Section 9

Regression Discontinuity Designs

Sooahn Shin

GOV 2003

Nov 11, 2021

Section 9

Regression Discontinuity Designs

Sooahn Shin

GOV 2003

Nov 11, 2021

Overview

- Logistics:
  - **Pset 8 (the last pset) released!**: Due at 11:59 pm (ET) on Nov 17
  - **November ~~12th~~ 19th**: Submit a brief (no longer than 5 page) page memo of your main results, including tables, figures, and brief analysis. For methodological projects, this should include a description of the method and any analytical/simulation results. You will be required to give feedback on another group's project, which will be counted toward the overall grade based on attentiveness and usefulness of the feedback provided.

- Today's topics:
  - Sharp RD
    - Identification
    - Estimation
    - Diagnostics

Sharp RD

- Finding exogenous variation in the treatment assignment
  - RD: a **discontinuity** in treatment *assignment*

- Example: incumbency advantage in the U.S. House (Lee 2008)
  - "The overall causal impact of being the current incumbent party in a district on the votes obtained in the district's election"
  - Treatment (D_i): being the current incumbent party
  - Forcing (X_i): margin of victory (at election t)
  - Outcome (Y_i): probability of winning (at election $t+1$)
  - Sharp RD: $D_i = 1\{X_i \geq c\} \forall i$

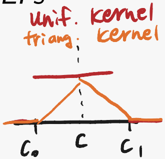
$$
\begin{tikzcd}
 & X_i \arrow[dl] \arrow[dr] & \\
D_i \arrow[rr] & & Y_i
\end{tikzcd}
$$

- Estimand: local average treatment effect **at the cutoff**

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

Sharp RD

- Identifying the effect at the cutoff with **continuity** of *CEFs*



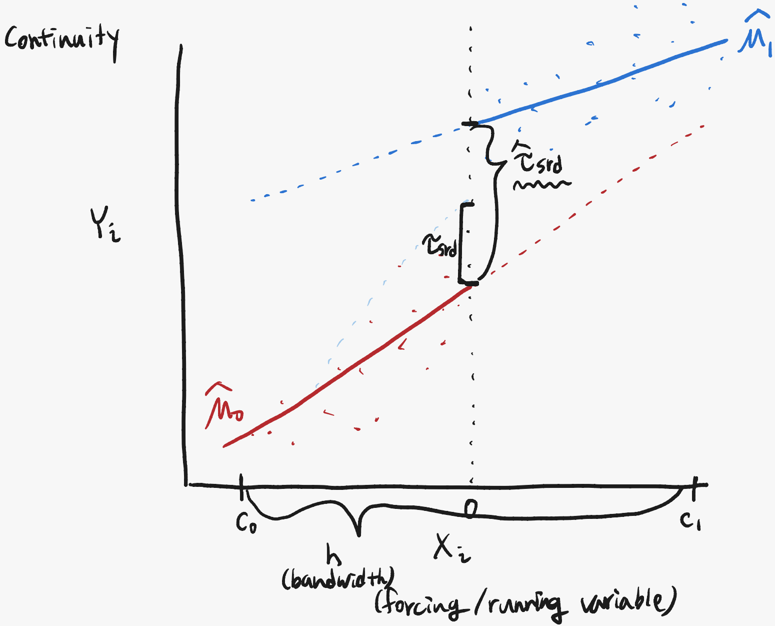
$$\tau_{\text{srd}} = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x]}_{=\mathbb{E}[Y_i(1)|X_i=c]} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}_{=\mathbb{E}[Y_i(0)|X_i=c]}$$

- Estimate the limit by the local linear regression

$$(\widehat{\alpha}_+, \widehat{\beta}_+) = \text{argmin} \sum_{i:X_i \geq c} \{Y_i - \alpha - \beta(X_i - c)\}^2 \underbrace{K\left(\frac{X_i - c}{h}\right)}_{\text{weights}}$$

- We take the estimated intercept: $\widehat{\alpha}_+ = \widehat{\mathbb{E}}[Y_i(1) \mid X_i = c]$
- Our point estimate is: $\widehat{\tau}_{\text{srd}} = \widehat{\alpha}_+ - \widehat{\alpha}_-$

Graphical Illustration



Identification: continuity v. local randomization

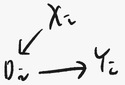
- Continuity assumption is not equivalent to local randomization

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

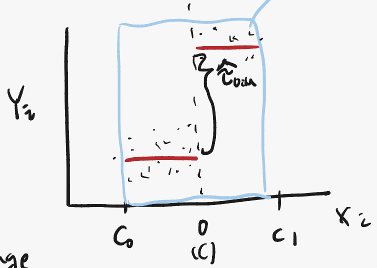
  - Stronger than continuity. Why?
  - Estimation and its visualization?

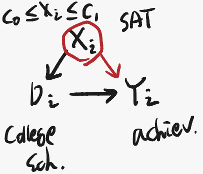
D_i

randomization "treatment"



Diff-in-means / local average





Estimation and visualization

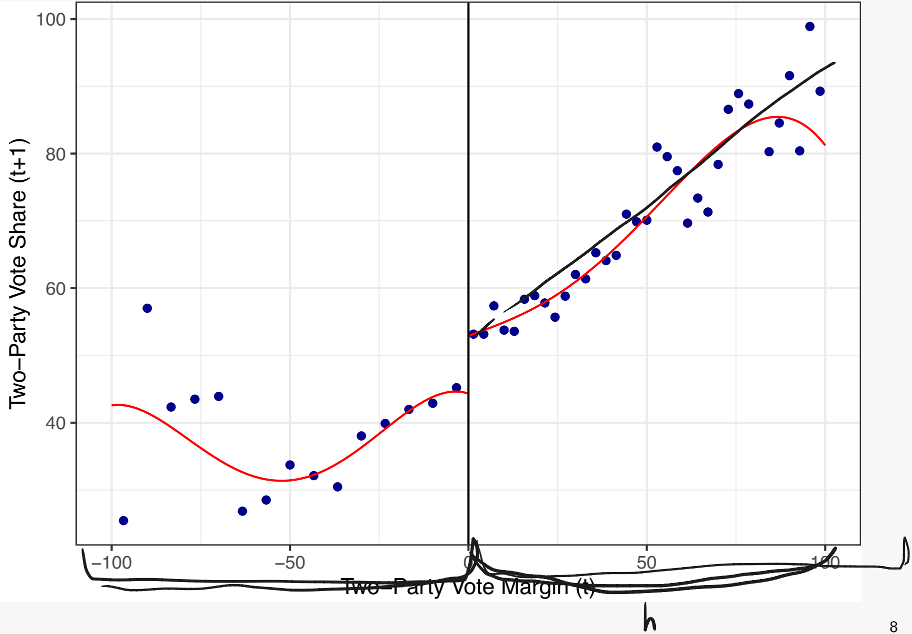
- Use `rdrobust` package (current standard)

1. Visualization: showing discontinuity at the cutoff

```r
# plot (data-driven Regression Discontinuity (RD) plots)
rdplot(y = vote, x = margin, kernel = "tri",
       title = "", y.label = "Two-Party Vote Share (t+1)",
       x.label = "Two-Party Vote Margin (t)")
```

RD Plot

p=3



100
80
60
40
Two–Party Vote Share (t+1)
−100
−50
0
50
100
Two–Party Vote Margin (t)
h

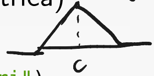
Estimation and visualization

2. Estimation: fit one linear regression with the interaction between $(X_i - c)$ and D_i

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

3. Optimal bandwidth, bias correction and robust standard errors

- Intuition:
    - find bandwidth that minimizes the estimation error MSE $E[E[\hat{\tau}_{srd}] - \tau]^2$
    - $\rightsquigarrow$ we don't know the true bias and have to estimate it
    - $\rightsquigarrow$ additional uncertainty
- Calonico, Cattaneo, and Titiunik (CCT, 2014, Econometrica) weights




```
# fit local linear regression
fit <- rdrobust(y = vote, x = margin, p = 1, kernel = "tri")
```

summary(fit)

fit$coef

linear

kernel = "uniform"

Results

	estimate	se
Conventional	7.414	1.459
Bias-Corrected	7.507	1.459
Robust	7.507	1.741 ← robust SE

bias-corrected

- Two types of point estimates:
  1. The standard local linear estimator $\widehat{\tau}_{\text{srd}}$
  2. The local linear estimator with bias-correction $\widehat{\tau}^{\text{rbc}}_{\text{srd}} = \widehat{\tau}_{\text{srd}} - \widehat{\text{bias}}$
- Two standard errors
  1. Standard SE $\widehat{\sigma}^2$
  2. "Robust" SE: accounts for uncertainty in bias estimation $\widehat{\sigma}^2_{\texttt{robust}}$
- We report the "Robust" estimate: $\widehat{\tau}_{\texttt{BC}}$ with $\widehat{\sigma}^2_{\texttt{robust}}$

Estimated effect along different bandwidths

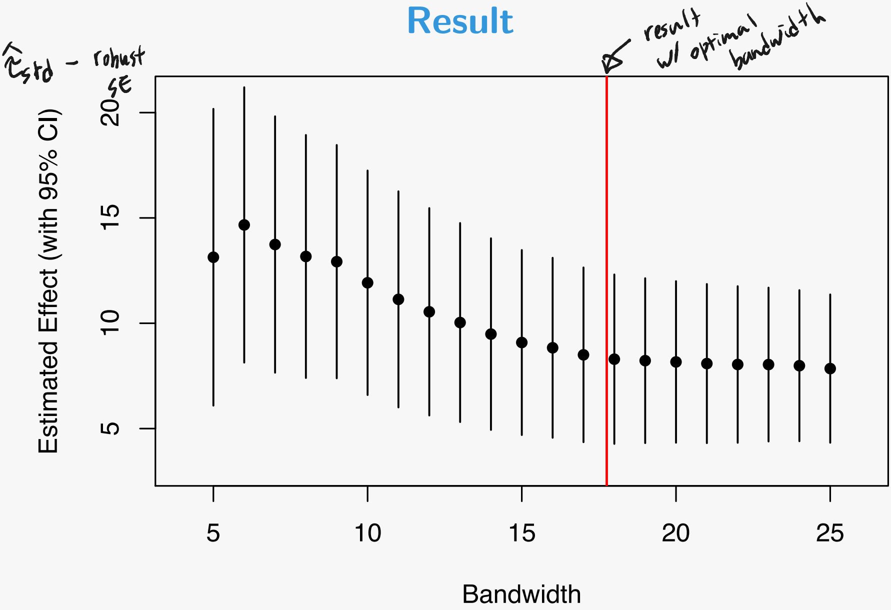
- We want to understand how results change along bandwidth

```r
# fit local linear regression with bandwidth bws[b]
bws <- seq(5, 25, by = 1); fits <- list()
for (b in 1:length(bws)) {
  fits[[b]] <- rdrobust(y = vote, x = margin, h = bws[b],
    p = 1, kernel = "tri")
}

# summarize result (use "robust")
plot(1, 1, type = 'n', xlim = c(4, 26), ylim  = c(3, 21),
     xlab = 'Bandwidth', ylab = 'Estimated Effect (with 95% CI)')
for (b in 1:length(bws)) {
  points(bws[b], fits[[b]]$coef[3], pch = 16)
  lines(c(bws[b], bws[b]), fits[[b]]$ci[3,], lwd = 1.2)
}
abline(v = fit$bws[1,1], col = 'red', lwd = 1.5)
```

optimal bandwidth

Result

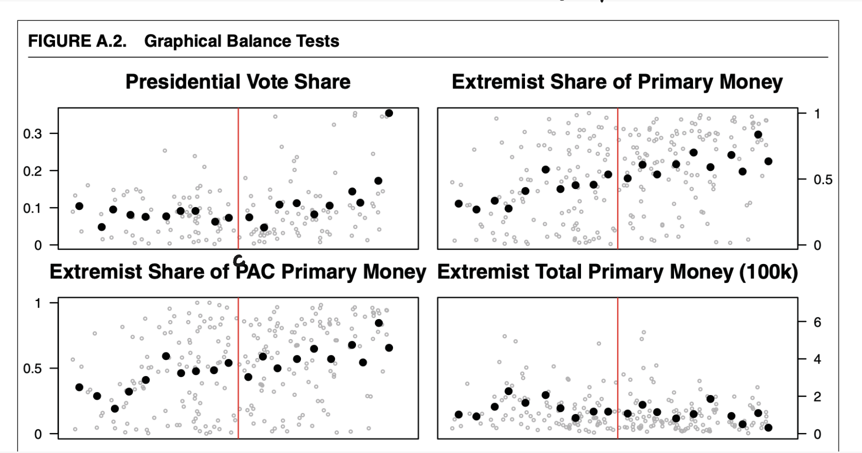


$\tilde{\tau}$ std - robust SE
result w/ optimal bandwidth
Estimated Effect (with 95% CI)
5
10
15
20
Bandwidth
5
10
15
20
25

Diagnostics: no-sorting?

- Using pre-treatment covariates

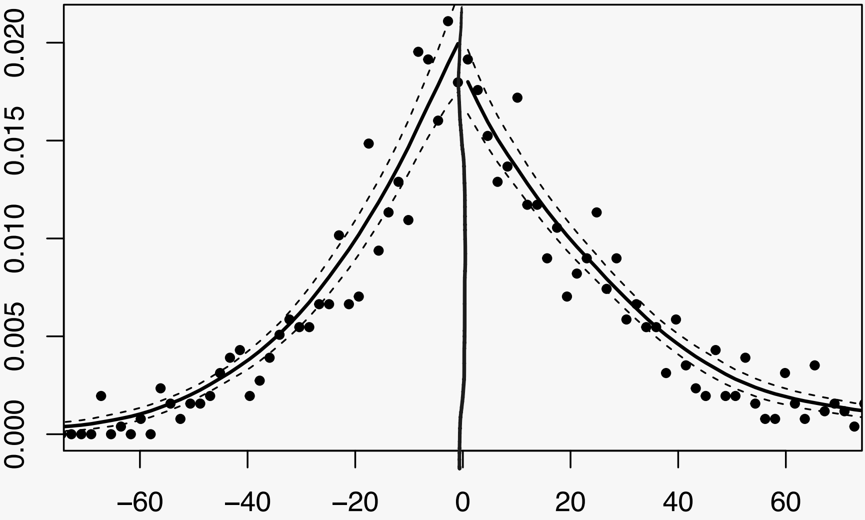
draw binned mean plot (rdplot()) using pre-treatment variables.



Diagnostics: no-sorting?

density

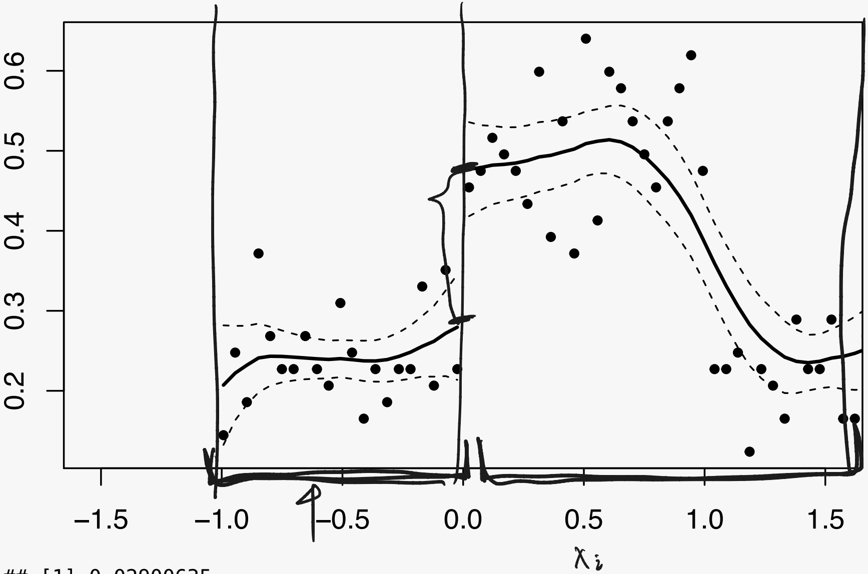
- McCrary test   density plot X_i ≠ rdplot Y_i / X_i




```
## [1] 0.3897849
```

Diagnostics: no-sorting?

- Example of discontinuity




```
## [1] 0.02900635
```