Lecture 14:
Regression discontinuity I

PPHA 34600
Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: wrapping up DD(D)

We met the DDD estimator:

$$\begin{aligned} Y_{ijt} = \beta_0 &+ \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Affected_j + \beta_4 (Treat_i \times Post_t) \\ &+ \beta_5 (Post_t \times Affected_j) + \beta_6 (Treat_i \times Affected_j) \\ &+ \tau (Treat_i \times Post_t \times Affected_j) + \varepsilon_{ijt} \end{aligned}$$

A tour of research designs

We've met several research designs this quarter:

1. **Randomized controlled trial**
   - Eliminates selection bias via randomization
2. **Regression adjustment & matching**
   - Selection on observables
   - Strong assumption: We observe everything that might matter
3. **Instrumental variables**
   - Use some (quasi)-random variation to move endogenous treatment
   - Strong assumption: Exclusion restriction
4. **Panel fixed effects**
   - Compare units to themselves over time
   - Strong assumption: Parallel counterfactual trends

My favorite selection on unobservables design

Today, we'll meet the regression discontinuity (RD):

- **Basic intuition:** Use a (policy-induced) cutoff to compare i and j

My favorite selection on unobservables design

Today, we'll meet the regression discontinuity (RD):

- **Basic intuition:** Use a (policy-induced) cutoff to compare i and j

→ Look at "barely-treated" units vs. "barely-untreated" units

My favorite selection on unobservables design

Today, we'll meet the regression discontinuity (RD):

- **Basic intuition:** Use a (policy-induced) cutoff to compare i and j

$\rightarrow$ Look at "barely-treated" units vs. "barely-untreated" units

- Enables us to come close to mimicking random assignment
- Identifying assumptions are transparent
- We can do a lot of this in pictures

Regression discontinuity design

As usual, we'd like to run:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

→ We will try to mimic random assignment

Regression discontinuity design

As usual, we'd like to run:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

→ We will try to mimic random assignment

The regression discontinuity:

- Suppose D_i is determined by whether or not X_i lies above a cutoff, c
- We call X_i the "running variable" here
- **Idea**: Having X_i just above or just below c is as good as random...
- ... And there is a discontinuous change in D_i as a result of crossing c

→ We can compare Y_i for units with X_i just above c to Y_i for units with X_i just below c

Sharp regression discontinuity

In the most straightforward, or "sharp" RD design:

- $Pr(D_i = 1 | X_i \geq c) = 1$ and $Pr(D_i = 1 | X_i < c) = 0$
- $Pr(D_i = 1 | X_i \geq c) - Pr(D_i = 1 | X_i < c) = 1$
- Nobody with $X_i < c$ gets treated
- Everybody with $X_i \geq c$ gets treated
- The probability of treatment jumps from 0 to 100% as X_i crosses c
- $D_i = \mathbf{1}(X_i \geq c)$

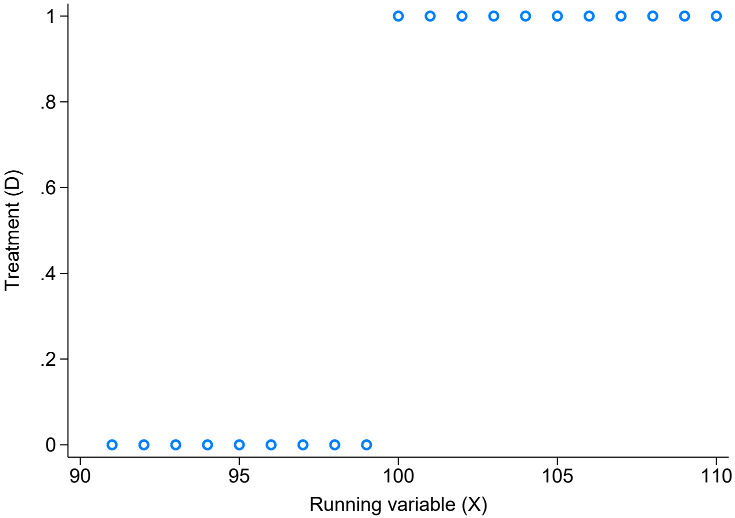
Sharp regression discontinuity

In the most straightforward, or "sharp" RD design:

- $Pr(D_i = 1 | X_i \geq c) = 1$ and $Pr(D_i = 1 | X_i < c) = 0$
- $Pr(D_i = 1 | X_i \geq c) - Pr(D_i = 1 | X_i < c) = 1$
- Nobody with $X_i < c$ gets treated
- Everybody with $X_i \geq c$ gets treated
- The probability of treatment jumps from 0 to 100% as X_i crosses c
- $D_i = \mathbf{1}(X_i \geq c)$

→ This is equivalent to **perfect compliance** in the RCT

Sharp regression discontinuity: Treatment assignment



Sharp regression discontinuity: Estimation

To get τ, compare units with $D_i = 0$ and $D_i = 1$ exactly at the cutoff:

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

- The estimator is defined exactly at the cutoff
- But we will never observe $Y_i(D_i = 1, X_i < c)$ or $Y_i(D_i = 0, X_i \geq c)$

$\rightarrow$ Even in RD land, we can't escape the FPCI!

Sharp regression discontinuity: Estimation

To get τ, compare units with $D_i = 0$ and $D_i = 1$ exactly at the cutoff:

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

- The estimator is defined exactly at the cutoff
- But we will never observe $Y_i(D_i = 1, X_i < c)$ or $Y_i(D_i = 0, X_i \geq c)$

$\rightarrow$ Even in RD land, we can't escape the FPCI!

To overcome this, get super close to c but not exactly there:

$$\underbrace{\lim_{x \downarrow c} E[Y_i | X_i = x]}_{\text{approach } c \text{ from above}} - \underbrace{\lim_{x \uparrow c} E[Y_i | X_i = x]}_{\text{approach } c \text{ from below}}$$

Sharp regression discontinuity: Estimation

To get τ, compare units with $D_i = 0$ and $D_i = 1$ exactly at the cutoff:

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

- The estimator is defined exactly at the cutoff
- But we will never observe $Y_i(D_i = 1, X_i < c)$ or $Y_i(D_i = 0, X_i \geq c)$

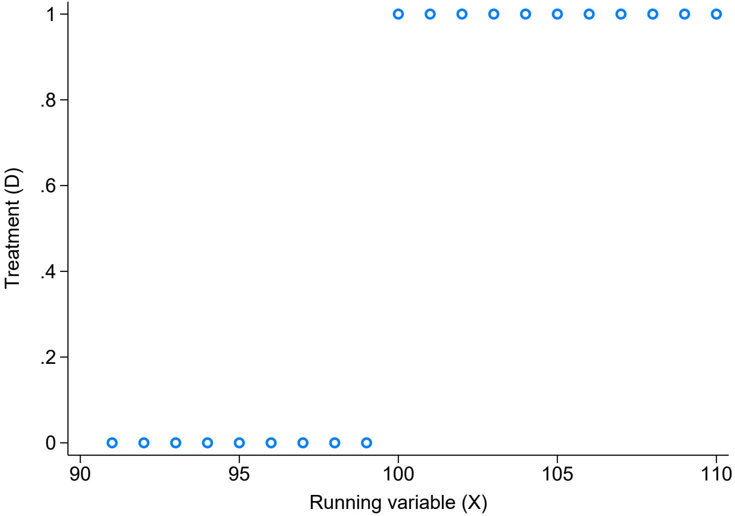
$\rightarrow$ Even in RD land, we can't escape the FPCI! ☠

To overcome this, get super close to c but not exactly there:

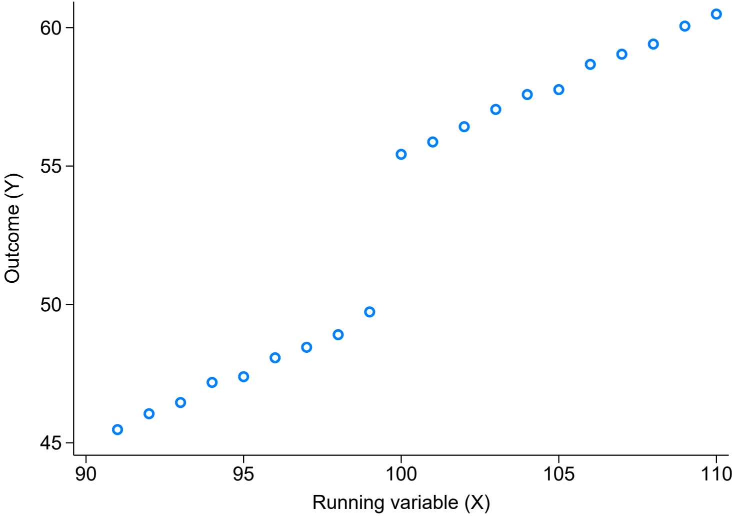
$$\underbrace{\lim_{x \downarrow c} E[Y_i | X_i = x]}_{\text{approach } c \text{ from above}} - \underbrace{\lim_{x \uparrow c} E[Y_i | X_i = x]}_{\text{approach } c \text{ from below}}$$

$$= \underbrace{\lim_{x \downarrow c} E[Y_i(1) | X_i = x]}_{X_i \geq c \text{ get treated}} - \underbrace{\lim_{x \uparrow c} E[Y_i(0) | X_i = x]}_{X_i \geq c \text{ don't get treated}}$$

Sharp regression discontinuity: Treatment assignment



Sharp regression discontinuity: Outcomes



Sharp regression discontinuity: Identifying assumption

We only need one identifying assumption for the sharp RD:

$$E[Y_i(1)|X_i = x] \text{ and } E[Y_i(0)|X_i = x] \text{ are continuous in } x$$

Sharp regression discontinuity: Identifying assumption

We only need one identifying assumption for the sharp RD:

$$E[Y_i(1)|X_i = x] \text{ and } E[Y_i(0)|X_i = x] \text{ are continuous in } x$$

With this assumption:

$$\hat{\tau}^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

Sharp regression discontinuity: Identifying assumption

We only need one identifying assumption for the sharp RD:

$$E[Y_i(1)|X_i = x] \text{ and } E[Y_i(0)|X_i = x] \text{ are continuous in } x$$

With this assumption:

$$\hat{\tau}^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

In words:

- We can compare units with X_i very close to, but not exactly at, c

Sharp regression discontinuity: Identifying assumption

We only need one identifying assumption for the sharp RD:

$$E[Y_i(1)|X_i = x] \text{ and } E[Y_i(0)|X_i = x] \text{ are continuous in } x$$

With this assumption:

$$\hat{\tau}^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

In words:

- We can compare units with X_i very close to, but not exactly at, c

In other words:

- The cutoff is as good as randomly assigned

In more other words:

- There are no discrete jumps in Y_i at c except due to D_i

In even more other words:

- All observed and unobserved determinants of Y_i (other than treatment) are smooth around the cutoff

RD validity tests

We can perform two major RD validity checks:

1. A "bunching" or "manipulation" test
2. A "covariate smoothness" test

→ As usual, we can't prove the identifying assumption!

→ We can just provide evidence in favor of it!

Manipulation tests

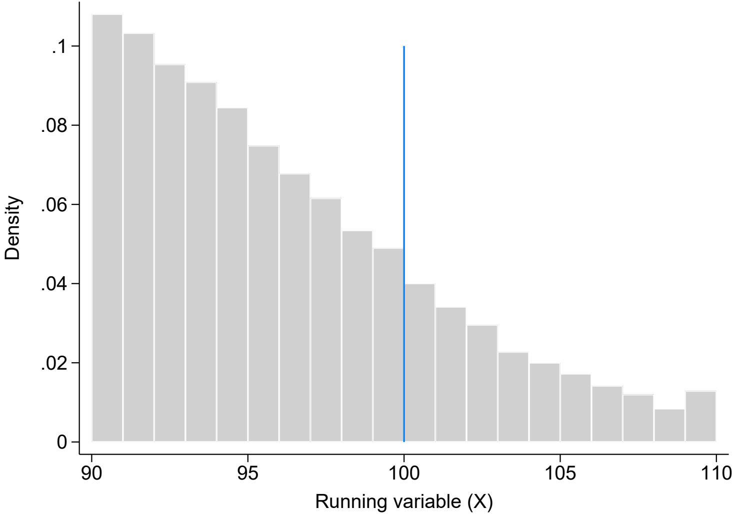
We are assuming that $X_i - c$ is as good as randomly assigned:

- (In the neighborhood of c)

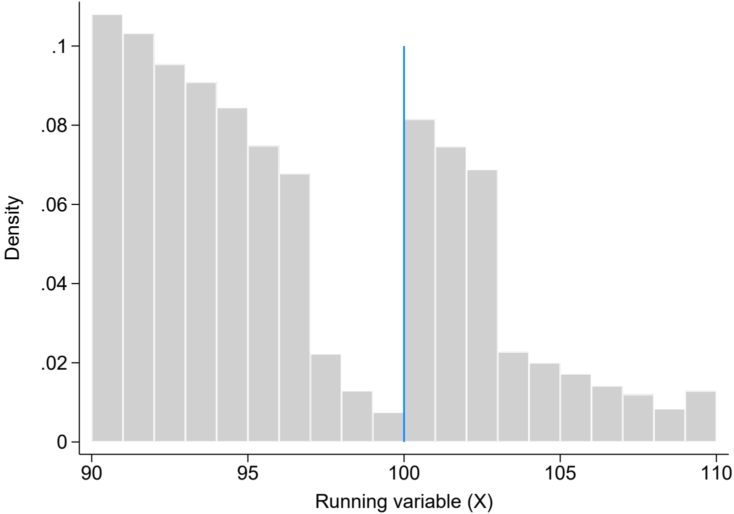
$\rightarrow$ We want to make sure units can't sort around c

$\rightarrow$ We test this by looking at the distribution of X_i

A manipulation test



A manipulation test



Covariate smoothness

We are assuming no other variables change discontinuously at c:

- We test this by looking at other variables around c
- This works for predetermeined variables only!

Covariate smoothness

We are assuming no other variables change discontinuously at c:

- We test this by looking at other variables around c
- This works for predetermeined variables only!
  - $\rightarrow$ Outcomes affected by treatment may well jump at c

Covariate smoothness

We are assuming no other variables change discontinuously at c:

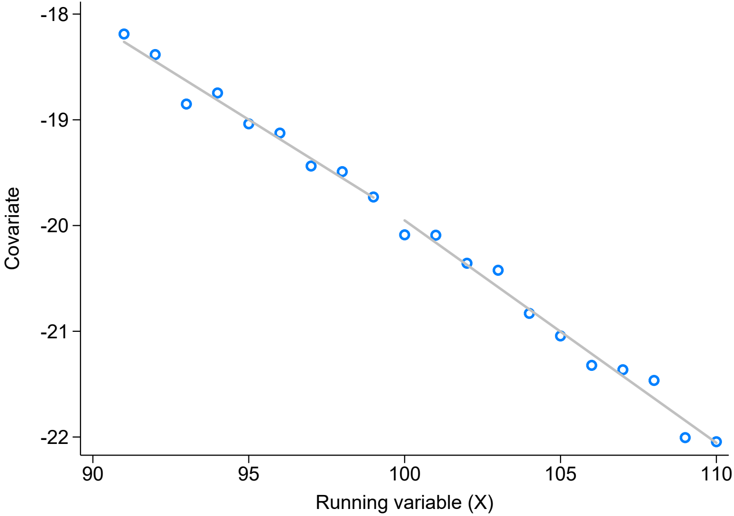
- We test this by looking at other variables around c
- This works for predetermeined variables only!
  - $\rightarrow$ Outcomes affected by treatment may well jump at c
- This test is imperfect (why?)

Covariate smoothness

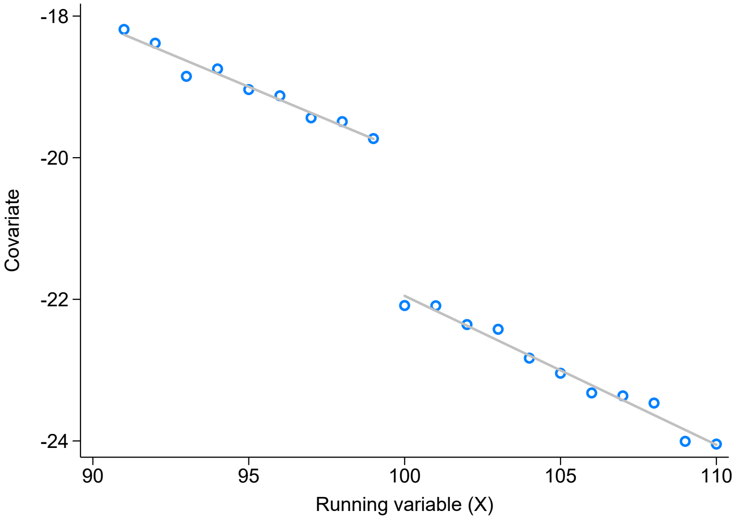
We are assuming no other variables change discontinuously at c:

- We test this by looking at other variables around c
- This works for predetermeined variables only!
  - $\rightarrow$ Outcomes affected by treatment may well jump at c
- This test is imperfect (why?)
  - $\rightarrow$ We can't check for smoothness of unobservables

A covariate smoothness test



A covariate smoothness test



Putting the "regression" in regression discontinuity

We want the difference in outcomes for just-treated vs. just-untreated:

$$\tau^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

Putting the "regression" in regression discontinuity

We want the difference in outcomes for just-treated vs. just-untreated:

$$\tau^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

We estimate average outcomes just below and above the cutoff:

$$\hat{\tau}^{SRD} = \bar{Y}(D_i = 1; c \leq X_i \leq c + h) - \bar{Y}(D_i = 0; c - h \leq X_i < c)$$

where $c - h \leq X_i \leq c + h$ is the bandwith in which we're "close" to c

Putting the “regression” in regression discontinuity

We want the difference in outcomes for just-treated vs. just-untreated:

$$\tau^{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]$$

We estimate average outcomes just below and above the cutoff:

$$\hat{\tau}^{SRD} = \bar{Y}(D_i = 1; c \leq X_i \leq c + h) - \bar{Y}(D_i = 0; c - h \leq X_i < c)$$

where $c - h \leq X_i \leq c + h$ is the bandwith in which we’re “close” to c

This leads to the regression-based RD:

$$Y_i = \alpha + \tau D_i + \varepsilon_i \text{ for } c - h \leq X_i \leq c + h$$

where $D_i = \mathbf{1}[X_i \geq c]$

How do we choose h?

$$Y_i = \alpha + \tau D_i + \varepsilon_i \text{ for } c - h \leq X_i \leq c + h$$

where $c - h \leq X_i \leq c + h$ is the bandwith in which we're "close" to c

- We want h to be small: the RD is identified only at c
    - If too small, we will get imprecision (no sample density)
- We want h to be big enough: our standard errors will be huge just using data at c
    - If too big, we will get bias (comparing dissimilar units)

How do we choose h?

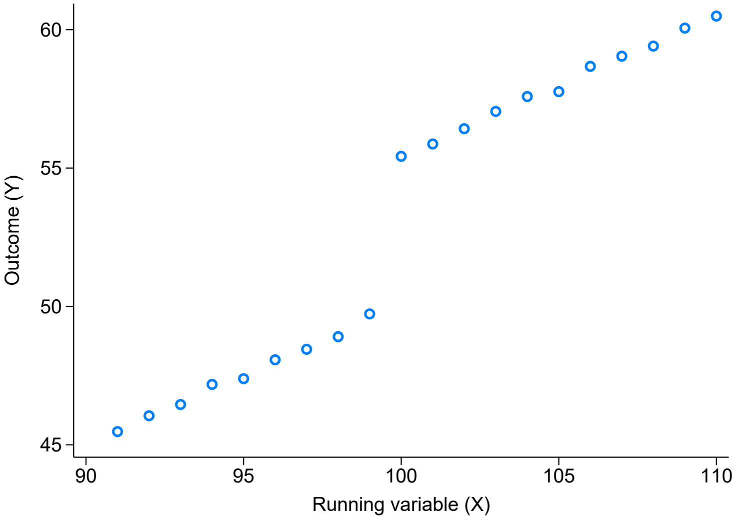
$$Y_i = \alpha + \tau D_i + \varepsilon_i \text{ for } c - h \leq X_i \leq c + h$$

where $c - h \leq X_i \leq c + h$ is the bandwith in which we're "close" to c

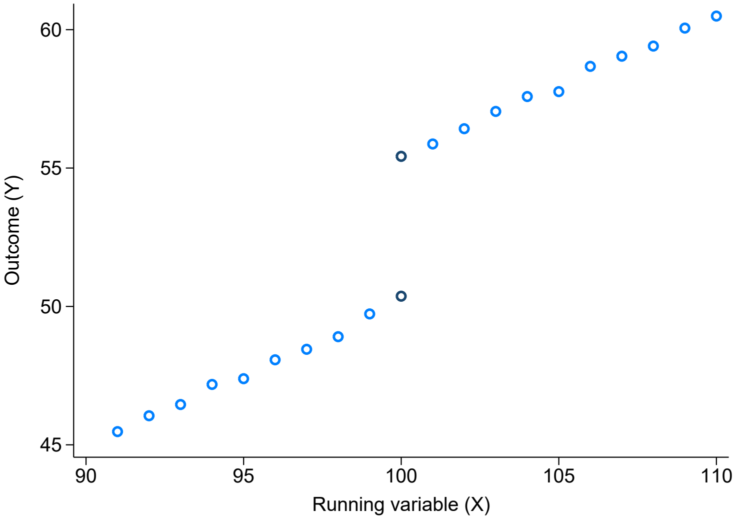
- We want h to be small: the RD is identified only at c
  - If too small, we will get imprecision (no sample density)
- We want h to be big enough: our standard errors will be huge just using data at c
  - If too big, we will get bias (comparing dissimilar units)

→ This is another example of the **bias-variance tradeoff**

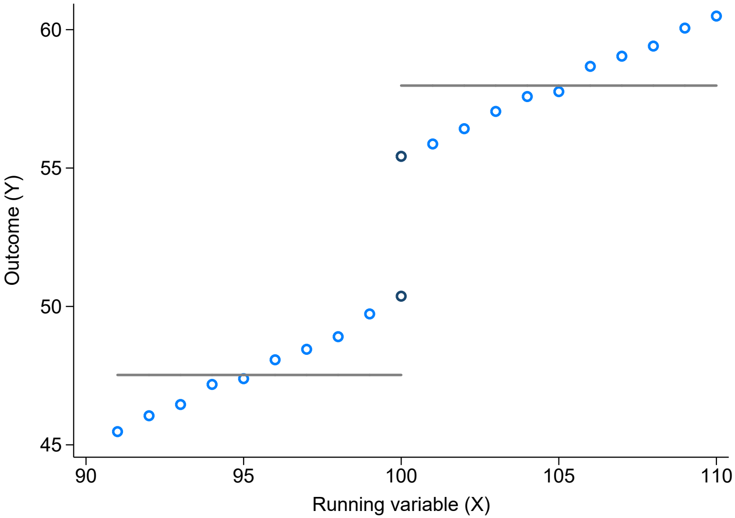
Bandwidth-induced bias in the RD



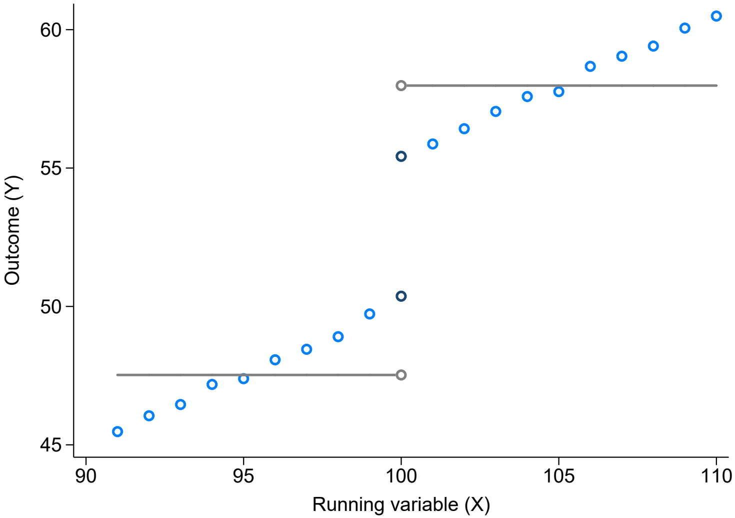
Bandwidth-induced bias in the RD



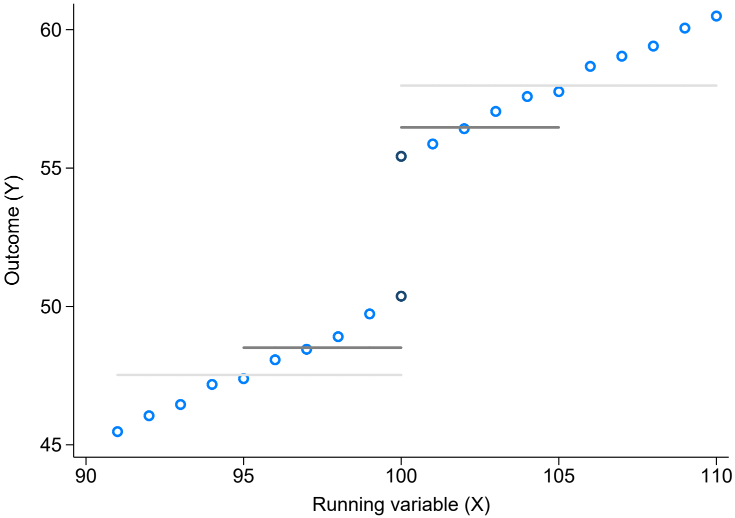
Bandwidth-induced bias in the RD



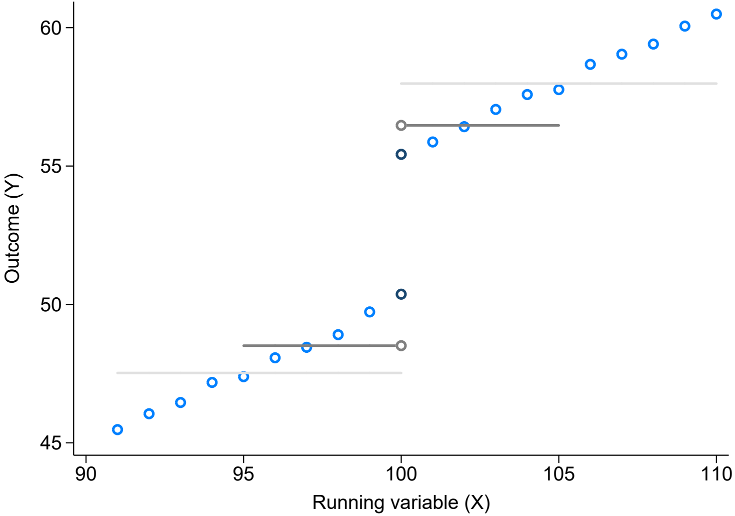
Bandwidth-induced bias in the RD



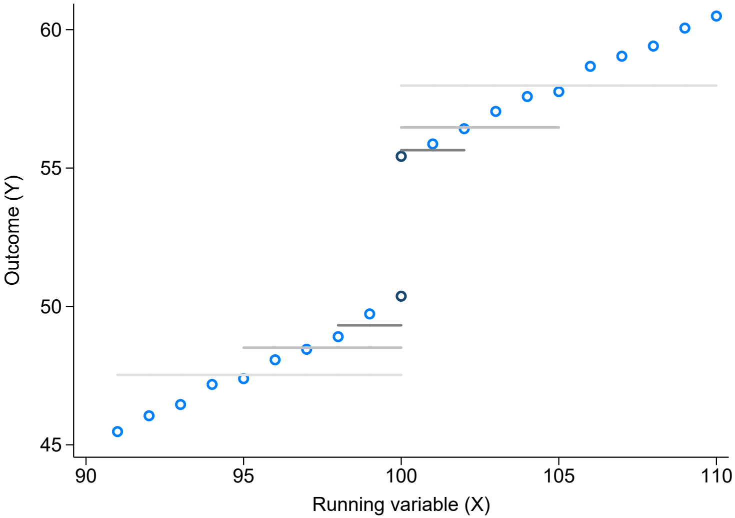
Bandwidth-induced bias in the RD



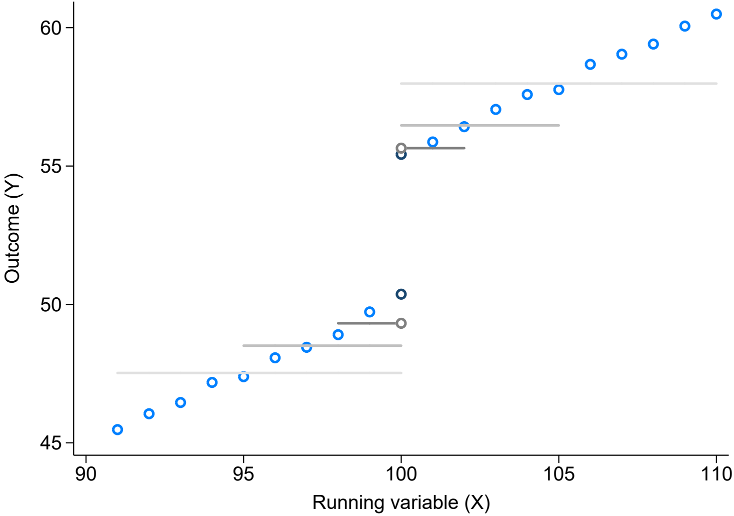
Bandwidth-induced bias in the RD



Bandwidth-induced bias in the RD



Bandwidth-induced bias in the RD



Can we do better than differences in means?

Simple differences in means ignores any relationship between Y_i and X_i:

- We can improve on this by controlling for the underlying relationship
- If we know $Y_i(X_i)$ is a linear function, we can just regress:

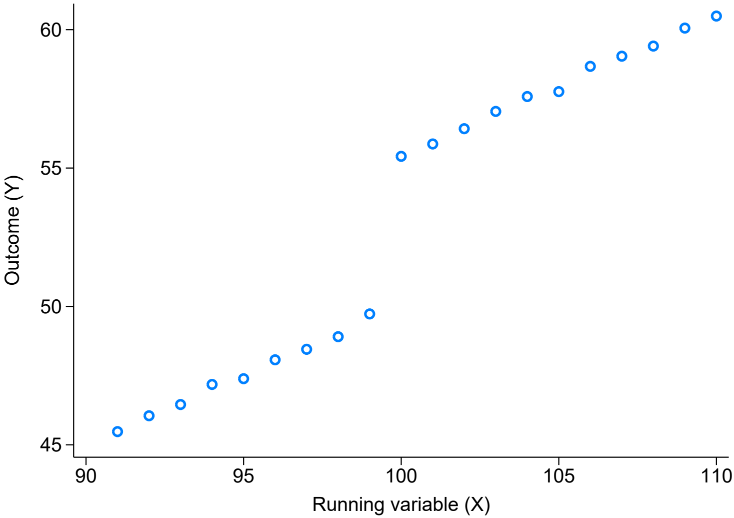
$$Y_i = \alpha + \tau D_i + \beta(X_i - c) + \varepsilon_i$$

$$= \alpha + \tau \mathbf{1}[X_i \geq c] + \beta(X_i - c) + \varepsilon_i$$

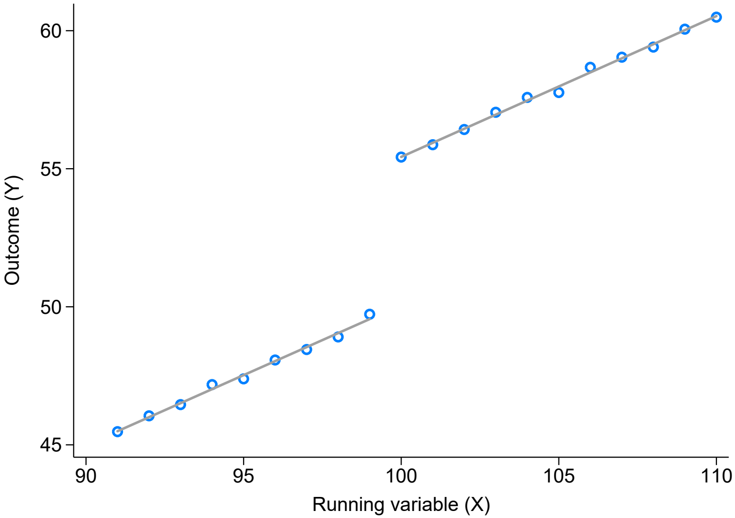
- We can also allow for different slopes above and below c:

$$Y_i = \alpha + \tau D_i + \underbrace{\beta_1(X_i - c)}_{\text{slope below}} + \underbrace{\beta_2(X_i - c)D_i}_{\text{slope above}} + \varepsilon_i$$

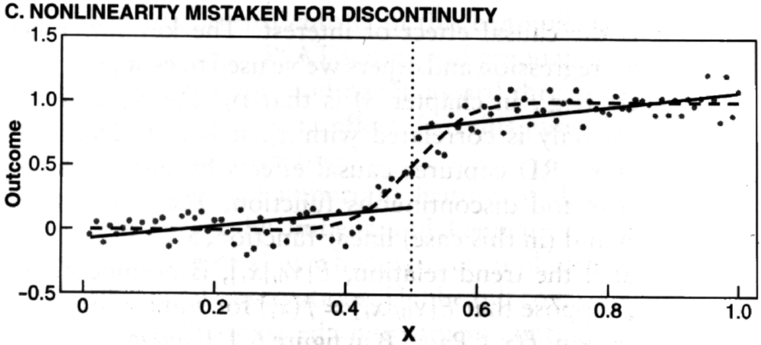
Controlling for the slope



Controlling for the slope



We have to be a bit careful about this!



A note on interpretation

External validity is an important consideration for RD:

- (This is true for all designs...)
- ... but in RD, we are estimating results *at the cutoff, c*
- In sharp RD, we're estimating a LATE around the cutoff!

→ This may be different (or not) from the ATE, or other LATEs from other cutoffs

Recap

TL;DR:

1. The regression discontinuity is great
2. We can mimic an RCT in observational data
3. And the tests are visual and transparent