Statistical Modeling and Causal Inference with R

Week 7: Regression Discontinuity Designs

Manuel Bosancianu

Max Schaub

October 26, 2020

Hertie School of Governance

Introduction

RDD features

- ✓ a score / running variable / forcing variable / index
- ✓ a cutoff / threshold
- ✓ a treatment

$P_{assignment}$ changes discontinuously at the threshold.

Sharp RDD

The Fujiwara (2015) study is an example of *sharp* RDD:

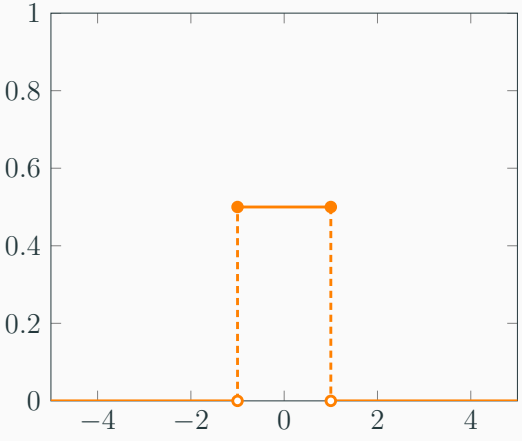
- ✓ all 307 municipalities above 40,500 registered voters used EV
- ✓ 4,967 out of 4,974 municipalities below 40,500 registered voters used paper ballots

$$\tau = \lim_{v_m \downarrow 40{,}500} E[Y_m | v_m] - \lim_{v_m \uparrow 40{,}500} E[Y_m | v_m] \tag{1}$$

Estimating RD effects

Uses linear regressions, without any weights ("rectangular" kernel).

Here, $c = 0$ and $h = 1$.



Estimating RD effects

Assume that:

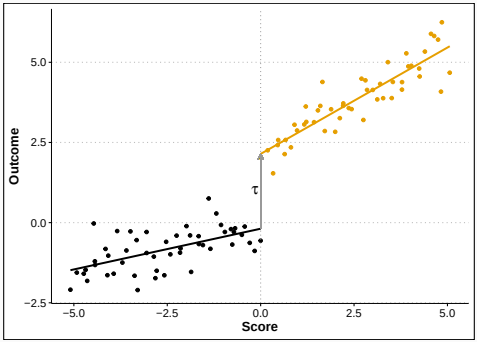
V_m are the registered voters in municipality m, and that

$$D = \begin{cases} 0, & \text{if } V_m < 40,500 \\ 1, & \text{if } V_m \geq 40,500 \end{cases} \tag{2}$$

$$Y_m = \beta_0 + \tau D_m + \beta_1 V_m + \beta_2 V_m D_m + \epsilon_m \tag{3}$$

What kind of regression characteristics is this assuming? What assumptions are implicit?

Linear model & different slope



- ✓ linearity: regressions are linear in V_m
- ✓ varying treatment effect (τ) along V

Results

Is there a jump at cutoff?

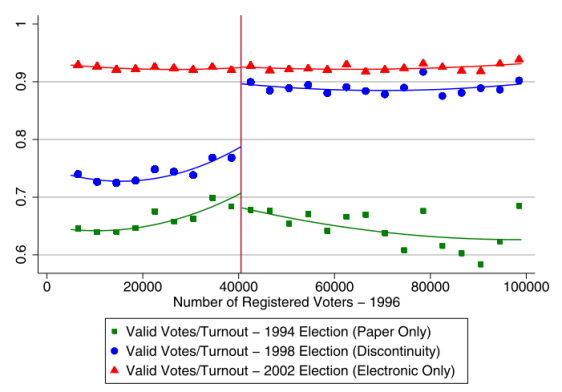


Figure 2 (p. 435)

Why are the other two election years provided here?

Effect of introducing EV

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	0.755	0.780	11,873	0.118	0.121	0.124
(1998 Election)	[0.087]	(0.013)	{265}	(0.015)	(0.016)	(0.025)
Turnout/Reg. Voters	0.765	0.785	12,438	−0.005	0.013	0.007
(1998 Election)	[0.091]	(0.011)	{283}	(0.019)	(0.021)	(0.033)
Reg. Voters/Population	0.748	0.737	15,956	−0.004	0.010	0.032
(1998 Election)	[0.141]	(0.010)	{388}	(0.027)	(0.034)	(0.044)

Table 2 (p. 436)

Placebo tests

Panel B: Placebo Tests (Election Years Without Discontinuous Assignment)						
Valid Votes/Turnout	0.653	0.697	17,111	−0.013	−0.008	0.006
(1994 Election)	[0.099]	(0.011)	{433}	(0.019)	(0.023)	(0.032)
Valid Votes/Turnout	0.928	0.921	17,204	0.005	0.008	0.009
(2002 Election)	[0.026]	(0.002)	{437}	(0.005)	(0.006)	(0.010)

Table 2 (p. 436)

Parties favored

TABLE III

TREATMENT EFFECTS OF ELECTRONIC VOTING, BY ILLITERACY RATE[a]

	Pre-Treat. Mean	IKBW {Obs.}	(1)	(2)	(3)	(4)
Panel A: Municipalities With Above-Median Illiteracy						
Valid Votes/Turnout	0.759	11,873	0.147	0.150	0.152	0.176
	(0.017)		(0.019)	(0.015)	(0.020)	(0.031)
N	—	—	116	279	103	49
Panel B: Municipalities With Below-Median Illiteracy						
Valid Votes/Turnout	0.799	11,873	0.092	0.113	0.096	0.089
	(0.018)		(0.020)	(0.016)	(0.022)	(0.032)
N	—	—	149	279	126	67
Test of Equality in TEs (*p*-Value)	—	—	0.049	0.090	0.056	0.054
Bandwidth	—	—	IKBW	20,000	10,000	5000

Table 3 (p. 439)

Validity checks

Falsification and validity

What were the 5 types of falsification and validity tests?

1. null effect on pre-treatment covariates and placebo outcomes
2. score density continuity around cutoff
3. treatment effect at artificial cutoff values
4. excluding observations near cutoff
5. sensitivity to bandwidth choices

Bandwidth sensitivity

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	0.755	0.780	11,873	0.118	0.121	0.124
(1998 Election)	[0.087]	(0.013)	{265}	(0.015)	(0.016)	(0.025)
Turnout/Reg. Voters	0.765	0.785	12,438	−0.005	0.013	0.007
(1998 Election)	[0.091]	(0.011)	{283}	(0.019)	(0.021)	(0.033)
Reg. Voters/Population	0.748	0.737	15,956	−0.004	0.010	0.032
(1998 Election)	[0.141]	(0.010)	{388}	(0.027)	(0.034)	(0.044)
Bandwidth				IKBW	10,000	5000
Specification				Linear	Linear	Linear
N	5281			—	229	116

Table 2 (p. 436)

Null effect on pre-treatment covariates

TABLE I

SUMMARY STATISTICS AND COVARIATE SMOOTHNESS (1991 CENSUS)[a]

	Full Sample Mean [Std. Dev.]	Pre-Treat. Mean	IKBW {Obs.}	(1)	(2)	(3)
Monthly Income (1991 *reais*)	123.13 [73.10]	174.83 (8.102)	20,000 {558}	0.908 (16.292)	6.096 (22.097)	14.017 (32.863)
Gini Index (Income)	0.559 [0.058]	0.575 (0.007)	15,596 {377}	0.005 (0.010)	0.002 (0.013)	−0.005 (0.017)
Latitude (Degrees)	−16.53 [8.23]	−16.40 (1.078)	16,547 {412}	0.174 (1.69)	0.361 (2.070)	−0.674 (2.998)
Longitude (Degrees)	46.36 [6.319]	45.18 (0.850)	14,531 {345}	0.419 (1.421)	0.550 (1.636)	2.685 (2.466)
Illiteracy Rate	0.360 [0.183]	0.274 (0.020)	16,068 {389}	−0.012 (0.020)	−0.076 (0.046)	−0.041 (0.065)
Share w/o 4 Years of Schooling	0.607 [0.179]	0.483 (0.020)	15,415 {372}	0.0006 (0.035)	−0.026 (0.041)	−0.041 (0.065)
Share w/o 8 Years of Schooling	0.876 [0.077]	0.788 (0.008)	20,000 {558}	−0.009 (0.015)	−0.017 (0.020)	−0.030 (0.032)
Population—1991 (Thousands)	24.80 [153.69]	58.35 (0.583)	20,000 {558}	0.653 (1.456)	1.066 (1.716)	0.962 (1.880)
Population—2000 (Thousands)	28.73 [170.91]	69.79 (1.257)	17,668 {454}	1.619 (3.043)	2.639 (3.937)	7.059 (5.011)
Share of Urban Population	0.507 [0.258]	0.237 (0.021)	20,000 {558}	0.004 (0.034)	−0.015 (0.048)	−0.069 (0.073)
Bandwidth	—	—	—	IKBW	10,000	5000
Observations	5281	—	—	—	229	116

Table 1 (p. 434)

Score density around cutoff

Danger here is that we may be dealing with sorting.

Discussed in text as implausible—why?

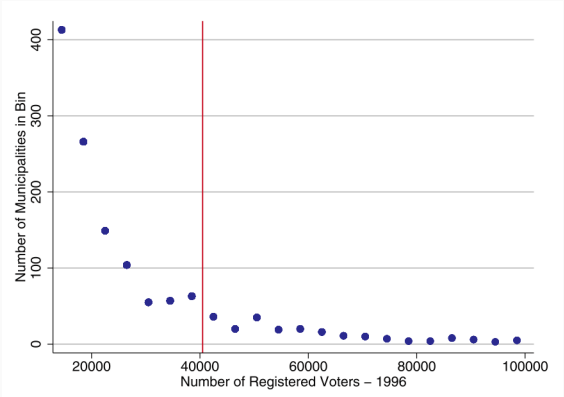


Figure A2

Score density: testing

Number of cases in bins

Registered voters	N
28,500-32,500	55
32,500-36,500	57
36,500-40,500	63
40,500-44,500	36
44,500-48,500	20
48,500-52,500	35

```
binom.test(36, 99, p = 0.5)
```

Score density: testing

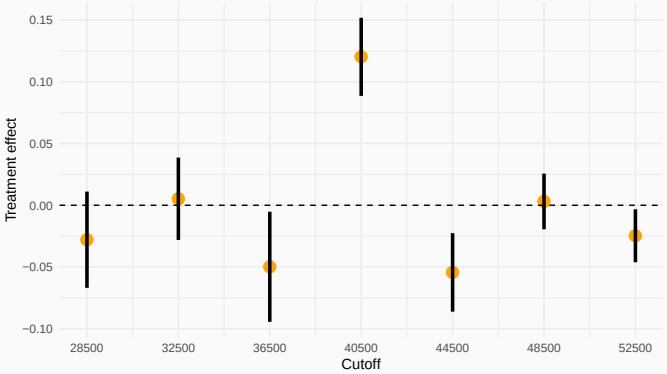
```
^^IExact binomial test

data:  36 and 99
number of successes = 36, number of trials = 99, p-value = 0.008634
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2692701 0.4663956
sample estimates:
probability of success
             0.3636364
```

Argument about timing of EV announcement cutoff is more convincing.

Artificial cutoffs

What was the logic here?

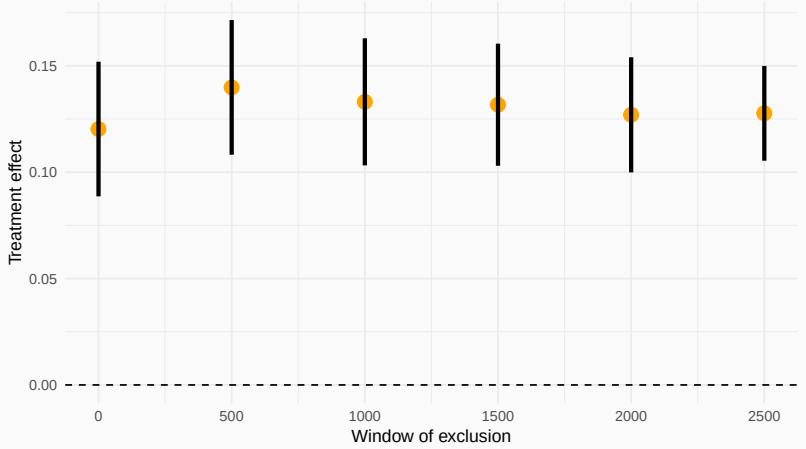


“Doughnut hole” test

What was the logic here?

I re-ran the model with gradually eliminating municipalities in bins of 500 registered voters on either side of cutoff.

“Doughnut hole” test



Thank you for the kind attention!

Fujiwara, T. (2015). Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil. *Econometrica*, *83*(2), 423–464.