Statistical Modeling and Causal Inference with R

Week 7: Regression Discontinuity Designs

Manuel Bosancianu

Max Schaub

October 26, 2020

Hertie School of Governance

Recap

Matching: “plugging in” the missing potential outcome for each treatment unit by using a control group unit.

How to choose: a control unit that is “closest” to the treatment unit on a set of covariates, X.

Two broad types of matching:

- ✓ exact: looking for identical values on X;
- ✓ approximate: looking for closest value on X.

Recap

Sophisticated implementations: propensity scores and coarsened exact matching.

One major problem is that matching can only address imbalance on observables.

If there is serious suspicion that unobserved covariates are responsible for imbalances, matching can't help.

Nothing beats randomization when looking to solve this kind of problem...

Today's focus

- ✓ Core features of the RDD
- ✓ Sharp RDD
- ✓ Estimation approaches in Sharp RDD
- ✓ Diagnostics of an RDD
- ✓ Example: female empowerment and its political determinants
- ✓ Fuzzy RDD

Introduction

RDD: core features

Introduced in educational studies by Thistlethwaite and Campbell (1960): rife with examples of cutoffs, tests, and selection procedures.

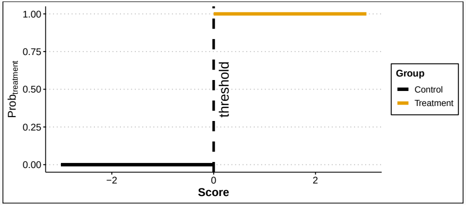
In economics, analyses only appear at the end of 90s (still dealing with education).

Setup is simple:

- ✓ a score (running/forcing variable, index): a variable which ranks units;
- ✓ a cutoff: a threshold for the score, which separates units in the treatment and control group;
- ✓ a treatment: a particular intervention.

RDD: core features

Defining feature: probability of treatment assignment as function of score changes discontinuously at cutoff.



A sharp RD design

Selection into treatment is non-random, but around the threshold we might benefit from *local randomization*.

RDD: benefits

- ✓ analysis relatively accessible to audience
- ✓ assumptions related to score, treatment and cutoff can usually be empirically tested
- ✓ extensive array of falsification tests and validity checks
- ✓ flexibility: time, geography, multiple scores, multiple cutoffs

Empirics I

Meyersson (2014) data

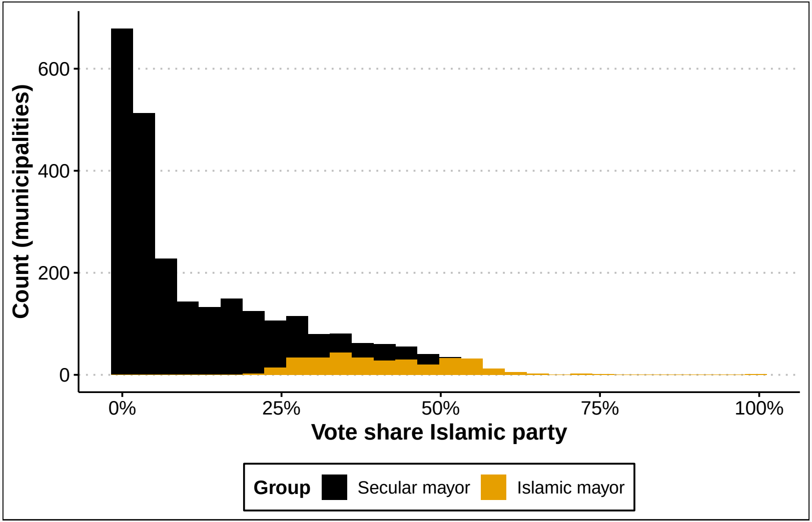
What effect does political control by Islamist parties have on the political empowerment of women?

1994 municipal elections in Turkey: 2 Islamic parties (*Refah* and *Büyük Birlik Partisi*) win the mayorship of 329 municipalities.

Outcomes targeted:

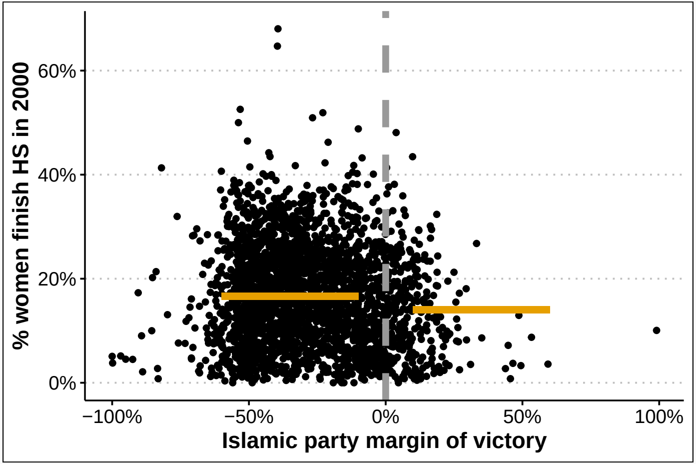
- ✓ educational enrollment of women;
- ✓ adolescent marriages;
- ✓ political participation of women.

Meyersson (2014) data



Distribution of municipalities

Women's empowerment



Association: female education and political power

Women's empowerment

HS completion below cutoff = 0.166152.

HS completion above cutoff = 0.140373.

Many ways in which municipalities where Islamic parties have margin above, say, 5% are different than the rest.

However, in a short window around the 0% cutoff, we might benefit from local randomization.

Sharp RDD

Sharp RDD

Two broad design types:

- ✓ sharp: treatment assigned = treatment received
- ✓ fuzzy: compliance with treatment assignment is imperfect

We use experimental language, but "treatment" is defined *ex post* by researcher.

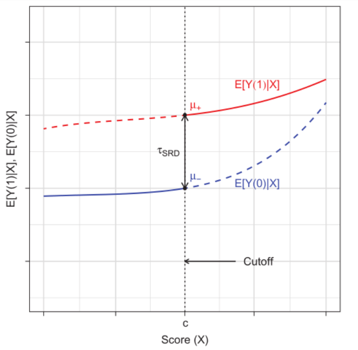
Observed outcomes (c is cutoff):

$$Y_i = (1 - T_i) * Y_{0i} + T_i * Y_{1i} = \begin{cases} Y_{0i}, & \textit{if } X_i < c \\ Y_{1i}, & \textit{if } X_i \geq c \end{cases} \tag{1}$$

Sharp RDD

Observed expected outcome is:

$$E[Y_i|X_i] = \begin{cases} E[Y_{0i}|X_i], & \textit{if } X_i < c \\ E[Y_{1i}|X_i], & \textit{if } X_i \geq c \end{cases} \quad (2)$$



Treatment effect in sharp RDD (Cattaneo et al., 2019)

RDD relies on extrapolation toward cutoff, to be able to compute τ_{SRD}.

Sharp RDD

We can't compute $E[Y_{1i}|X_i = x] - E[Y_{0i}|X_i = x]$ for almost any value of X.

The cutoff c is the only exception to this (we "almost" observe both lines).

$$\tau_{SRD} = E[Y_{1i}|X_i = c] - E[Y_{0i}|X_i = c] \equiv \mu_+ - \mu_- \tag{3}$$

Treatment is local in nature (LATE)!

Continuity assumption

A function $f(x)$ is continuous at the point $x = a$ if $f(x)$ and $f(a)$ get closer to each other as x gets closer to a.

If $E[Y_{1i}|X_i = x]$ and $E[Y_{0i}|X_i = x]$ are continuous at $x = c$, then:

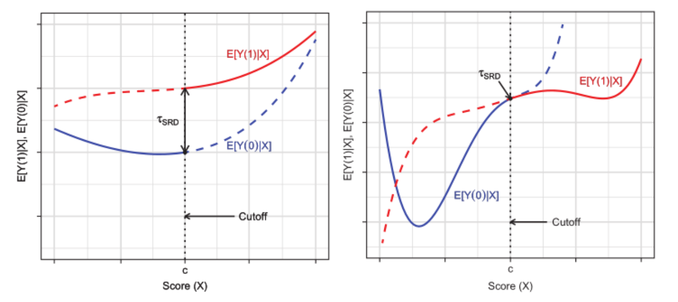
$$E[Y_{1i} - Y_{0i}|X_i = c] = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x] \tag{4}$$

Conditioning on X_i is impossible (no common support), but extrapolation allows us to compensate for this at $X_i = c$.

Local nature of effects

We compute the effect, τ_{SRD}, at a single point: c.

When treatment effect varies as a function of score X, τ_{SRD} not informative outside of c.



Treatment effect heterogeneity (Cattaneo et al., 2019)

Estimation

Estimating RD effects

Two strategies:

- ✓ continuity-based: using local polynomial methods to approximate $E[Y_i|X_i = x]$ on each side of cutoff;
- ✓ randomization-based: using tools from analysis of experiments in the area around the cutoff.

First set of tools is flexible and easy to implement, but not always justifiable.

Local polynomial approach

Usually, no observations where $X_i = c \Rightarrow$ makes extrapolation necessary.

Fundamentally, involves approximating these two regression functions: $E[Y_{0i}|X_i = x]$ and $E[Y_{1i}|X_i = x]$.

Using all the data for this produces a poor approximation at the boundary (Runge's phenomenon) $\Rightarrow$ use only observations close to cutoff.

Stages in approach

Estimation only uses observations between $c - h$ and $c + h$, where $h > 0$ is the *bandwidth*.

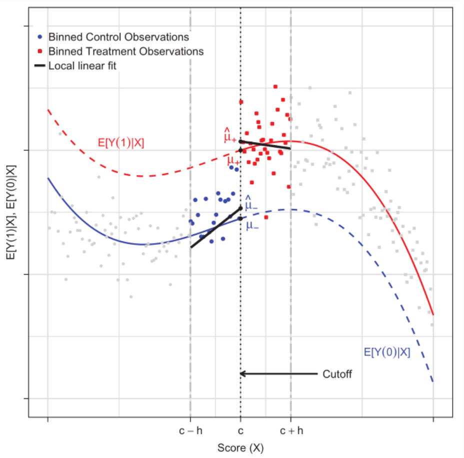
Cases are weighted as a function of distance to c. Those closer to c get higher weight.

- ✓ choose polynomial order p, and kernel function: $K(\cdot)$ (for weights);
- ✓ choose bandwidth h;
- ✓ fit WLS regression with p polynomial terms in the $[c, c+h]$ region and keep intercept $(\hat{\mu}_+)$;
- ✓ fit WLS regression with p polynomial terms in the $[c-h, c]$ region and keep intercept $(\hat{\mu}_-)$;
- ✓ $\hat{\tau}_{SRD} = \hat{\mu}_+ - \hat{\mu}_+$.

Choices to be made

3 choices:

- ✓ polynomial order: p
- ✓ bandwidth: h
- ✓ kernel function: $K(\cdot)$

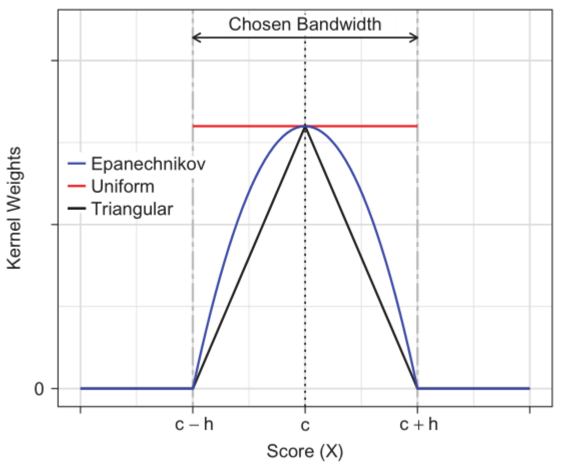


(Cattaneo et al., 2019)

Kernel function

Triangular is typically default.

In practice, does not make a big difference.



(Cattaneo et al., 2019)

Polynomial order

First order: $Y_i = \beta_0 + \beta_1 X + \epsilon_i$.

Second order: $Y_i = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon_i$.

The default tends to be the linear RD estimator (first order).

Bandwidth selection

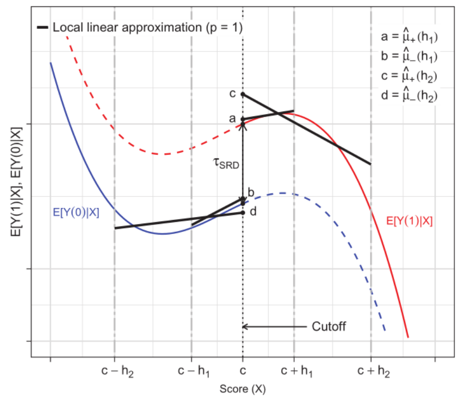
Very consequential for estimation.

Accuracy of approximation can be improved by reducing bandwidth.

The downside is that variance of estimator increases because fewer observations make it in.

“Bias-variance tradeoff” in bandwidth choice.

Bandwidth selection



(Cattaneo et al., 2019)

Software will handle this automatically.

Regression: linear model & common slope

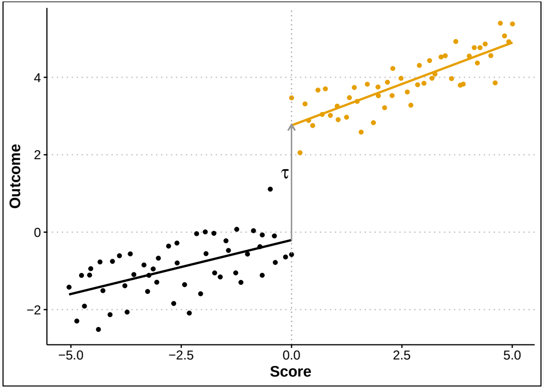
Assumptions:

- linearity: $E[Y_{0i}|X_i = x]$ and $E[Y_{1i}|X_i = x]$ are linear in x
- constant treatment effect (τ)

Implication:

$$\begin{cases} E[Y_{0i}|X_i] = \beta_0 + \beta_1 * X_i \\ E[Y_{1i}|X_i] = \beta_0 + \tau + \beta_1 * X_i \end{cases} \tag{5}$$

Regression: linear model & common slope



Model is $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \epsilon_i$

Use transformed version of X: deviation from c.

Regression: linear model & different slope

Assumptions:

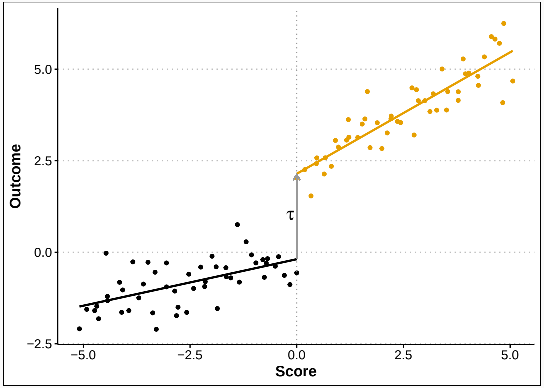
- ✓ linearity: $E[Y_{0i}|X_i = x]$ and $E[Y_{1i}|X_i = x]$ are linear in x
- ✓ varying treatment effect (τ) along X

Implication:

$$\begin{cases} E[Y_{0i}|X_i] = \beta_0 + \beta_1 * X_i \\ E[Y_{1i}|X_i] = \beta_0 + \tau + (\beta_1 + \phi) * X_i \end{cases} \tag{6}$$

ϕ can be either positive or negative.

Regression: linear model & different slope



Model is $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \phi D_i X_i + \epsilon_i$

Use transformed version of X: deviation from c.

Regression: nonlinear model

Assumptions:

- ✓ non-linearity: $E[Y_{0i}|X_i = x]$ and $E[Y_{1i}|X_i = x]$ are non-linear in x
- ✓ varying treatment effect (τ) along X

Include quadratic version of score, X_i^2, and interaction with D_i, but venture further very carefully (Gelman & Imbens, 2019).

Model: $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i D_i + \beta_4 X_i^2 D_i + \epsilon_i$

Empirics II

Meyersson (2014) data

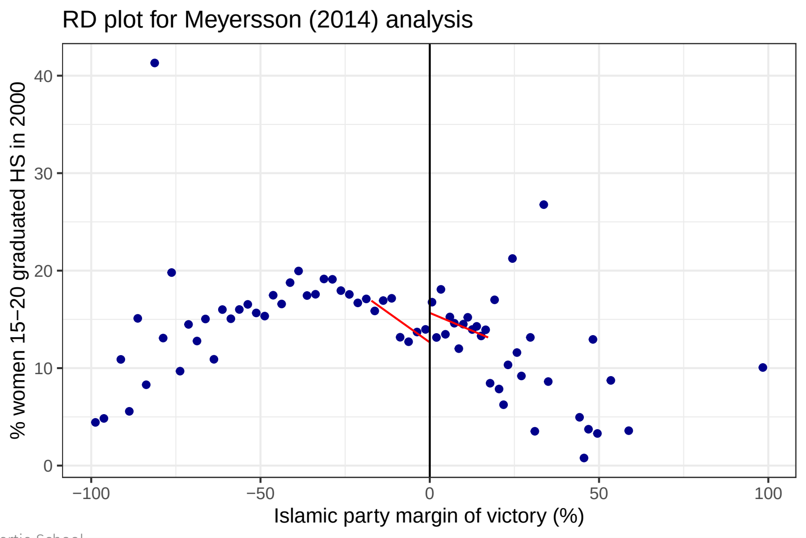
What effect does political control by Islamist parties have on the political empowerment of women?

```r
df_rdd <- df_rdd %>%
  mutate(iwm94 = iwm94 * 100,
         hischshr1520f = hischshr1520f * 100)

out <- rdrobust(df_rdd$hischshr1520f,
                df_rdd$iwm94,
                kernel = "triangular",
                p = 1,
                bwselect = "mserd")
```

Effect of Islamic party power

```
summary(out)


Sharp RD estimates using local polynomial regression.

Number of Obs.                 2630
BW type                       mserd
Kernel                   Triangular
VCE method                       NN

Number of Obs.                 2315             315
Eff. Number of Obs.             529             266
Order est. (p)                    1               1
Order bias  (q)                   2               2
BW est. (h)                  17.243          17.243
BW bias (b)                  28.575          28.575
rho (h/b)                     0.603           0.603
Unique Obs.                    2313             315

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     3.020     1.427     2.116     0.034     [0.223 , 5.816]
        Robust         -         -     1.776     0.076    [-0.309 , 6.276]
=============================================================================
```

Effect of Islamic party power



Alternative specifications

TABLE II

ISLAMIC RULE AND HIGH SCHOOL EDUCATION[a]

Outcome	Completed High School in 2000								Enrollment
Age Cohort	15–20								15–30
Control Function	None		Linear				Quadratic	Cubic	Linear
Bandwidth	Global		$\hat{h}$		$\hat{h}/2$	$2\hat{h}$	$\hat{h}$	$\hat{h}$	$\hat{h}$
Covariates	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Women									
Outcome mean	0.163	0.163	0.152	0.152	0.144	0.166	0.152	0.152	0.127
Islamic mayor in 1994	−0.026***	0.012**	0.032***	0.028***	0.032***	0.022***	0.028***	0.043***	0.014***
	(0.006)	(0.006)	(0.010)	(0.007)	(0.011)	(0.006)	(0.011)	(0.016)	(0.005)
Bandwidth	1.000	1.000	0.240	0.240	0.120	0.480	0.240	0.240	0.205
R^2	0.01	0.55	0.03	0.65	0.65	0.58	0.65	0.65	0.48
Observations	2629	2629	1020	1020	589	2049	1020	1020	904

Diagnostics and Falsification Checks

Diagnosing an RD design

Treatment assignment mechanism is known to researcher, and based on observable features.

A whole array of approaches!

- ✓ null effect on pre-treatment covariates and placebo outcomes
- ✓ score density continuity around cutoff
- ✓ treatment effect at artificial cutoff values
- ✓ excluding observations near cutoff
- ✓ sensitivity to bandwidth choices

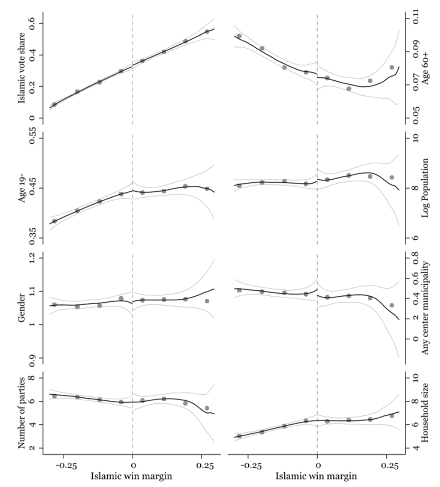
Pre-treatment covariates and placebos

Are treatment and control units similar around the cutoff on observables?

Two types:

- ✓ pre-treatment covariates: determined before assignment to treatment;
- ✓ placebo outcomes: post-treatment, but not affected by treatment.

Pre-treatment covariates: Meyersson (2014)



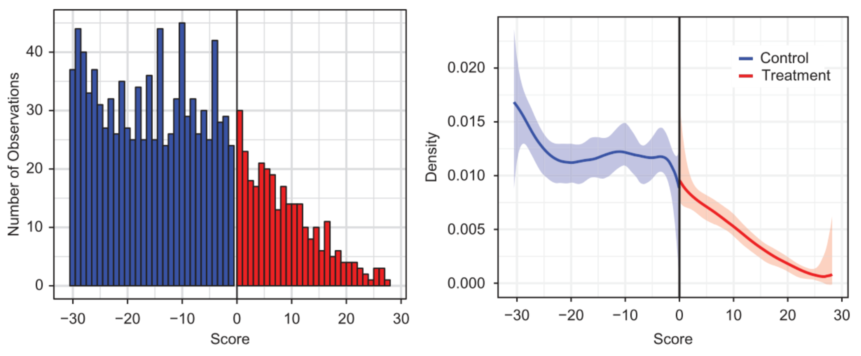
Score density

Is number of observations different below and above the cutoff? (if local randomization holds, it shouldn't be)

Could indicate active manipulation of score (e.g. contesting test results below passing threshold).

Easily done with a density test, to test for sorting.

Score density: Meyersson (2014)



(Cattaneo et al., 2019)

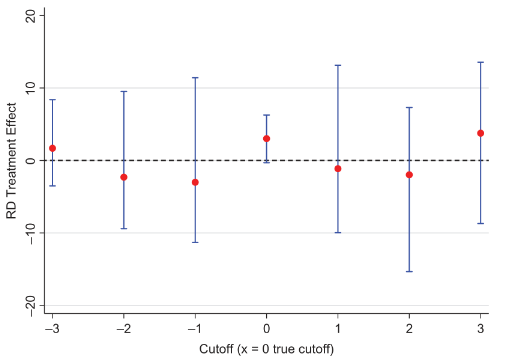
Artificial cutoff values

Key identifying assumption: continuity of regression function at cutoff in the absence of treatment.

Impossible to test at cutoff, but the opposite can be tested outside of it.

Are there discontinuities in regression functions away from cutoff which can't be explained?

Artificial cutoff values: Meyersson (2014)



(Cattaneo et al., 2019)

No evidence of discontinuous jump at artificial cutoffs.

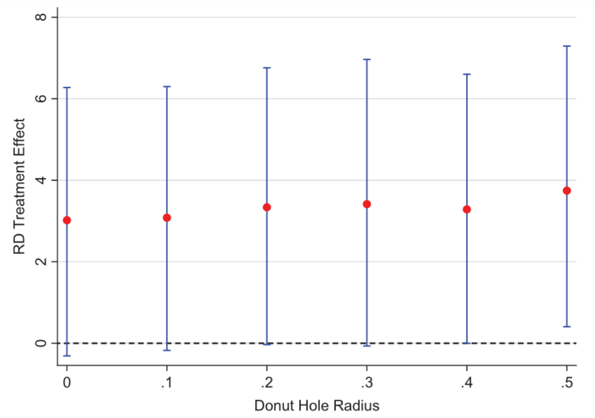
Sensitivity to cases near cutoff

Are results sensitive to excluding cases near the cutoff?

If any score manipulation took place, these would be the most likely units to engage in this.

Gradually remove observations in a window around the cutoff, $[c - w; c + w]$, and re-run analysis.

Sensitivity to cases near cutoff: Meyersson (2014)



(Cattaneo et al., 2019)

Sensitivity to bandwidth choice: Meyersson (2014)

TABLE III
ALTERNATIVE RD SPECIFICATIONS[a]

	Bandwidth				
	1	0.5	0.25	0.1	0.05
	(1)	(2)	(3)	(4)	(5)
			Panel A: Women		
Polynomial order of control function					
None	0.012**	0.015**	0.018***	0.025***	0.018*
	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)
Linear	0.014**	0.021***	0.025***	0.028**	0.039**
	(0.007)	(0.006)	(0.007)	(0.012)	(0.019)
Quadratic	0.027***	0.030***	0.033***	0.032*	0.051
	(0.007)	(0.007)	(0.010)	(0.018)	(0.032)
Cubic	0.031***	0.026***	0.036**	0.057**	0.054
	(0.007)	(0.010)	(0.015)	(0.028)	(0.042)
Quartic	0.030***	0.032**	0.044**	0.067**	0.028
	(0.009)	(0.012)	(0.017)	(0.033)	(0.056)
Observations	2628	2177	1049	489	257

In this instance, results are largely insensitive to bandwidth.

Fuzzy RD

Fuzzy RD: features

In sharp RD we assumed:

- ✓ all units assigned to treatment actually take it
- ✓ no units assigned to control take treatment

In fuzzy RD, those assumptions are no longer met:

- ✓ some units assigned to treatment fail to receive it
- ✓ some control units manage to get treatment

Fuzzy RD: features

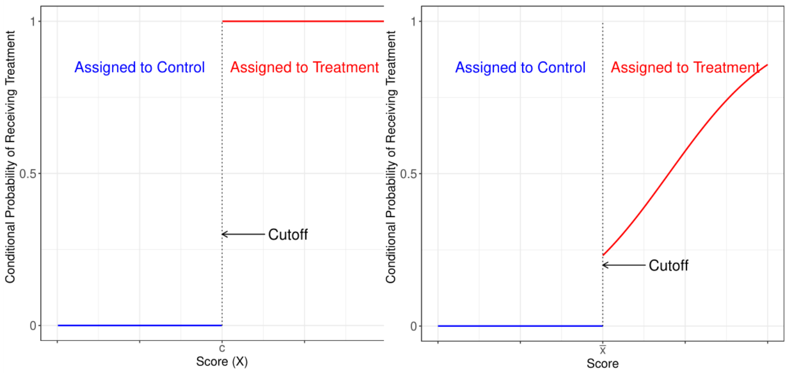
Probability of *receiving* treatment still jumps at cutoff, but is no longer either 0 or 1.

T_i: treatment assignment.

D_i: treatment take-up.

For some units $T_i \neq D_i$.

Fuzzy RD: features



(a) Sharp RD (b) Fuzzy RD (One-Sided)

Treatment taken: $D_i = T_i D_{1i} + (1 - T_i) D_{0i}$.

IV similarity

Similarity to IV: T_i only impacts Y_i through its effect on D_i (*exclusion restriction*).

Interest is in both effect of T_i (which is a standard sharp RD estimation), and of D_i.

Can be estimated using either the continuity-based approach or the local randomization one.

Estimation: continuity-based

Effect of T_i is clear, as compliance with the assignment rule is perfect (unlike take-up).

$$\tau_{ITT} = \lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x] =$$
$$= E[(D_{1i} - D_{0i})(Y_{1i} - Y_{0i}) | X_i = c]$$

$\tau_{SRD} \neq \tau_{ITT}$, since the latter also includes $D_{1i} - D_{0i}$.

With perfect compliance, $D_{1i} - D_{0i} = 1 - 0 = 1$ for all i.

Estimation: continuity-based

We can also define the first-stage effect: effect (at the cutoff) of being assigned to treatment on treatment take-up.

$$\tau_{FS} = \lim_{x \downarrow c} E[D_i | X_i = x] - \lim_{x \uparrow c} E[D_i | X_i = x] =$$
$$= E[D_{1i} - D_{0i} | X_i = c]$$

Both τ_{ITT} and τ_{FS} are sharp RD parameters, and can be estimated as we saw above.

Estimation: treatment effect

Additional assumption: monotonicity.

Unit i that refuses treatment at cutoff c_1 must refuse it for any cutoff $c_2 > c_1$. Similarly, treatment taken at cutoff c_1 should also be taken at cutoff $c_2 < c_1$.

It can be shown that under this condition (plus continuity):

$$LATE = \tau_{FRD} = \frac{\tau_{ITT}}{\tau_{FS}} = \frac{\lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x]}{\lim_{x \downarrow c} E[D_i|X_i = x] - \lim_{x \uparrow c} E[D_i|X_i = x]} \tag{7}$$

Estimation is performed with 2SLS.

Wrap-up

Validity

Though assignment is “as-if-random”, it’s not purposeful random assignment.

Internal validity is good: with enough data around the cutoff, we can estimate τ at cutoff.

External validity is less good: effect is local, at the cutoff. We cannot extrapolate to other values $X \neq c$ except by making additional assumptions.

However, a very flexible framework: score can be categorical, multiple scores and multiple cutoffs can be accommodated.

Summary

Though data-intensive, the RDD framework is very powerful.

It comes with a host of tools for model assessment and validation.

Can be adapted to a host of policy-relevant empirical settings: educational achievement, corruption, distributive politics, political accountability.

Thank you for the kind attention!

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. New York: Cambridge University Press.

Gelman, A., & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456.

Meyersson, E. (2014). Islamic Rule and the Empowerment of the Poor and Pious. *Econometrica*, *82*(1), 229–269.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*, *51*(6), 309–317.