Political Methodology II
RDD

Apoorva Lal
March 11, 2022
Stanford University

Roadmap

Roadmap

Recap of the course

Sharp RD

- Sharp RD Identification
- Sharp RD Estimation
  - Parametric Models

Fuzzy RDD

Example: Electronic Voting in Brazil

Crash course in DAGs

The experimental ideal

Recall the fundamental problem of causal inference:

The experimental ideal

Recall the fundamental problem of causal inference:

- We can never observe both potential outcomes for any given unit.

The experimental ideal

Recall the fundamental problem of causal inference:

- We can never observe both potential outcomes for any given unit.
- In the ideal world, we randomly assign the treatment D_i so that it is independent of all observed and unobserved confounders, i.e., $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i$. Then it's easy to get unbiased estimates of $\mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]$.

The experimental ideal

Recall the fundamental problem of causal inference:

- We can never observe both potential outcomes for any given unit.
- In the ideal world, we randomly assign the treatment D_i so that it is independent of all observed and unobserved confounders, i.e., $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i$. Then it's easy to get unbiased estimates of $\mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]$.
- This works because we know the assignment mechanism and we design it to enable causal inference.

Causal inference techniques

We've learned methods to approximate this experimental ideal (to get 'as-if' random assignment) under certain assumptions:

Causal inference techniques

We've learned methods to approximate this experimental ideal (to get 'as-if' random assignment) under certain assumptions:

- Selection on observables
    - Find some covariates X_i such that treatment is as good as randomly assigned after conditioning on the covariates: $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i \mid X_i$.
    - Then we can get unbiased estimates of $\mathbb{E}[Y_{1i} - Y_{0i} \mid X_i]$, and then take a weighted average over the values of X_i to get an overall treatment effect.

Causal inference techniques

We've learned methods to approximate this experimental ideal (to get 'as-if' random assignment) under certain assumptions:

- Selection on observables
  - Find some covariates X_i such that treatment is as good as randomly assigned after conditioning on the covariates: $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i \mid X_i$.
  - Then we can get unbiased estimates of $\mathbb{E}[Y_{1i} - Y_{0i} \mid X_i]$, and then take a weighted average over the values of X_i to get an overall treatment effect.
- Difference-in differences
  - We assume that the treatment assignment is independent of the trajectory of potential outcome under control that units are following over time ("parallel trends"): $\mathbb{E}[Y_{0i}(1) - Y_{0i}(0)] \perp\!\!\!\perp D_i$.
  - Then we can use untreated units in time $t = 1$ to impute what the potential outcomes for the treated units would have been at $t = 1$ if they hadn't been treated.

Causal inference techniques (cont'd)

- Panel methods
  - Assume that treatment is independent of particular kinds of unobserved determinants of the outcome.
  - Then use de-meaning procedures or fixed effects to estimate what would have happened to treated units if they hadn't been treated (i.e. estimate the ATT).

Causal inference techniques (cont'd)

- Panel methods
  - Assume that treatment is independent of particular kinds of unobserved determinants of the outcome.
  - Then use de-meaning procedures or fixed effects to estimate what would have happened to treated units if they hadn't been treated (i.e. estimate the ATT).
- Instrumental variables
  - Find another variable Z_i that affects the treatment but is independent of the potential outcomes: $(Y_{0i}, Y_{1i}, D_{0i}, D_{1i}) \perp\!\!\!\perp Z_i$ and $P(D_{1i} = 1) \neq P(D_{0i} = 1)$.
  - Use various techniques to estimate a local average treatment effect for the compliers.

Causal inference techniques (cont'd)

- Panel methods
  - Assume that treatment is independent of particular kinds of unobserved determinants of the outcome.
  - Then use de-meaning procedures or fixed effects to estimate what would have happened to treated units if they hadn't been treated (i.e. estimate the ATT).
- Instrumental variables
  - Find another variable Z_i that affects the treatment but is independent of the potential outcomes: $(Y_{0i}, Y_{1i}, D_{0i}, D_{1i}) \perp\!\!\!\perp Z_i$ and $P(D_{1i} = 1) \neq P(D_{0i} = 1)$.
  - Use various techniques to estimate a local average treatment effect for the compliers.

All of these methods rely on some (generally untestable) assumptions about how the treatment is assigned to various units.

Roadmap

Recap of the course

Sharp RD

- Sharp RD Identification
- Sharp RD Estimation
  - Parametric Models

Fuzzy RDD

Example: Electronic Voting in Brazil

Crash course in DAGs

Setup

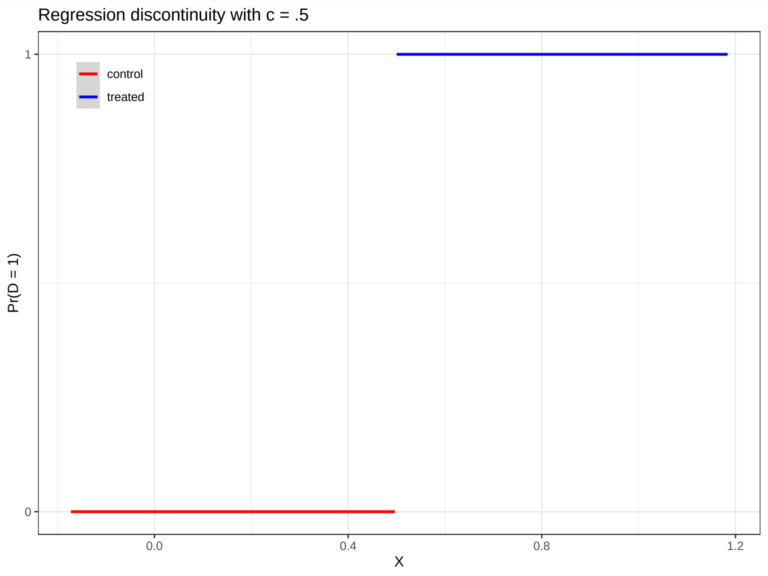
RDD Setup:

- X covariate, also called the forcing variable with a cutoff c that determines assignment to treatment
- Binary treatment variable:

$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

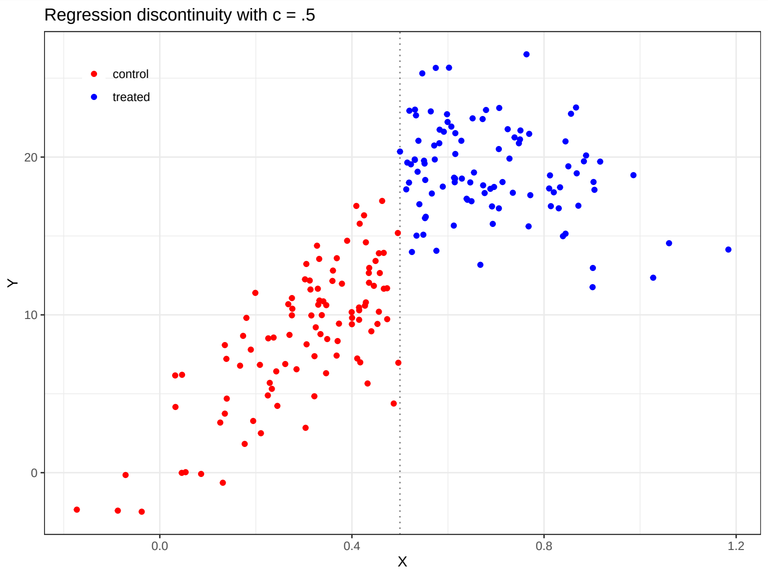
- Potential outcomes: Y_{1i}, Y_{0i}
- Y is the observed outcome: $Y_i = D_i \cdot Y_1 + (1 - D_i) \cdot Y_0$
- We allow the potential outcomes to be correlated with X
- We do require that potential outcomes be smooth around the threshold, c to identify the effect of treatment at the threshold

Example



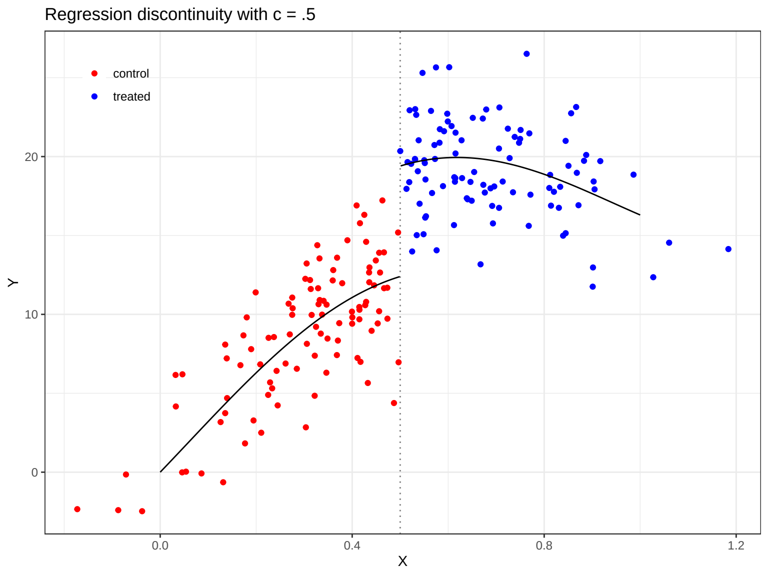
Regression discontinuity with c = .5
1
0
control
treated
Pr(D = 1)
0.0
0.4
0.8
1.2
X

Example



Regression discontinuity with c = .5
control
treated
20
10
0
Y
0.0
0.4
0.8
1.2
X

Example


Regression discontinuity with c = .5
control
treated
Y
X
0
10
20
0.0
0.4
0.8
1.2

Identification

Identification Assumption

$\mathbb{E}[Y_{1i}|X, D]$ and $\mathbb{E}[Y_{0i}|X, D]$ are continuous around the threshold, “imprecise control over assignment”

We have a new estimand: average treatment effect **at the threshold**:

$$\begin{aligned}\tau_{RDD} &= \mathbb{E}[Y_{1i} - Y_{0i}|X = c] \\ &= \mathbb{E}[Y_{1i}|X = c] - \mathbb{E}[Y_{0i}|X = c] \\ &= \lim_{x \to c^-} \mathbb{E}[Y_{1i}|X = x] - \lim_{x \to c^+} \mathbb{E}[Y_{0i}|X = x]\end{aligned}$$

Why is identification challenging in this case?

Identification

Identification Assumption

$\mathbb{E}[Y_{1i}|X,D]$ and $\mathbb{E}[Y_{0i}|X,D]$ are continuous around the threshold, "imprecise control over assignment"

We have a new estimand: average treatment effect **at the threshold**:

$$\begin{aligned}\tau_{RDD} &= \mathbb{E}[Y_{1i} - Y_{0i}|X = c] \\ &= \mathbb{E}[Y_{1i}|X = c] - \mathbb{E}[Y_{0i}|X = c] \\ &= \lim_{x \to c^-} \mathbb{E}[Y_{1i}|X = x] - \lim_{x \to c^+} \mathbb{E}[Y_{0i}|X = x]\end{aligned}$$

Why is identification challenging in this case? **Lack of overlap**: There is no value of X at which we get to observe both treatment and control observations.

Identification

Identification Assumption

$\mathbb{E}[Y_{1i}|X, D]$ and $\mathbb{E}[Y_{0i}|X, D]$ are continuous around the threshold, "imprecise control over assignment"

We have a new estimand: average treatment effect **at the threshold**:

$$\begin{aligned} \tau_{RDD} &= \mathbb{E}[Y_{1i} - Y_{0i}|X = c] \\ &= \mathbb{E}[Y_{1i}|X = c] - \mathbb{E}[Y_{0i}|X = c] \\ &= \lim_{x \to c^-} \mathbb{E}[Y_{1i}|X = x] - \lim_{x \to c^+} \mathbb{E}[Y_{0i}|X = x] \end{aligned}$$

Why is identification challenging in this case? **Lack of overlap**: There is no value of X at which we get to observe both treatment and control observations.The effect is identified at the limit. This is why functional form is going to matter.

Local Randomisation

- Sometimes see RD model/method analyzed and described in terms of ”local randomization” (Cattaneo Frandsen Titiunik 2015)
- Near cutoff, whether you are on one side or another is “”effectively random” and you can treat data “as if” an experiment near the cutoff
- Formalization: $Y^{d,x} = Y^d$ for $x \in [\underline{c}, \bar{c}]$ $\underline{c} < c < \bar{c}$ and $X \perp (Y^0, Y^1) | X \in [\underline{c}, \bar{c}]$
- In this situation, **no** effect of or association of running variable with outcomes in a window
- This is much stronger assumption than continuity
    - Asks for potential outcomes functions to be exactly flat near the cutoff
    - Maybe plausible if assignment rule based on variable not at all important to outcome
- Aside from deliberate random assignment, this is implausible
- But if you do have that, `rdlocrand` implements Fisher-style tests

Estimation

Given this identification result, how should we try to estimate the jump at the discontinuity?

Estimation

Given this identification result, how should we try to estimate the jump at the discontinuity?

- Recall the we need to estimate

$$\lim_{x \to c^+} \mathbb{E}[Y_{1i} \mid X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_{0i} \mid X_i = x].$$

Estimation

Given this identification result, how should we try to estimate the jump at the discontinuity?

- Recall the we need to estimate

$$\lim_{x \to c^+} \mathbb{E}[Y_{1i} \mid X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_{0i} \mid X_i = x].$$

- When $X_i > c$, we know that $D_i = 1$, and when $X_i < c$, we know that $D_i = 0$.

Estimation

Given this identification result, how should we try to estimate the jump at the discontinuity?

- Recall the we need to estimate

$$\lim_{x \to c^+} \mathbb{E}[Y_{1i} \mid X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_{0i} \mid X_i = x].$$

- When $X_i > c$, we know that $D_i = 1$, and when $X_i < c$, we know that $D_i = 0$.
- This means that we can use the data points above and below the cutoff to model $\mathbb{E}[Y_{1i} \mid X_i]$ and $[Y_{0i} \mid X_i]$, respectively.

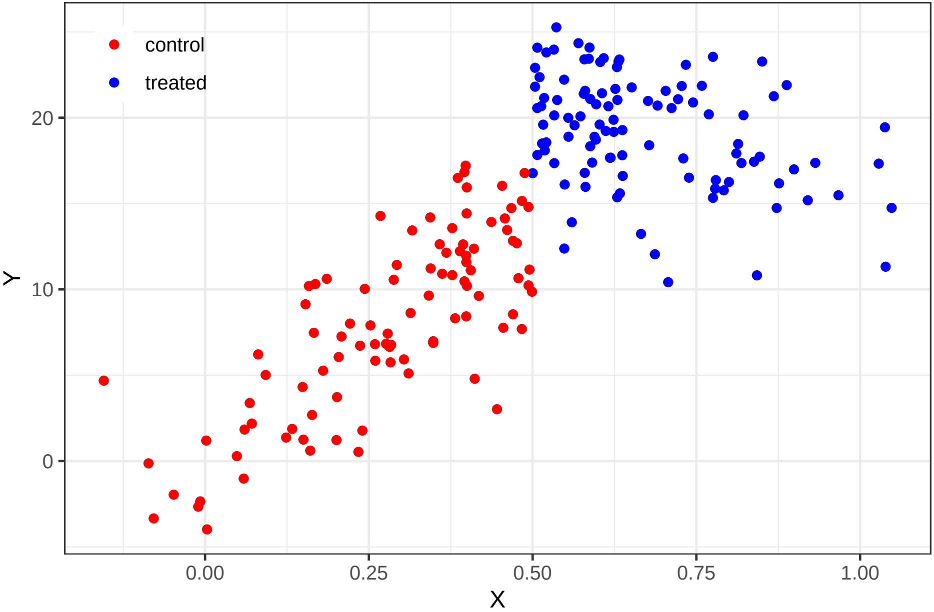
Estimation

Given this identification result, how should we try to estimate the jump at the discontinuity?

- Recall the we need to estimate

$$\lim_{x \to c^+} \mathbb{E}[Y_{1i} \mid X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_{0i} \mid X_i = x].$$

- When $X_i > c$, we know that $D_i = 1$, and when $X_i < c$, we know that $D_i = 0$.
- This means that we can use the data points above and below the cutoff to model $\mathbb{E}[Y_{1i} \mid X_i]$ and $[Y_{0i} \mid X_i]$, respectively.
- This is a curve fitting problem. Out of habit, we start with linear functions, then use more flexible methods [**450c** is about this]
- Intuition is to estimate a regression function on either side of the cutoff, then to find the gap between the predicted values at the cutoff c from the right side regression and the left side regression.


control
treated
Y
20
10
0
X
0.00
0.25
0.50
0.75
1.00

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α:

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1:

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$
- β_2:

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$
- β_2: The expected difference in Y between treated and control units when $X = 0$

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$
- β_2: The expected difference in Y between treated and control units when $X = 0$
- β_3:

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$
- β_2: The expected difference in Y between treated and control units when $X = 0$
- β_3: The difference in the marginal effect of X between the treated and control groups

Estimation: Linear model

Let's estimate the following regression:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

- α: The expected value of Y when $X = 0$ and $D = 0$
- β_1: The expected increase in Y for a one-unit change in X when $D = 0$
- β_2: The expected difference in Y between treated and control units when $X = 0$
- β_3: The difference in the marginal effect of X between the treated and control groups

Estimation: Linear model

Regression equation:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

What's the effect of a one-unit increase in X for treated units?

Estimation: Linear model

Regression equation:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

What's the effect of a one-unit increase in X for treated units?

$$\frac{\partial \ \mathbb{E}[Y_i \mid X_i, D_i = 1]}{\partial X_i} = \beta_1 + \beta_3 \cdot 1$$

Estimation: Linear model

Regression equation:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

What's the effect of a one-unit increase in X for treated units?

$$\frac{\partial \, \mathbb{E}[Y_i \mid X_i, D_i = 1]}{\partial X_i} = \beta_1 + \beta_3 \cdot 1$$

What's the partial effect of D at the threshold?

Estimation: Linear model

Regression equation:

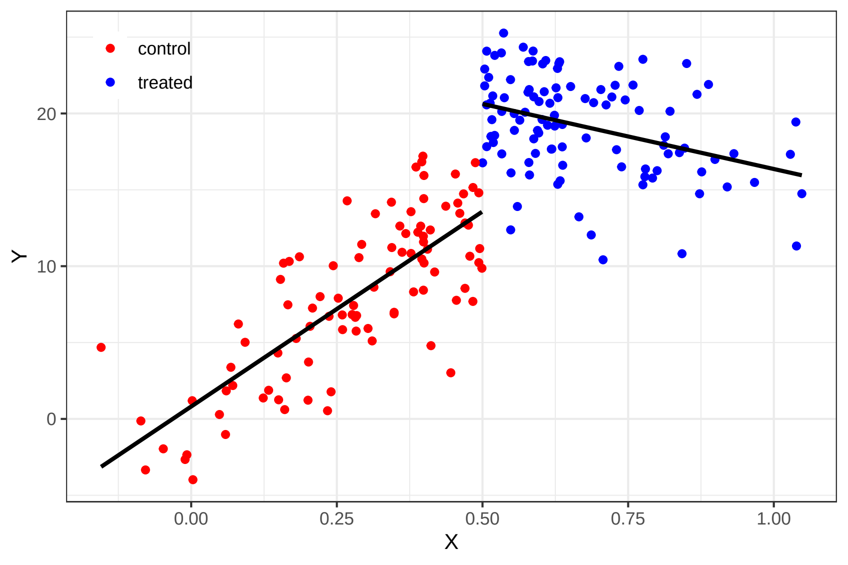
$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i \cdot D_i + u_i$$

What's the effect of a one-unit increase in X for treated units?

$$\frac{\partial\ \mathbb{E}[Y_i \mid X_i, D_i = 1]}{\partial X_i} = \beta_1 + \beta_3 \cdot 1$$

What's the partial effect of D at the threshold?

$$\frac{\partial\ \mathbb{E}[Y_i \mid D_i, X_i = c]}{\partial D_i} = \beta_2 + \beta_3 \cdot c$$


control
treated
20
10
0
Y
0.00
0.25
0.50
0.75
1.00
X

Estimation: Linear model

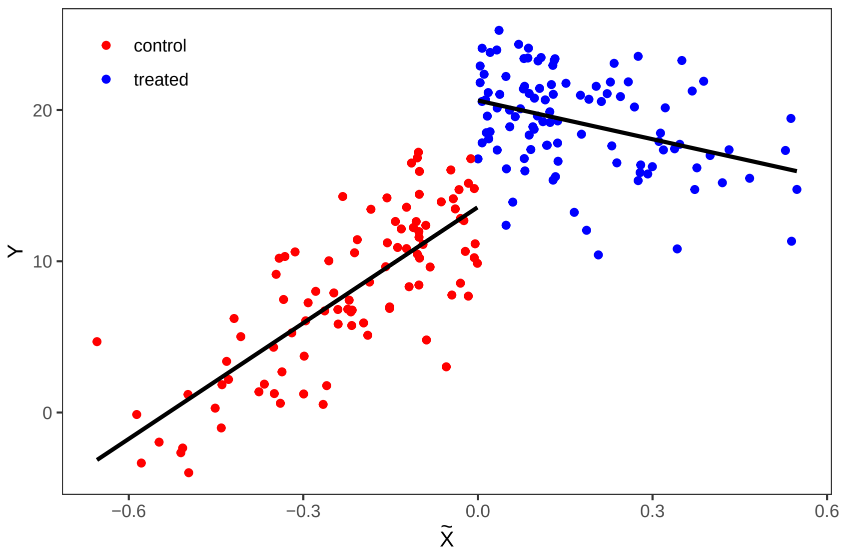
It's common practice to subtract c from X, so the cutoff is redefined as $\tilde{c} = 0$. Then we can run the regression

$$Y_i = \alpha + \beta_1 \tilde{X}_i + \beta_2 D_i + \beta_3 \tilde{X}_i \cdot D_i + u_i,$$

and now β_2 is directly interpretable as the differences in $\mathbb{E}[Y_1 \mid X]$ and $\mathbb{E}[Y_0 \mid X]$ at the cutoff.

```r
dat$x.tilde = dat$X - .5
mod = lm(Y ~ D + x.tilde + D:x.tilde, dat)
coeftest(mod, vcov. = vcovHC(mod, "HC2"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  13.58153    0.61555  22.064 < 2.2e-16 ***
## D             7.04357    0.76171   9.247 < 2.2e-16 ***
## x.tilde      25.53883    2.39909  10.645 < 2.2e-16 ***
## D:x.tilde   -34.06120    3.15705 -10.789 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

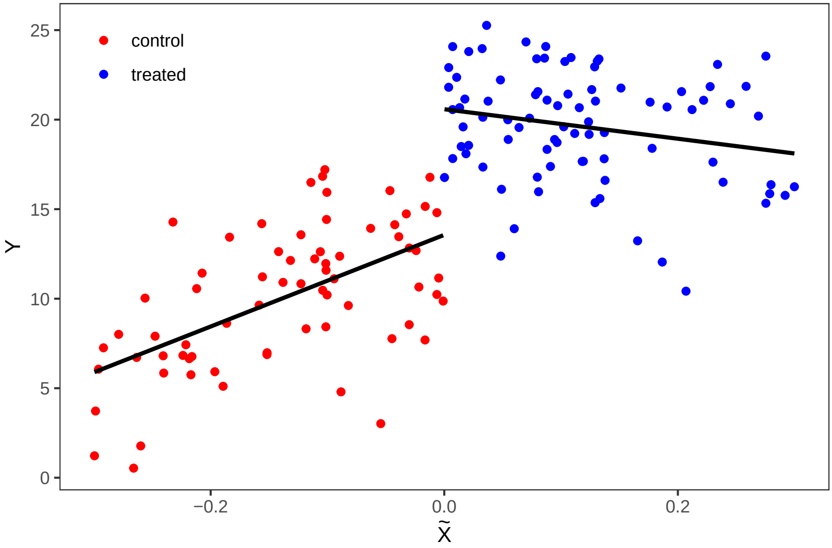

control
treated
20
10
0
Y
−0.6
−0.3
0.0
0.3
0.6
$\tilde{X}$

Estimation: Narrowing the bandwidth

We don't have global unconfoundedness, so using a bandwidth that is 'too wide' will bias the estimate. Common practice to use local methods to estimate the conditional expectation function.

```
# trim data to X within .3 of cutoff
trimdat = dat[abs(dat$x.tilde) < .3, ]
trimmod = lm(Y ~ D + x.tilde + D:x.tilde, trimdat)
coeftest(trimmod, vcov. = vcovHC(trimmod, "HC2"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  13.56527    0.76122 17.8205 < 2.2e-16 ***
## D             7.01668    0.93197  7.5289 4.531e-12 ***
## x.tilde      25.59844    4.48616  5.7061 6.039e-08 ***
## D:x.tilde   -33.84314    6.04762 -5.5961 1.023e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


control
treated
25
20
15
10
5
0
Y
−0.2
0.0
0.2
$\tilde{X}$

Adding flexibility

We can model the conditional expectation function more flexibly by specifying higher-order polynomials or using kernel methods:

```
quadmod = lm(Y ~ D * (x.tilde + I(x.tilde^2)), trimdat)
coeftest(quadmod, vcov. = vcovHC(quadmod, "HC2"))
```

```
##
## t test of coefficients:
##
##                  Estimate Std. Error t value  Pr(>|t|)
## (Intercept)       11.8269     0.9580 12.3454 < 2.2e-16 ***
## D                  8.6089     1.2091  7.1201 4.436e-11 ***
## x.tilde          -11.9285    14.2598 -0.8365  0.404223
## I(x.tilde^2)    -128.6189    46.7517 -2.7511  0.006687 **
## D:x.tilde          7.0535    19.1304  0.3687  0.712876
## D:I(x.tilde^2)   116.5644    63.5090  1.8354  0.068467 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Why stop at quadratic?

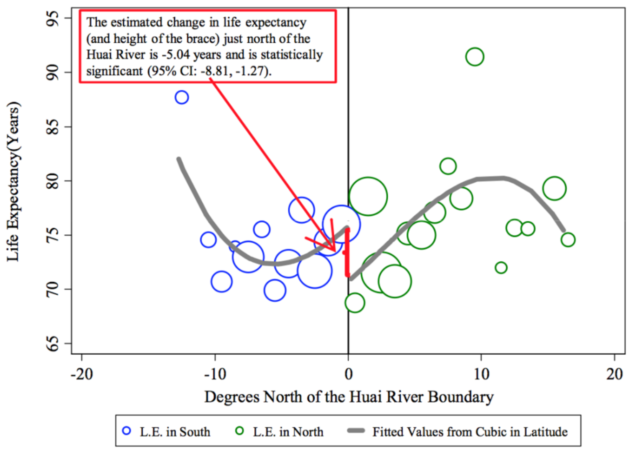


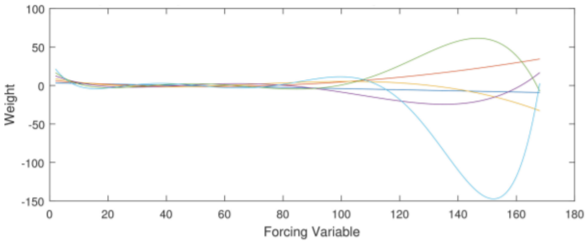
Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

This introduces strong functional form dependence

Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs

Andrew Gelman
Department of Statistics and Department of Political Science, Columbia University, New York, NY, 10027 (*gelman@stat.columbia.edu*)

Guido Imbens
Graduate School of Business, Stanford University, Stanford, CA 94305, and NBER, Stanford University, Stanford, CA 94305 (*imbens@stanford.edu*)



Estimation: Local regression

- Continuity is fairly weak notion formalizing small changes in x leading to small changes in Y^d
- Simple implementation is to take window of width δ and estimate expectations by conditional means

$$\widehat{\beta}_0^- = \frac{\sum_{i=1}^n Y_i 1\{c-\delta < X_i < c\}}{\sum_{i=1}^n 1\{c-\delta < X_i < c\}}$$
$$\widehat{\beta}_0^+ = \frac{\sum_{i=1}^n Y_i 1\{c \le X_i < c+\delta\}}{\sum_{i=1}^n 1\{c \le X_i < c+\delta\}}$$
$$\widehat{\tau}^{RD} = \widehat{\beta}_0^+ - \widehat{\beta}_0^-$$

- This gets called the local constant or Nadaraya-Watson estimator
- Issue of what to assume about local response of conditional mean, and based on it, how much to shrink window, is main source of difficulty

Estimation: Local Polynomials

- Idea: to estimate function near a point, use function estimation methods, but use only data near that point. Choose a window around x and run (weighted) regression in that window
- Account for slope and curvature and extrapolate locally by using polynomial of order p
  - $p = 0$ is local constant, $p = 1$ is local linear (default), $p = 2$ is local quadratic
- Let $K(u)$ be a kernel function weighting points within window
- Boxcar/Uniform: $K(u) = 1\{|u| \leq 1\}$, Triangular: $K(u) = \max\{0, (1 - |u|)\}$, Epanechnikov: $K(u) = \max\{0, \frac{3}{4}(1 - u^2)\}$
- Choose how spread out to make the average by adjusting bandwidth h

$$\widehat{\beta}^-(x) := (\widehat{\beta}_0^-(x), \hat{\beta}_1^-(x), \ldots) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} 1\{x_i < c\} K\left(\frac{x_i - x}{h}\right) \left(Y_i - \sum_{j=1}^{p} \beta_j (x_i - x)^j\right)^2$$

$$\widehat{\beta}^+(x) := (\widehat{\beta}_0^+(x), \hat{\beta}_1^+(x), \ldots) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} 1\{x_i \geq c\} K\left(\frac{x_i - x}{h}\right) \left(Y_i - \sum_{j=1}^{p} \beta_j (x_i - x)^j\right)^2$$

$$\widehat{\tau}^{RD} = \widehat{\beta}_0^+(c) - \widehat{\beta}_0^-(c)$$

- With uniform kernel, this is just regression in a window

The ‘right’ way to add flexibility: Nonparametric CEF estimators

Local-Linear Regression : `rdrobust` from Calonico, Cattaneo, and Titiunik (2014)

Separate bandwidth selection and estimation steps.

The ‘right’ way to add flexibility: Nonparametric CEF estimators

Local-Linear Regression : `rdrobust` from Calonico, Cattaneo, and Titiunik (2014)

Separate bandwidth selection and estimation steps. Antecedents:

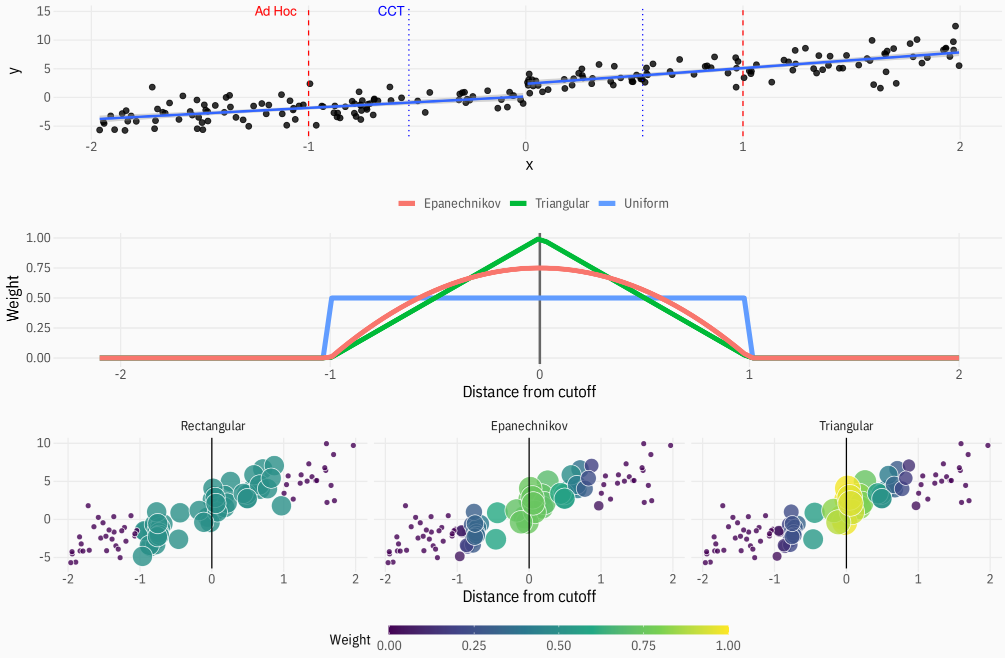
Imbens-Kalyanaraman (IK) 2012

```
library(rdrobust)
nonparametric = rdrobust(dat$Y, dat$x.tilde)
cbind(nonparametric$coef, nonparametric$se, nonparametric$z)


##                     Coeff Std. Err.        z
## Conventional    8.430291  1.173172 7.185892
## Bias-Corrected 8.788016  1.173172 7.490814
## Robust          8.788016  1.319503 6.660096
```

3

Subsequent work: **Imbens-Wager 2018**


Ad Hoc
CCT
y
x
Epanechnikov
Triangular
Uniform
Weight
Distance from cutoff
Rectangular
Epanechnikov
Triangular
Distance from cutoff
Weight

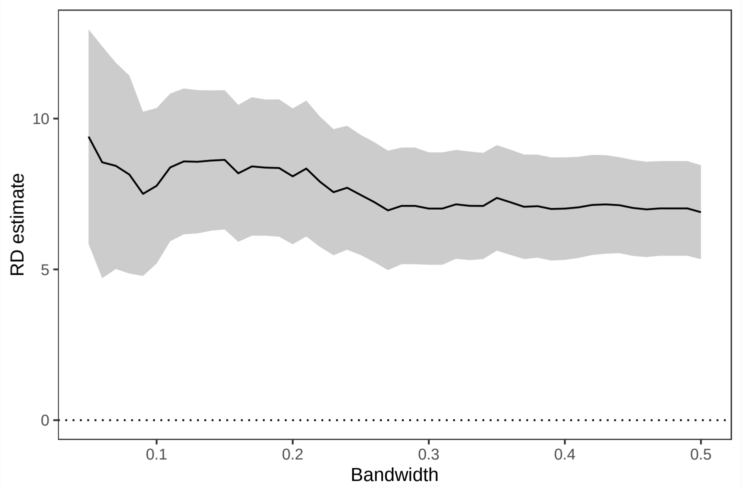
Assessing sensitivity

Results can change based on the data you include, so it's common to re-estimate the model using different bandwidths.

```
# create matrix w/ grid of bandwidths
sens = data.frame(bw = seq(.05, .5, .01), est = NaN, se=NaN)

# re-estimate linear RD model for each bandwidth
for (i in 1:nrow(sens)){
  mod = lm(Y~D*x.tilde, dat[abs(dat$x.tilde) < sens$bw[i], ])
  se = sqrt(vcovHC(mod, "HC2")[2,2])
  sens$est[i] = mod$coefficients["D"]
  sens$se[i] = se
}
```

```r
ggplot(sens, aes(x = bw, y = est, ymin = est-2*se, ymax = est+2*se)) +
  geom_hline(yintercept = 0, lty=3) +
  geom_ribbon(fill = "grey80") +
  geom_line() +
  labs(x = "Bandwidth", y = "RD estimate")
```



Roadmap

Fuzzy RDD Setup

- Fuzzy RDD is similar to instrumental variable encouragement treatments
- At the cutoff probability of getting treated jumps, but it is not 1.
- Z is the binary encouragement indicator that captures **only** whether units are above or below the threshold c: $Z = 1\{X > c\}$
- D_i remains the binary treatment indicator
- Observed treatment is:

$$D = Z \cdot D_1 + (1 - Z) \cdot D_0; \text{ so } D_i = \begin{cases} D_{1i} & \text{if } Z_i = 1 \\ D_{0i} & \text{if } Z_i = 0 \end{cases}$$

IV Assumptions

We keep the same terminology as in case of instrumental variables.
Who are compliers, always-takers, never-takers, defiers?

Estimation

$$\begin{aligned}\alpha_{FRDD} &= \mathbb{E}[Y_{1i} - Y_{0i} | X = c \text{ and i is a complier}] \\ &= \frac{\text{outcome discontinuity}}{\text{treatment discontinuity}} \\ &= \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}\end{aligned}$$

Estimation

$$\begin{aligned}\alpha_{FRDD} &= \mathbb{E}[Y_{1i} - Y_{0i} | X = c \text{ and i is a complier}] \\ &= \frac{\text{outcome discontinuity}}{\text{treatment discontinuity}} \\ &= \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}\end{aligned}$$

You can use two-stage least squares regression to estimate the treatment effect:

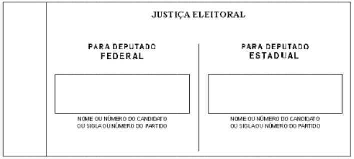
$$Y = \beta_0 + \beta_1 \tilde{X} + \beta_2 (Z \cdot \tilde{X}) + \alpha D + \epsilon$$

where D is instrumented with Z.

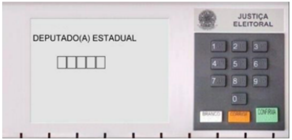
Roadmap

Electronic Voting in Brazil: Fujiwara (2015) ECTA

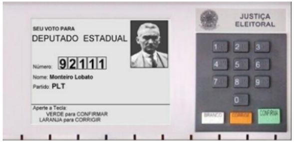
- Franchise expansions have substantial political consequences, but are hard to study
- De-facto franchise extension for the poor in Brazil (and many other developing countries) in the 1990s when voting was made electronic
  - Before: Voters expected to write name and five-digit code for preferred candidate
  - After: Voters enter a number, see picture and party ID, confirm vote
- **Assignment**: municipalities with population $\geq 40,500$ got EVMs in 1998



Paper ballot



Initial screen of the voting technology



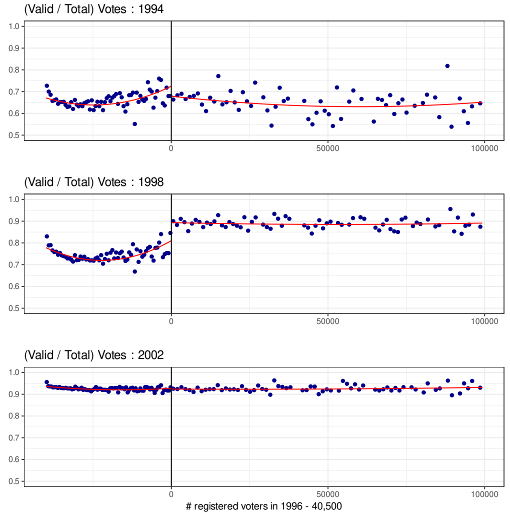
Voting for (fictional) candidate number 92111 (name: Monteiro Lobato, party: PLT)

Fujiwara (2015) ‘First Stage’

TABLE II

TREATMENT EFFECTS OF ELECTRONIC VOTING[a]

	Full Sample Mean	Pre-Treat. Mean	IKBW {Obs.}	(1)	(2)	(3)
Panel A: Baseline Results						
Valid Votes/Turnout (1998 Election)	0.755 [0.087]	0.780 (0.013)	11,873 {265}	0.118 (0.015)	0.121 (0.016)	0.124 (0.025)
Turnout/Reg. Voters (1998 Election)	0.765 [0.091]	0.785 (0.011)	12,438 {283}	−0.005 (0.019)	0.013 (0.021)	0.007 (0.033)
Reg. Voters/Population (1998 Election)	0.748 [0.141]	0.737 (0.010)	15,956 {388}	−0.004 (0.027)	0.010 (0.034)	0.032 (0.044)
Panel B: Placebo Tests (Election Years Without Discontinuous Assignment)						
Valid Votes/Turnout (1994 Election)	0.653 [0.099]	0.697 (0.011)	17,111 {433}	−0.013 (0.019)	−0.008 (0.023)	0.006 (0.032)
Valid Votes/Turnout (2002 Election)	0.928 [0.026]	0.921 (0.002)	17,204 {437}	0.005 (0.005)	0.008 (0.006)	0.009 (0.010)
Panel C: Do Left-Wing Parties Benefit Disproportionately From Electronic Voting?						
Vote-Weighted Party Ideology (1998 Elec.)	5.397 [0.692]	5.162 (0.094)	20,000 {558}	−0.222 (0.100)	−0.250 (0.081)	−0.108 (0.170)
Bandwidth Specification				IKBW Linear	10,000 Linear	5000 Linear
N	5281			—	229	116



Roadmap

DAGworld

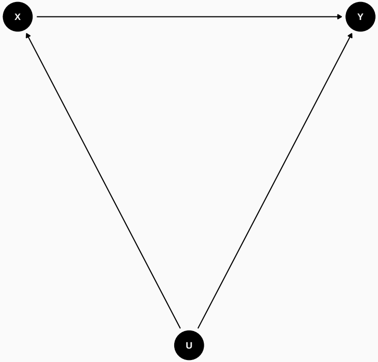
- http://dagitty.net/primer/
- https://donskerclass.github.io/CausalEconometrics.html Ch 4

Fork Structure

Chain Structure

Collider structure

Confounding of effect of X on Y by U



Realistic−ish Example

A: Mediator; B: Collider; C: Confounder