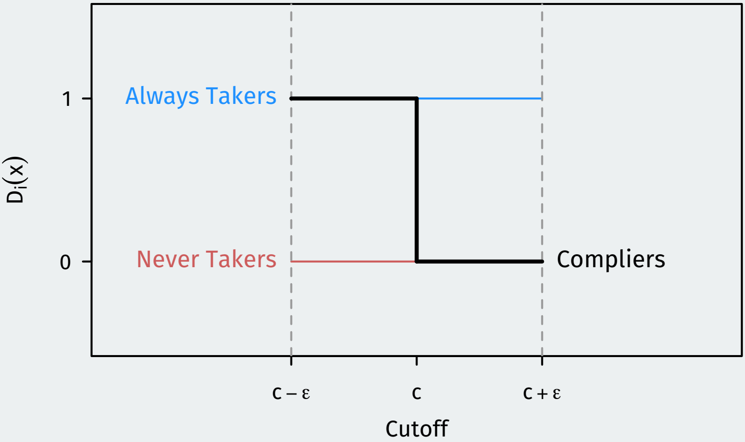
Compliance groups

- 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: $D_i(c+e)=D_i(c-e)=0$



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: $D_i(c + e) = D_i(c - e) = 0$



LATE in the Fuzzy RD

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

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

LATE in the Fuzzy RD

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

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

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

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

LATE in the Fuzzy RD

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

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

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

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

- 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:

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

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

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

- Proof is very similar to the LATE proof
- External validity? Doubly local $\rightsquigarrow$ careful about generalizing.

Estimation in FRD

- Remember that we had:

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

Estimation in FRD

- Remember that we had:

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

- 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:

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

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

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

- CCT provides (more complicated) robust bias correction, bandwidths.

More practical FRD estimation

- The ratio estimator above is equivalent to a TSLS approach.

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:

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

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.

Kink RD

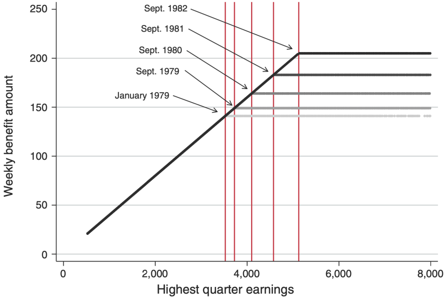


FIGURE 2. LOUISIANA: SCHEDULE OF UI WEEKLY BENEFIT AMOUNT, JAN. 1979–DEC. 1983

- **Sharp Kink RD**: discontinuities in the first derivatives rather than levels.

Kink RD

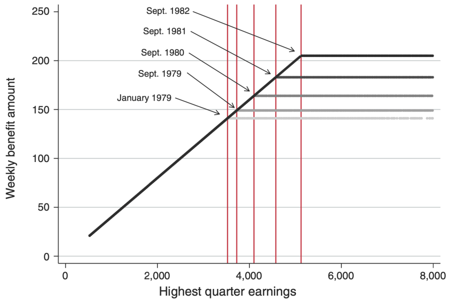


Figure 2. Louisiana: Schedule of UI Weekly Benefit Amount, Jan. 1979–Dec. 1983

- **Sharp Kink RD**: discontinuities in the first derivatives rather than levels.
    - Unemployment benefits as a function of prior earnings.

Kink RD

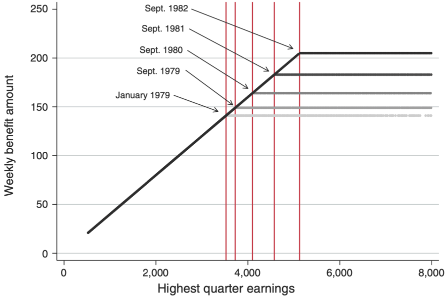


Figure 2. Louisiana: Schedule of UI Weekly Benefit Amount, Jan. 1979–Dec. 1983

- **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.

Kink RD

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

- **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.

Kink RD

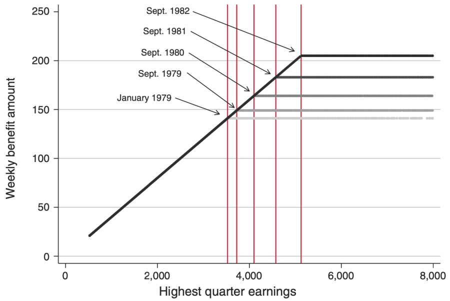


Figure 2. Louisiana: Schedule of UI Weekly Benefit Amount, Jan. 1979–Dec. 1983

- **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.