Gov 2002 - Causal Inference III: Regression Discontinuity Designs

Matthew Blackwell   Arthur Spirling

October 16th, 2014

Introduction

- Causal for us so far: selection of observables, instrumental variables for when this doesn't hold

Introduction

- Causal for us so far: selection of observables, instrumental variables for when this doesn't hold
- Basic idea behind both: find some plausibly exogeneous variation in the treatment assignment

Introduction

- Causal for us so far: selection of observables, instrumental variables for when this doesn't hold
- Basic idea behind both: find some plausibly exogeneous variation in the treatment assignment
- Selection on observables: treatment as-if random conditional on X_i

Introduction

- Causal for us so far: selection of observables, instrumental variables for when this doesn't hold
- Basic idea behind both: find some plausibly exogeneous variation in the treatment assignment
- Selection on observables: treatment as-if random conditional on X_i
- IV: instrument provides exogeneous variation

Introduction

- Causal for us so far: selection of observables, instrumental variables for when this doesn't hold
- Basic idea behind both: find some plausibly exogeneous variation in the treatment assignment
- Selection on observables: treatment as-if random conditional on X_i
- IV: instrument provides exogeneous variation
- Regression Discontinuity: exogeneous variation from a discontinuity in treatment assignment

Plan of attack

Sharp Regression Discontinuity Designs

Estimation in the SRD

Readings

Fuzzy Regression Discontinuity Designs

Sharp Regression Discontinuity Designs

Setup

- The basic idea behind regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.

Setup

- The basic idea behind regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.
- This variable may or may not be related to the potential outcomes, but we assume that relationship is smooth, so that changes in the outcome around the threshold can be interpretted as a causal effect.

Setup

- The basic idea behind regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.
- This variable may or may not be related to the potential outcomes, but we assume that relationship is smooth, so that changes in the outcome around the threshold can be interpretted as a causal effect.
- The classic example of this is in the educational context:

Setup

- The basic idea behind regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.
- This variable may or may not be related to the potential outcomes, but we assume that relationship is smooth, so that changes in the outcome around the threshold can be interpretted as a causal effect.
- 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 regression discontinuity designs is that we have a variable, X_i, that we call the **forcing variable**, which determines (partly or wholly) the treatment assignment on either side of a fixed threshold.
- This variable may or may not be related to the potential outcomes, but we assume that relationship is smooth, so that changes in the outcome around the threshold can be interpretted as a causal effect.
- 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)

Notation

- Treatment: $A_i = 1$ or $A_i = 0$

Notation

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

Notation

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

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

Notation

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

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

- Forcing variable: $X_i \in \mathbb{R}$

Notation

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

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

- Forcing variable: $X_i \in \mathbb{R}$
- Covariates: an M-length vector $Z_i = (Z_{1i}, \ldots, Z_{Mi})$

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

- When test scores are above 1500 $\rightarrow$ offered scholarship

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

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

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

- When test scores are above 1500 $\rightarrow$ offered scholarship
- When test scores are below 1500 $\rightarrow$ not offered scholarship
- Key assumption: no compliance problems (deterministic)

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

- When test scores are above 1500 $\rightarrow$ offered scholarship
- When test scores are below 1500 $\rightarrow$ not offered scholarship
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

- When test scores are above 1500 $\rightarrow$ offered scholarship
- When test scores are below 1500 $\rightarrow$ not offered scholarship
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

Design

- In a sharp RD design, the treatment assignment is a deterministic function of the forcing variable and the threshold, c so that:

Assumption SRD

$$A_i = 1\{X_i \geq c\} \qquad \forall i$$

- When test scores are above 1500 $\rightarrow$ offered scholarship
- When test scores are below 1500 $\rightarrow$ not offered scholarship
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(A_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(A_i = 1 | X_i = c - \varepsilon) = 0$$

Threshold

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.

Threshold

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We want to investigate the behavior of the outcome around the threshold:

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

Threshold

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We want to investigate the behavior of the outcome around the threshold:

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

- Under certain assumptions, this quantity identifies the ATE at the threshold:

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

Plotting the RDD (Imbens and Lemieux, 2008)

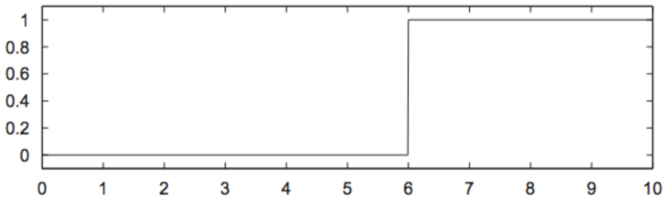


Fig. 1. Assignment probabilities (SRD).

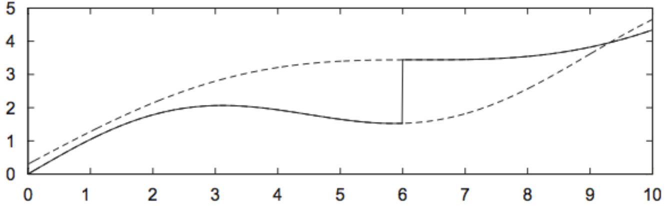


Fig. 2. Potential and observed outcome regression functions.

Comparison to traditional setup

- Note that ignorability here hold by design, because condition on the forcing variable, the treatment is deterministic.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp A_i | X_i$$

Comparison to traditional setup

- Note that ignorability here hold by design, because condition on the forcing variable, the treatment is deterministic.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp A_i | X_i$$

- Again, we can't directly use this because we know that the usual posivity assumption is violated. Remember that positivity is an overlap condition:

$$0 < \Pr[A_i = 1 | X_i = x] < 1$$

Comparison to traditional setup

- Note that ignorability here hold by design, because condition on the forcing variable, the treatment is deterministic.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp A_i | X_i$$

- Again, we can't directly use this because we know that the usual posivity assumption is violated. Remember that positivity is an overlap condition:

$$0 < \Pr[A_i = 1 | X_i = x] < 1$$

- Here, obviously, the propensity score is only 0 or 1, depending on the value of the forcing variable.

Comparison to traditional setup

- Note that ignorability here hold by design, because condition on the forcing variable, the treatment is deterministic.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp A_i | X_i$$

- Again, we can't directly use this because we know that the usual posivity assumption is violated. Remember that positivity is an overlap condition:

$$0 < \Pr[A_i = 1 | X_i = x] < 1$$

- Here, obviously, the propensity score is only 0 or 1, depending on the value of the forcing variable.
- Thus, we need to extrapolate from the treated to the control group and vice versa.

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

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

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

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

- But we don't observe $E[Y_i(0)|X_i = c]$ ever due to the design, so we're going to extrapolate from $E[Y_i(0)|X_i = c - \varepsilon]$.

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

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

- But we don't observe $E[Y_i(0)|X_i = c]$ ever due to the design, so we're going to extrapolate from $E[Y_i(0)|X_i = c - \varepsilon]$.
- Extrapolation, even at short distances, requires a certain smoothness in the functions we are extrapolating.

Continuity of the CEFs

Assumption 1: Continuity

The functions

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

are continuous in x.

Continuity of the CEFs

Assumption 1: Continuity

The functions

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

are continuous in x.

- 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 1: Continuity

The functions

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

are continuous in x.

- 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 1: Continuity

The functions

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

are continuous in x.

- This continuity implies the following:

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

Continuity of the CEFs

Assumption 1: Continuity

The functions

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

are continuous in x.

- This continuity implies the following:

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

Continuity of the CEFs

Assumption 1: Continuity

The functions

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

are continuous in x.

- This continuity implies the following:

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

- 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

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

- Note that each of these is identified at least with infinite data, as long as X_i has positive density around the cutpoint

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

- Note that each of these is identified at least with infinite data, as long as X_i has positive density around the cutpoint
- Why? With arbitrarily high N, we'll get an arbitrarily good approximations to the expectation of the line

Identification results

- Thus, under the ignorability assumption, the sharp RD assumption, and the continuity assumption, we have:

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

- Note that each of these is identified at least with infinite data, as long as X_i has positive density around the cutpoint
- Why? With arbitrarily high N, we'll get an arbitrarily good approximations to the expectation of the line
- How to estimate these nonparametrically is difficult as we'll see (endpoints are a big problem)

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.
- For instance, if people sort around threshold, then you might get jumps other than the one you care about.

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.
- For instance, if people sort around threshold, then you might get jumps other than the one you care about.
- If things other than the treatment change at the threshold, then that might cause discontinuities in the potential outcomes.

Estimation in the SRD

Graphical approaches

- Simple plot of mean outcomes within bins of the forcing variable:
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$
where N_k is the number of units within bin k and b_k are the bin cutpoints.

Graphical approaches

- Simple plot of mean outcomes within bins of the forcing variable:
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$
where N_k is the number of units within bin k and b_k are the bin cutpoints.
- Obvious discontinuity at the threshold?

Graphical approaches

- Simple plot of mean outcomes within bins of the forcing variable:
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$
where N_k is the number of units within bin k and b_k are the bin cutpoints.
- Obvious discontinuity at the threshold?
- Are there other, unexplained discontinuities?

Graphical approaches

- Simple plot of mean outcomes within bins of the forcing variable:
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$
where N_k is the number of units within bin k and b_k are the bin cutpoints.
- Obvious discontinuity at the threshold?
- Are there other, unexplained discontinuities?
- As Imbens and Lemieux say:

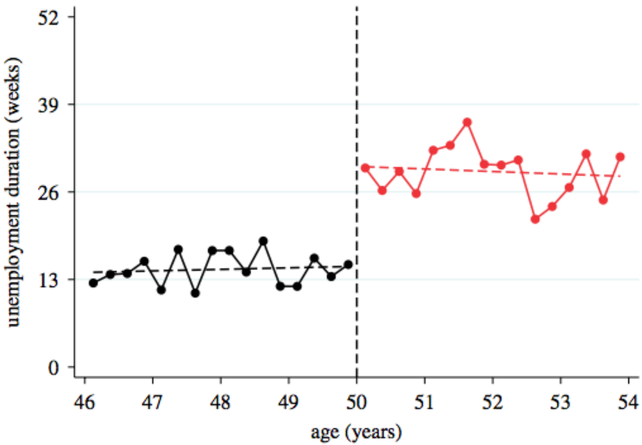
Graphical approaches

- Simple plot of mean outcomes within bins of the forcing variable:
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^{N} Y_i \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$
where N_k is the number of units within bin k and b_k are the bin cutpoints.
- Obvious discontinuity at the threshold?
- Are there other, unexplained discontinuities?
- As Imbens and Lemieux say:

  > *The formal statistical analyses discussed below are essentially just sophisticated versions of this, and if the basic plot does not show any evidence of a discontinuity, there is relatively little chance that the more sophisticated analyses will lead to robust and credible estimates with statistically and substantially significant magnitudes.*

Example from RD on extending unemployment

R. Lalive / Journal of Econometrics 142 (2008) 785–806



Discontinuity at threshold = 14.798; with std. err. = 1.928.

Other graphs to include

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity.

Other graphs to include

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity.
- Same binning strategy:

$$\overline{Z}_{km} = \frac{1}{N_k} \sum_{i=1}^{N} Z_{im} \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$

Other graphs to include

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity.
- Same binning strategy:

$$\overline{Z}_{km} = \frac{1}{N_k} \sum_{i=1}^{N} Z_{im} \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$

- Intuition: our key assumption is that the potential outcomes are smooth in the forcing variable.

Other graphs to include

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity.
- Same binning strategy:

$$\overline{Z}_{km} = \frac{1}{N_k} \sum_{i=1}^{N} Z_{im} \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$

- Intuition: our key assumption is that the potential outcomes are smooth in the forcing variable.
- Discontinuities in covariates unaffected by the threshold could be indications of discontinuities in the potential outcomes.

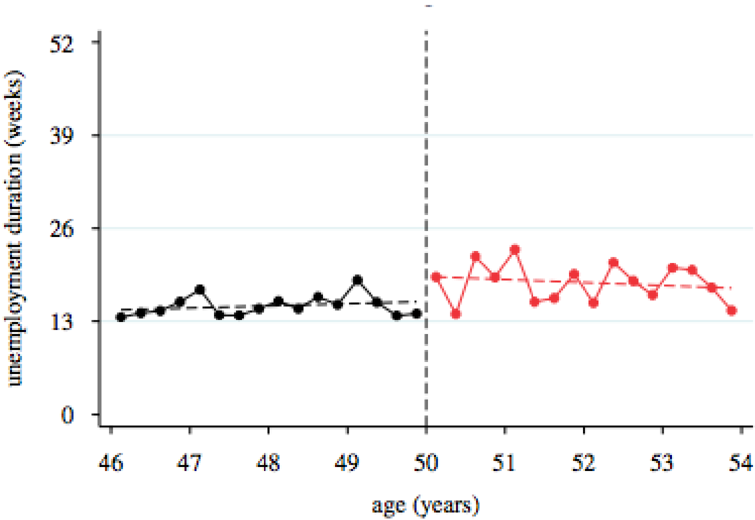
Other graphs to include

- Next, it's a good idea to plot covariates by the forcing variable to see if these covariates also jump at the discontinuity.
- Same binning strategy:

$$\overline{Z}_{km} = \frac{1}{N_k} \sum_{i=1}^{N} Z_{im} \cdot \mathbb{I}(b_k < X_i \leq b_{k+1})$$

- Intuition: our key assumption is that the potential outcomes are smooth in the forcing variable.
- Discontinuities in covariates unaffected by the threshold could be indications of discontinuities in the potential outcomes.
- Similar to balance tests in matching

Checking covariates at the discontinuity



Discontinuity at threshold = 3.442; with std. err. = 1.416.

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:
$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$
- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:
$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$
- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly.

General estimation strategy

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

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

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly.
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.

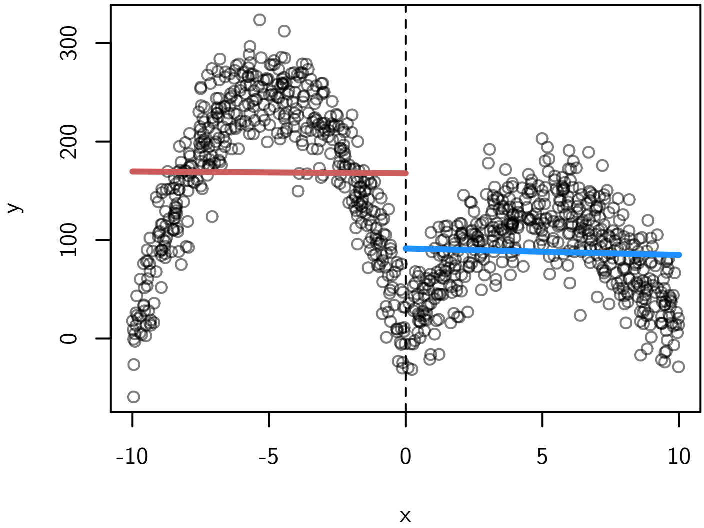
General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:
$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$
- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly.
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.
- Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:
$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$
- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly.
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.
- Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint.
- $\rightarrow$ restrict our estimation to units close to the threshold.

Example of misleading trends



Nonparametric and semiparametric approaches

- Let's define
$$\mu_R(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_L(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$

Nonparametric and semiparametric approaches

- Let's define
$$\mu_R(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_L(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$
- For the SRD, we have $\tau_{SRD} = \mu_1(x) - \mu_0(x)$.

Nonparametric and semiparametric approaches

- Let's define
$$\mu_R(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_L(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$
- For the SRD, we have $\tau_{SRD} = \mu_1(x) - \mu_0(x)$.
- One nonparametric approach is to estimate nonparametrically $\mu_L(x)$ with a **uniform kernel**:

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

Nonparametric and semiparametric approaches

- Let's define

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

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

- For the SRD, we have $\tau_{SRD} = \mu_1(x) - \mu_0(x)$.
- One nonparametric approach is to estimate nonparametrically $\mu_L(x)$ with a **uniform kernel**:

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

- Here, h is a bandwidth parameter, selected by you.

Nonparametric and semiparametric approaches

- Let's define

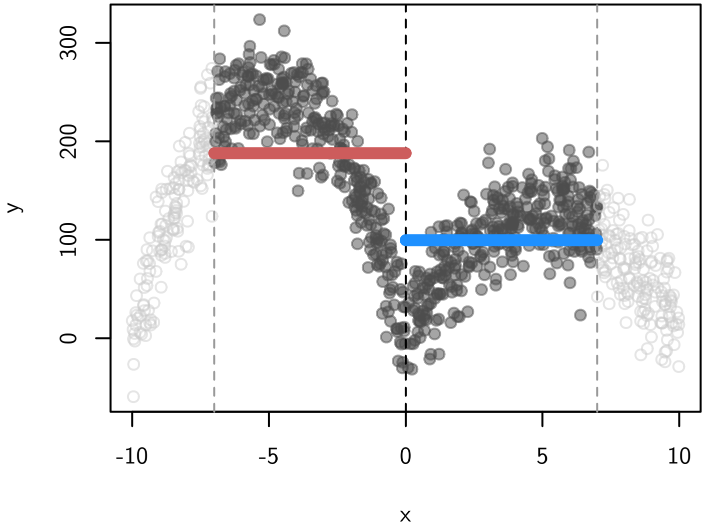
$$\mu_R(x) = \lim_{z \downarrow x} E[Y_i(1)|X_i = z]$$
$$\mu_L(x) = \lim_{z \uparrow x} E[Y_i(0)|X_i = z]$$

- For the SRD, we have $\tau_{SRD} = \mu_1(x) - \mu_0(x)$.
- One nonparametric approach is to estimate nonparametrically $\mu_L(x)$ with a **uniform kernel**:

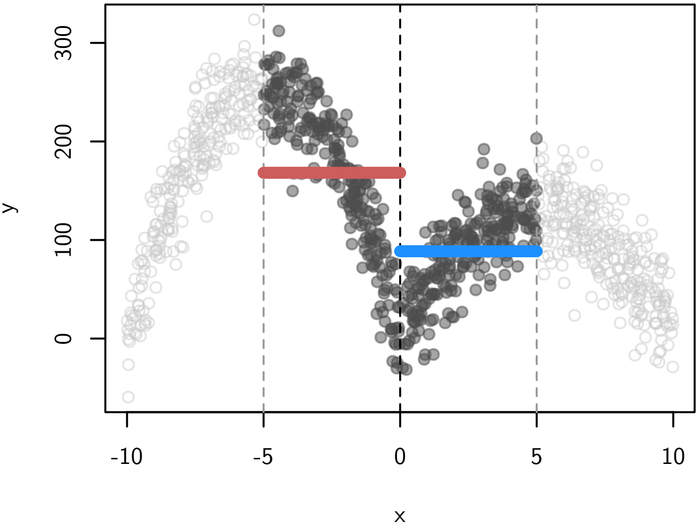
$$\widehat{\mu}_L(c) = \frac{\sum_{i=1}^{N} Y_i \cdot \mathbb{I}\{c - h \leq X_i < c\}}{\sum_{i=1}^{N} \mathbb{I}\{c - h \leq X_i < c\}}$$

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

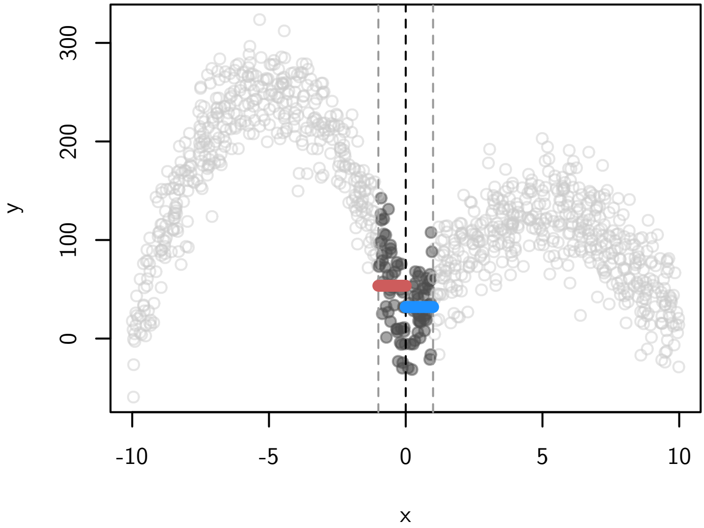
Bandwidth equal to 7



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 do this with the following approach regression on those units less than h away from c:

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

Local averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.
- Can do this with the following approach regression on those units less than h away from c:

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

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

Local averages

- Estimate mean of Y_i when $X_i \in [c, c+h]$ and when $X_i \in [c-h, c)$.
- Can do this with the following approach regression on those units less than h away from c:

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

- Here, $\widehat{\tau}_{SRD} = \widehat{\tau}$.
- This turns out to have very large bias as the we increase the bandwidth.

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 $X_i \in [c, c+h]$ to estimate $\mu_1(x)$ and the same regression for group with $X_i \in [c-h, c)$:

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

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

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 $X_i \in [c, c+h]$ to estimate $\mu_1(x)$ and the same regression for group with $X_i \in [c-h, c)$:

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

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

- Our estimate is

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

More practical estimation

- We can estimate this local linear regression by dropping observations more than h away from c and then running the following regression:

$$Y_i = \alpha + \beta(X_i - c) + \tau A_i + \gamma(X_i - c)A_i + \eta_i$$

More practical estimation

- We can estimate this local linear regression by dropping observations more than h away from c and then running the following regression:

$$Y_i = \alpha + \beta(X_i - c) + \tau A_i + \gamma(X_i - c)A_i + \eta_i$$

- Here we just have an interaction term between the treatment status and the forcing variable.

More practical estimation

- We can estimate this local linear regression by dropping observations more than h away from c and then running the following regression:

$$Y_i = \alpha + \beta(X_i - c) + \tau A_i + \gamma(X_i - c)A_i + \eta_i$$

- Here we just have an interaction term between the treatment status and the forcing variable.
- Here, $\widehat{\tau}_{SRD} = \widehat{\tau}$ which is the coefficient on the treatment.

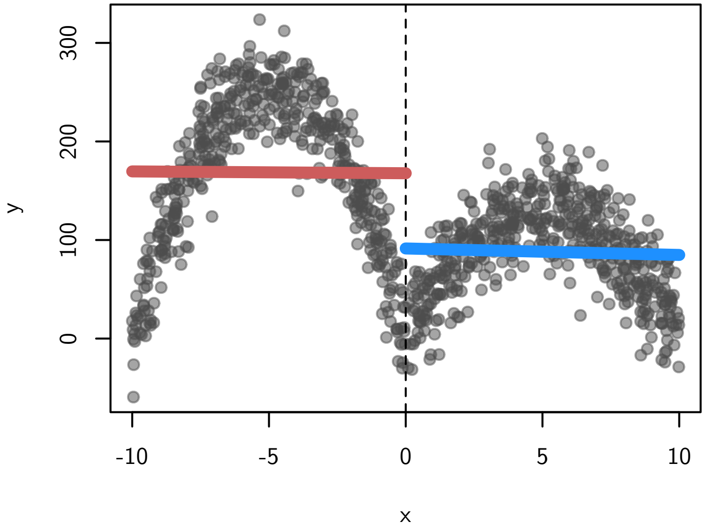
More practical estimation

- We can estimate this local linear regression by dropping observations more than h away from c and then running the following regression:

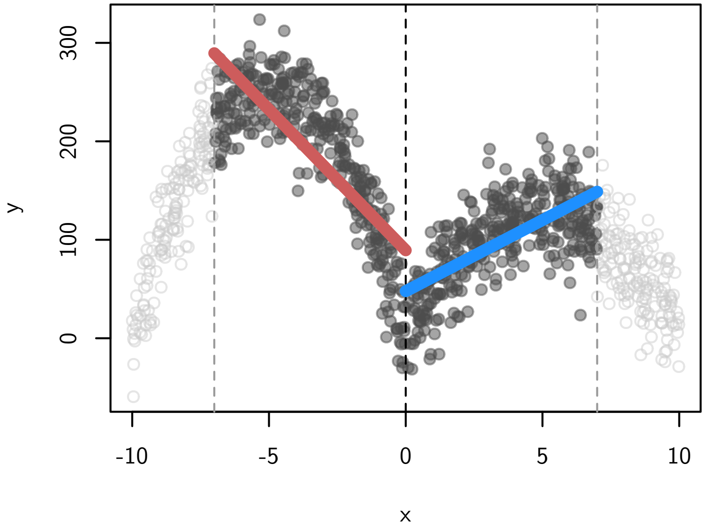
$$Y_i = \alpha + \beta(X_i - c) + \tau A_i + \gamma(X_i - c)A_i + \eta_i$$

- Here we just have an interaction term between the treatment status and the forcing variable.
- Here, $\widehat{\tau}_{SRD} = \widehat{\tau}$ which is the coefficient on the treatment.
- Yields numerically the same as the separate regressions.

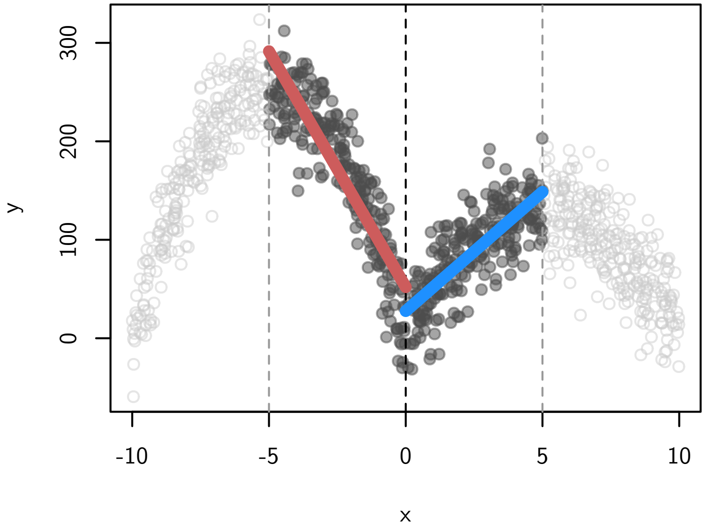
Bandwidth equal to 10 (Global)



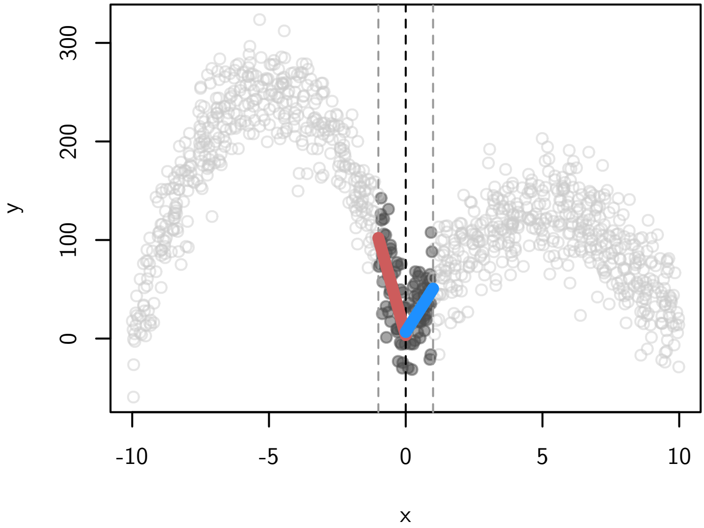
Bandwidth equal to 7



Bandwidth equal to 5



Bandwidth equal to 1



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.
- Covariates: shouldn't matter, but can include them for increased precision.

Odds and ends for the SRD

- Standard errors: robust standard errors from local OLS are valid.
- Covariates: shouldn't matter, but can include them for increased precision.
- ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST

Odds and ends for the SRD

- Standard errors: robust standard errors from local OLS are valid.
- Covariates: shouldn't matter, but can include them for increased precision.
- ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST
- You can include polynomials of the forcing variable in the local regression. Let $\tilde{X}_i = X_i - c$

$$Y_i = \alpha + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \tau A_i + \gamma_1 \tilde{X}_i A_i + \gamma_2 \tilde{X}_i^2 A_i + \eta_i$$

Odds and ends for the SRD

- Standard errors: robust standard errors from local OLS are valid.
- Covariates: shouldn't matter, but can include them for increased precision.
- ALWAYS REPORT MODELS WITHOUT COVARIATES FIRST
- You can include polynomials of the forcing variable in the local regression. Let $\tilde{X}_i = X_i - c$

$$Y_i = \alpha + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \tau A_i + \gamma_1 \tilde{X}_i A_i + \gamma_2 \tilde{X}_i^2 A_i + \eta_i$$

- Make sure that your effects aren't dependent on the polynomial choice.

Bandwidth selection

- The choice of bandwidth is fairly important here and we want it to be smaller as N grows.

Bandwidth selection

- The choice of bandwidth is fairly important here and we want it to be smaller as N grows.
- In general, we can use cross-validation techniques to choose the optimal bandwidth.

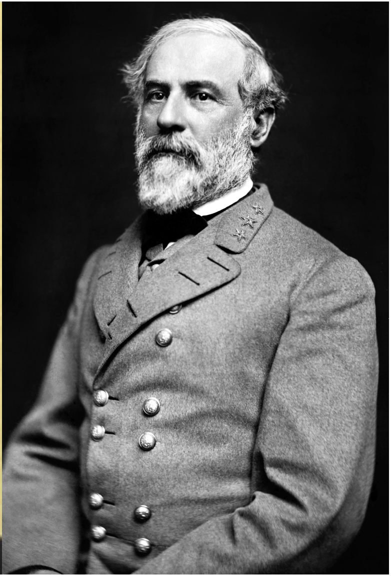
Bandwidth selection

- The choice of bandwidth is fairly important here and we want it to be smaller as N grows.
- In general, we can use cross-validation techniques to choose the optimal bandwidth.
- See Imbens and Kalyanaraman (2012) for optimal bandwidth selection.

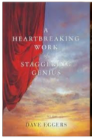
Readings

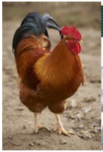
Reading 1

Reading 1

Reading 2





Fuzzy Regression Discontinuity Designs

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

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

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

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

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

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

- In the sharp RD, this is also true, but it further requried the jump in probability to be from 0 to 1.

Setup

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

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

- In the sharp RD, this is also true, but it further requried the jump in probability to be from 0 to 1.
- Fuzzy RD is often useful when the a threshold encourages participation in program, but does not actually force units to participate.

Fuzzy RD in a graph

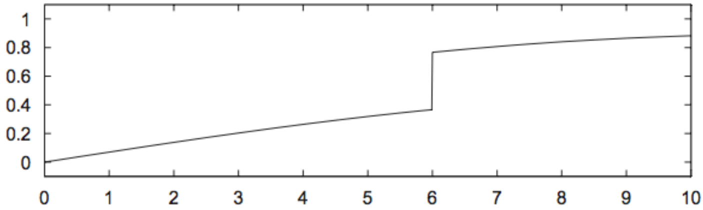


Fig. 3. Assignment probabilities (FRD).

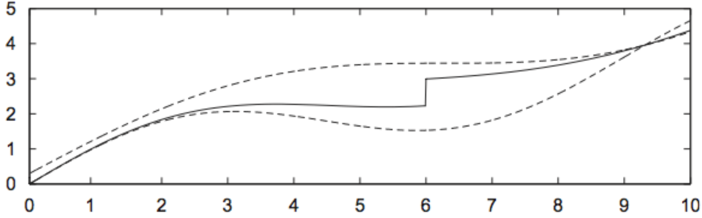


Fig. 4. Potential and observed outcome regression (FRD).

Fuzzy RD is IV

- Forcing variable is an **instrument**:

Fuzzy RD is IV

- Forcing variable is an **instrument**:
  - affects Y_i, but only through A_i (at the threshold)

Fuzzy RD is IV

- Forcing variable is an **instrument**:
  - affects Y_i, but only through A_i (at the threshold)
- Let $A_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.

Fuzzy RD is IV

- Forcing variable is an **instrument**:
  - affects Y_i, but only through A_i (at the threshold)
- Let $A_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.
- $A_i(x) = 1$ if unit i would take treatment when X_i was x

Fuzzy RD is IV

- Forcing variable is an **instrument**:
  - affects Y_i, but only through A_i (at the threshold)
- Let $A_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.
- $A_i(x) = 1$ if unit i would take treatment when X_i was x
- $A_i(x) = 0$ if unit i would take control when X_i was x

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

- Increasing the forcing variable doesn't encourage people to take the treatment

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

- Increasing the forcing variable doesn't encourage people to take the treatment

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

- Increasing the forcing variable doesn't encourage people to take the treatment

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, A_i(x)\} \perp\!\!\!\perp X_i$$

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

- Increasing the forcing variable doesn't encourage people to take the treatment

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, A_i(x)\} \perp\!\!\!\perp X_i$$

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $A_i(c+e) \geq A_i(c-e)$ for all $0 < e < \varepsilon$

- Increasing the forcing variable doesn't encourage people to take the treatment

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, A_i(x)\} \perp\!\!\!\perp X_i$$

- Basically, in an ε-ball around c, the forcing variable is randomly assigned.

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

- Think about college students that get above a certain GPA are encouraged to apply to grad school.

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

- Think about college students that get above a certain GPA are encouraged to apply to grad school.
- Compliers would:

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

- Think about college students that get above a certain GPA are encouraged to apply to grad school.
- Compliers would:
  - apply to grad school if their GPA was just above the threshold

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

- Think about college students that get above a certain GPA are encouraged to apply to grad school.
- Compliers would:
  - apply to grad school if their GPA was just above the threshold
  - not apply to grad school if their GPA was just below the threshold

Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

- Think about college students that get above a certain GPA are encouraged to apply to grad school.
- Compliers would:
  - apply to grad school if their GPA was just above the threshold
  - not apply to grad school if their GPA was just below the threshold
- We don't get to see their compliance status because due to the fundamental problem of causal inference

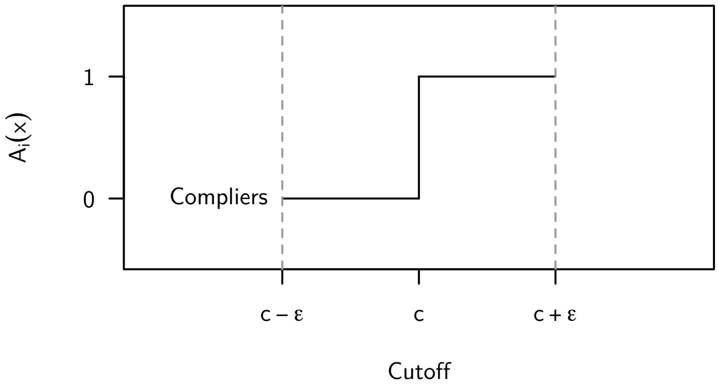
Compliance in Fuzzy RDs

- Compliers are those i such that for all $0 < e < \varepsilon$:

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

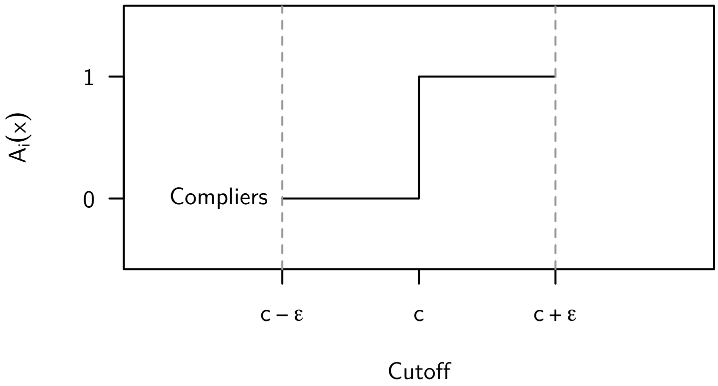
- Think about college students that get above a certain GPA are encouraged to apply to grad school.
- Compliers would:
  - apply to grad school if their GPA was just above the threshold
  - not apply to grad school if their GPA was just below the threshold
- We don't get to see their compliance status because due to the fundamental problem of causal inference
- Could also think about this as changing the threshold instead of changing X_i

Compliance graph



- Compliers would not take the treatment if they had $X_i = c$ and we increased the cutoff by some small amount

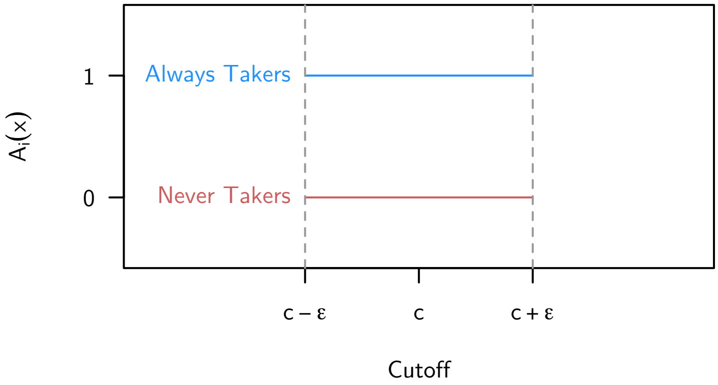
Compliance graph



- 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

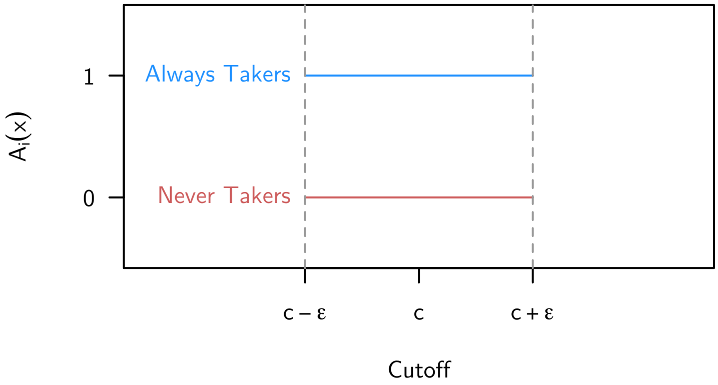
Compliance groups

- Compliers: $A_i(c + e) = 1$ and $A_i(c - e) = 0$



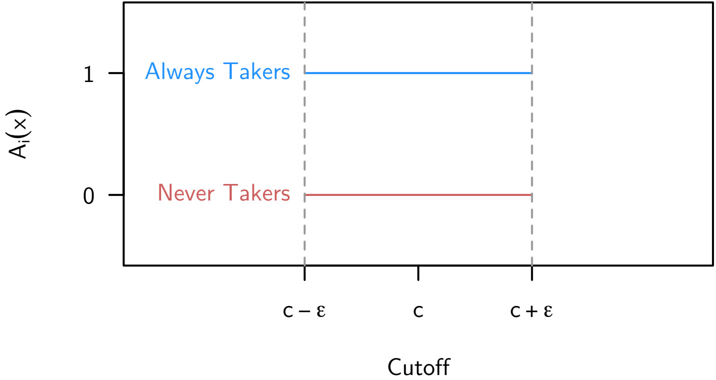
Compliance groups

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



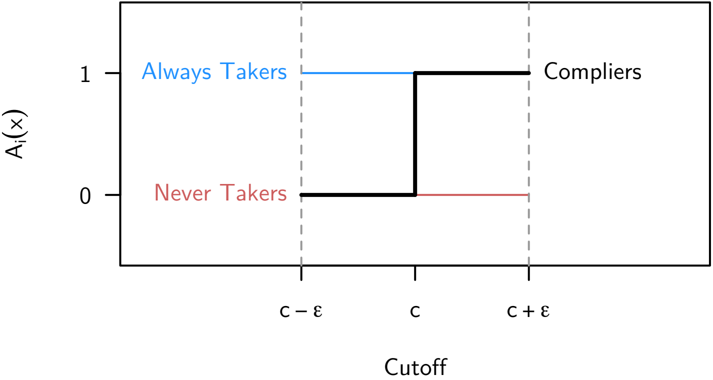
Compliance groups

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



Compliance groups

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



LATE in the Fuzzy RD

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

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

LATE in the Fuzzy RD

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

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

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

$$\tau_{FRD} = \lim_{e \downarrow 0} E[\tau_i | A_i(c+e) > A_i(c-e)]$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i | X_i = c + e] - E[Y_i | X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \end{aligned}$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \\ &= Y_i(0) + \tau_i A_i \end{aligned}$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \\ &= Y_i(0) + \tau_i A_i \end{aligned}$$

- Plug this into the CEF of the outcome:

$$E[Y_i|X_i = c + e] = E[Y_i(0) + \tau_i A_i|X_i = c + e]$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \\ &= Y_i(0) + \tau_i A_i \end{aligned}$$

- Plug this into the CEF of the outcome:

$$E[Y_i|X_i = c + e] = E[Y_i(0) + \tau_i A_i|X_i = c + e]$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \\ &= Y_i(0) + \tau_i A_i \end{aligned}$$

- Plug this into the CEF of the outcome:

$$\begin{aligned} E[Y_i|X_i = c + e] &= E[Y_i(0) + \tau_i A_i|X_i = c + e] \\ &= E[Y_i(0) + \tau_i A_i(c + e)] \end{aligned}$$

Proof

- To prove this, we'll look at the discontinuity in Y_i in a window around the threshold and then shrink that window:

$$E[Y_i|X_i = c + e] - E[Y_i|X_i = c - e]$$

- First, remember that by consistency,

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))A_i \\ &= Y_i(0) + \tau_i A_i \end{aligned}$$

- Plug this into the CEF of the outcome:

$$\begin{aligned} E[Y_i|X_i = c + e] &= E[Y_i(0) + \tau_i A_i|X_i = c + e] \\ &= E[Y_i(0) + \tau_i A_i(c + e)] \end{aligned}$$

- Thus, we can write the difference around the threshold as:

$$E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] = E[\tau_i(A_i(c+e) - A_i(c-e))]$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

$$\begin{aligned}
&E[\tau_i \times 1 \mid \text{complier}] \times \Pr[\text{complier}] \\
+&E[\tau_i \times -1 \mid \text{defier}] \times \Pr[\text{defier}] \\
+&E[\tau_i \times (A_i(c+e) - A_i(c-e)) \mid \text{always}] \times \Pr[\text{always}] \\
+&E[\tau_i \times (A_i(c+e) - A_i(c-e)) \mid \text{never}] \times \Pr[\text{never}]
\end{aligned}$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

$$\begin{aligned}
&E[\tau_i \times 1 \mid \text{complier}\,] \times \Pr[\text{complier}] \\
+&E[\tau_i \times \; -1 \mid \text{defier}] \times 0 \\
+&E[\tau_i \times (A_i(c+e) - A_i(c-e)) \mid \text{always}] \times \Pr[\text{always}] \\
+&E[\tau_i \times (A_i(c+e) - A_i(c-e)) \mid \text{never}] \times \Pr[\text{never}]
\end{aligned}$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

$$\begin{aligned}
&E[\tau_i \times 1 \mid \text{complier }] \times \Pr[\text{complier}] \\
+&E[\tau_i \times \; -1 \mid \text{defier}] \times 0 \\
+&E[\tau_i \times 0 \mid \text{always}] \times \Pr[\text{always}] \\
+&E[\tau_i \times 0 \mid \text{never}] \times \Pr[\text{never}]
\end{aligned}$$

Proof (cont)

- Let's break this expectation apart using the law of iterated expectations:

$$E[\tau_i(A_i(c+e) - A_i(c-e))] =$$

$$\begin{aligned}
&E[\tau_i \times 1 \mid \text{complier}\ ] \times \Pr[\text{complier}] \\
+&E[\tau_i \times\ -1 \mid \text{defier}] \times 0 \\
+&E[\tau_i \times 0 \mid \text{always}] \times \Pr[\text{always}] \\
+&E[\tau_i \times 0 \mid \text{never}] \times \Pr[\text{never}]
\end{aligned}$$

$$= E[\tau_i \mid \text{complier}] \times \Pr[\text{complier}]$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e]-E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}]\times\Pr[\text{complier}]$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e]-E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}]\times \Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\Pr[\text{complier}] = \Pr[A_i(c+e) - A_i(c-e) = 1]$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}] \times \Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\Pr[\text{complier}] = \Pr[A_i(c+e) - A_i(c-e) = 1]$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}] \times \Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\begin{aligned} \Pr[\text{complier}] &= \Pr[A_i(c+e) - A_i(c-e) = 1] \\ &= E[A_i(c+e) - A_i(c-e)] \end{aligned}$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}] \times \Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\begin{aligned} \Pr[\text{complier}] &= \Pr[A_i(c+e) - A_i(c-e) = 1] \\ &= E[A_i(c+e) - A_i(c-e)] \\ &= E[A_i(c+e)] - E[A_i(c-e)] \end{aligned}$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e]-E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}]\times\Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\begin{aligned}
\Pr[\text{complier}] &= \Pr[A_i(c+e) - A_i(c-e) = 1] \\
&= E[A_i(c+e) - A_i(c-e)] \\
&= E[A_i(c+e)] - E[A_i(c-e)] \\
&= E[A_i(c+e)|X_i = c+e] - E[A_i(c-e)|X_i = c-e]
\end{aligned}$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e]-E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}]\times\Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\begin{aligned}
\Pr[\text{complier}] &= \Pr[A_i(c+e) - A_i(c-e) = 1] \\
&= E[A_i(c+e) - A_i(c-e)] \\
&= E[A_i(c+e)] - E[A_i(c-e)] \\
&= E[A_i(c+e)|X_i = c+e] - E[A_i(c-e)|X_i = c-e] \\
&= E[A_i|X_i = c+e] - E[A_i|X_i = c-e]
\end{aligned}$$

Proof (cont)

- So far, we've shown that the outcome jump at the discontinuity is the LATE times the probability of compliance:

$$E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e] = E[\tau_i \mid \text{complier}] \times \Pr[\text{complier}]$$

- What is the probability of compliance though?

$$\begin{aligned}
\Pr[\text{complier}] &= \Pr[A_i(c+e) - A_i(c-e) = 1] \\
&= E[A_i(c+e) - A_i(c-e)] \\
&= E[A_i(c+e)] - E[A_i(c-e)] \\
&= E[A_i(c+e)|X_i = c+e] - E[A_i(c-e)|X_i = c-e] \\
&= E[A_i|X_i = c+e] - E[A_i|X_i = c-e]
\end{aligned}$$

- Thus,

$$\frac{E[Y_i|X_i = c+e] - E[Y_i|X_i = c-e]}{E[A_i|X_i = c+e] - E[A_i|X_i = c-e]} = E[\tau_i \mid A_i(c+e) > A_i(c-e)]$$

Misc notes

- Taking the limit as $e \to 0$, we've shown that:

$$\begin{aligned}\tau_{FRD} &= \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]} \\ &= \lim_{e \downarrow 0} E[\tau_i|A_i(c+e) > A_i(c-e)]\end{aligned}$$

Misc notes

- Taking the limit as $e \to 0$, we've shown that:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]}$$
$$= \lim_{e \downarrow 0} E[\tau_i | A_i(c+e) > A_i(c-e)]$$

- Note that the FRD estimator emcompasses the SRD estimator because with a sharp design:

$$\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x] = 1$$

Misc notes

- Taking the limit as $e \to 0$, we've shown that:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]}$$
$$= \lim_{e \downarrow 0} E[\tau_i | A_i(c+e) > A_i(c-e)]$$

- Note that the FRD estimator emcompasses the SRD estimator because with a sharp design:

$$\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x] = 1$$

- A note on external validity: obsviously, FRD puts even more restrictions on the external validity of our estimates because not only are we discussing a LATE, but also the effect is at the threshold. That might give us pause about generalizing other populations for the both the SRD and FRD.

Estimation in FRD

- Remember that we had:
$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]}$$

Estimation in FRD

- Remember that we had:
$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]}$$
- We can estimate the numerator using the SRD approaches we just outlined, $\widehat{\tau}_{SRD}$.

Estimation in FRD

- Remember that we had:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]}{\lim_{x \downarrow c} E[A_i | X_i = x] - \lim_{x \uparrow c} E[A_i | X_i = x]}$$

- We can estimate the numerator using the SRD approaches we just outlined, $\widehat{\tau}_{SRD}$.
- For the denominator, we simply apply the local linear regression to the A_i:

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

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

Estimation in FRD

- Remember that we had:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[A_i|X_i = x] - \lim_{x \uparrow c} E[A_i|X_i = x]}$$

- We can estimate the numerator using the SRD approaches we just outlined, $\widehat{\tau}_{SRD}$.
- For the denominator, we simply apply the local linear regression to the A_i:

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

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

- Use this to calculate the effect of threshold on A_i: $\widehat{\tau}_a = \widehat{\alpha}_{aR} - \widehat{\alpha}_{aL}$

Estimation in FRD

- Remember that we had:

$$\tau_{FRD} = \frac{\lim_{x\downarrow c} E[Y_i|X_i = x] - \lim_{x\uparrow c} E[Y_i|X_i = x]}{\lim_{x\downarrow c} E[A_i|X_i = x] - \lim_{x\uparrow c} E[A_i|X_i = x]}$$

- We can estimate the numerator using the SRD approaches we just outlined, $\widehat{\tau}_{SRD}$.
- For the denominator, we simply apply the local linear regression to the A_i:

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

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

- Use this to calculate the effect of threshold on A_i: $\widehat{\tau}_a = \widehat{\alpha}_{aR} - \widehat{\alpha}_{aL}$
- Calculate ratio estimator:

$$\widehat{\tau}_{FRD} = \frac{\widehat{\tau}_{SRD}}{\widehat{\tau}_a}$$

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:

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

More practical FRD estimation

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

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

- First stage:

$$A_i = \delta_1' V_i + \rho \mathbb{I}\{X_i \geq c\} + \nu_i$$

More practical FRD estimation

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

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

- First stage:

$$A_i = \delta_1' V_i + \rho \mathbb{I}\{X_i \geq c\} + \nu_i$$

- Second stage:

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

More practical FRD estimation

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

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

- First stage:

$$A_i = \delta_1' V_i + \rho \mathbb{I}\{X_i \geq c\} + \nu_i$$

- Second stage:

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

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