Regression Discontinuity Design

PP346

Harris School of Public Policy
University of Chicago

Topic 5

Regression Discontinuity Design (RD or RDD)

Suppose you wanted to estimate the effect of attending Chicago Harris. How would you do it?

First thing that comes to mind is to compare applicants who were admitted to applicants who were not.

Is that a good idea?

Now let's refine that comparison. Assume that the HS admissions office:

- (i) summarizes everything we know about an applicant into a score, which we will denote by H, then
- (ii) ranks all applicants according to H, then
- (iii) draws a line separating the acceptances from the rejections.

What if we compared the the last applicant admitted to the first person denied? Would that be a better comparison?

What about the precision of the estimates? How could we improve it?

This is an example of a regression discontinuity design.

Regression discontinuities identify treatment effects by exploiting the tendency for people close to the margin to be similar, differing primarily by treatment status.

Treatment is assigned on the basis of one continuous variable called the assignment variable. It is also called the running or forcing variable. In the hypothetical admissions example, the running variable is H.

Everybody on one side of the cutoff is assigned to the treatment group; everybody on the other side is assigned to the comparison group.

The assignment variable could be any measure taken prior to treatment.

RDDs take two forms: sharp and fuzzy.

Sharp RDD

In a sharp RDD design, treatment probability goes from 0 to 100% at the cutoff.

Define X to be the assignment or running variable. We say that there is a sharp regression discontinuity if there exists a cutoff c such that

$$Pr(D = 1|X \geq c) = 1 \text{ and}$$
$$Pr(D = 1|X < c) = 0$$

so $Pr(D = 1|X \geq c) - Pr(D = 1|X < c) = 1$.

In a sharp RD, there is complete compliance with treatment. No one below the cutoff gets treatment, and everyone above takes it up.

This is a picture of a sharp design:

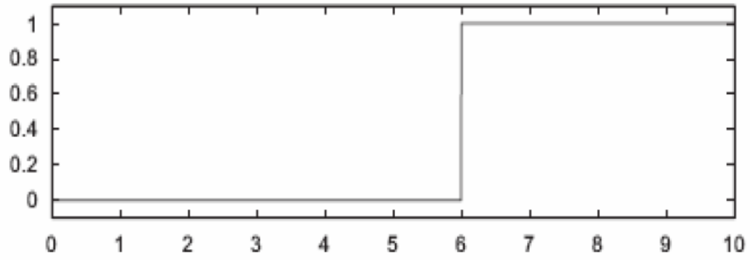
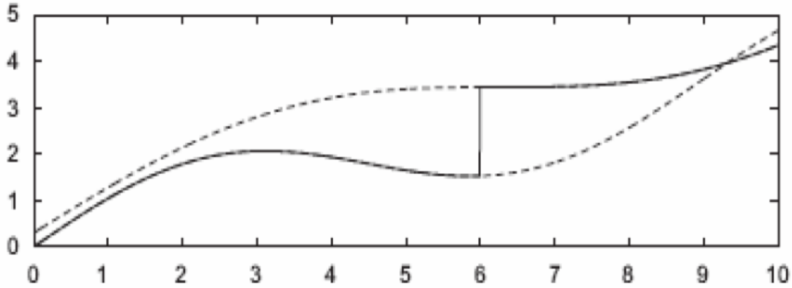


Fig. 1. Assignment probabilities (SRD).



Fuzzy RDD

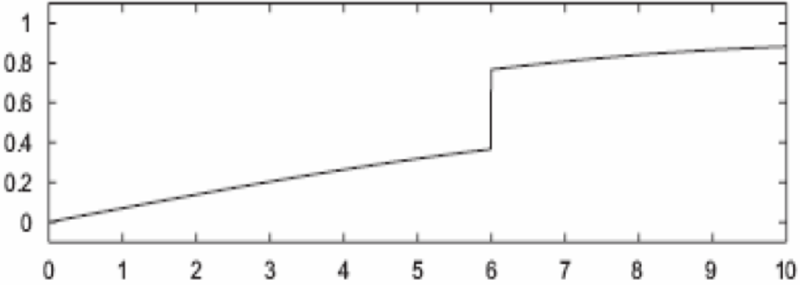
In a fuzzy RDD design, treatment probability changes discontinuously, but by less than 100% at the cutoff.

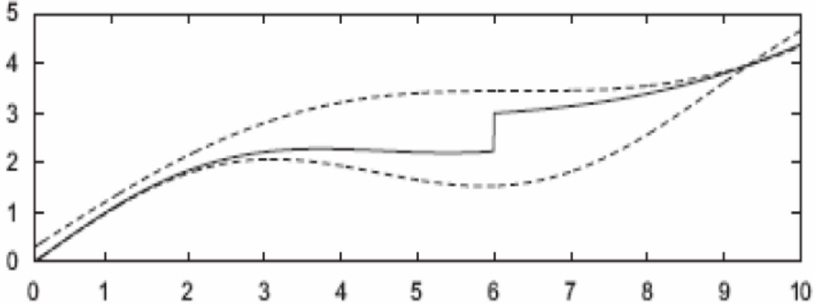
There is a fuzzy regression discontinuity if there exists a c such that

$$Pr(D = 1|X \geq c) - Pr(D = 1|X < c) = k$$
$$\text{where } 0 < k < 1$$

In a fuzzy RD, there is non-compliance, potentially in both directions. People below the cutoff may get treatment, whereas those above may not take it up.

This is a picture of a fuzzy design:





Deshpande (2016)

- General question is whether transfer programs inhibit labor market success
- Specific question is whether being dropped from SSI rolls increases earnings and income
- Study population is people who received SSI while children

Deshpande (2016) cont.

Discontinuity stems from policy change

- Kids who turned 18 before August 22, 1996 not subject to medical review
- Kids who turned 18 after that date required medical review in order to stay on SSI
- Result: many kids were dropped from rolls, since child and adult eligibility reqs differ

Is Deshpande's design sharp or fuzzy?

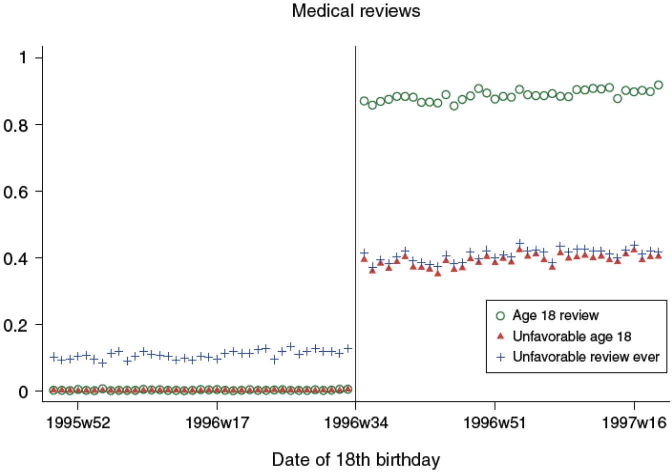


FIGURE 2. EMPIRICAL STRATEGY USING VARIATION IN ELIGIBILITY FOR MEDICAL REVIEWS

Notes: Figure plots proportion of SSI children in each birthweek bin who receive an age 18 medical review, receive an unfavorable age 18 medical review, and ever receive an unfavorable medical review (through 2013). Sample is SSI children with an 18th birthday within 37 weeks of the August 22, 1996 cutoff.

Intuition

In an RDD framework, people who are similar to each other end up being treated very differently.

This helps to estimate the effect of adult SSI on earnings and income.

Think of comparing the adult outcomes of two people who both received SSI as kids. One leaves the program, one remains on.

In general, have to be concerned about selection: the adult incomes of the two might have been different, even if they had both remained on the program.

With welfare reform, people who turned 18 on August 21, 1996 were much more likely to remain on SSI as adults than people who turned 18 on August 22.

If we focus on individuals "close" to the cutoff, they might not vary much in terms of potential outcomes. Differences in adult outcomes can be more convincingly attributed to leaving SSI.

Identifying assumption

This crucial assumption for the RDD to be valid is that the mean potential outcomes are continuous in the running variable X.

In the case of SSI receipt, there can be no discontinuous jumps in the potential earnings of people who turned 18 just before or after August 22, 1996.

Is this observable?

RDD Estimation

Sharp Design

We want to estimate the mean difference in outcomes for people on either side of the cutoff c:

$$\delta(c) = \lim_{x \uparrow c} E(Y_1|X = x) - \lim_{x \downarrow c} E(Y_0|X = x)$$

One can estimate the effect with a regression.

The regression looks like this:

$$Y = \beta_0 + \beta_1(X_1 - c) + \delta D + \beta_2(X_1 - c)D + \varepsilon$$

where X_1 is the running variable, and D is the treatment dummy.

How do you interpret these variables? The parameters?

Sometimes it's advisable to use a higher-order polynomial, such as a quadratic or cubic, in $X - c$.

You also have to choose the "caliper," that is, how far to extend the sample on either side of c.

How to think about the caliper?

The smaller the caliper, the more similar are the control and treatment groups. More convincing that you are isolating the effect of treatment. At the same time, the sample can get small.

A larger caliper gives you a bigger sample, but treatment and control groups are less similar (in the limit, you're just running OLS on your full sample).

The caliper involves a bias-variance tradeoff. Good practice is to experiment, report results for alternative values.

Adding controls

The OLS estimator can be augmented with additional covariate controls for other Xs.

- They can help reduce standard errors
- These should have little impact on estimates if your caliper is small
- Indeed, if they affect the estimates much, that's a sign that something is wrong

RDD Estimation

Fuzzy Design

Now we want to estimate:

$$\delta(c) = \frac{\lim_{x \downarrow c} E(Y_1|X=x) - \lim_{x \uparrow c} E(Y_0|X=x)}{\lim_{x \downarrow c} E(D|X=x) - \lim_{x \uparrow c} E(D|X=x)}$$

The numerator is the same as above. Here it's the reduced form.

The denominator adjusts for the fact that we no longer have complete compliance, that is, that the probability of treatment changes discontinuously at the cutoff, but not from 0 to 1. It's the first stage.

How to estimate δ? Run a 2SLS regression of the form:

$$Y = \beta_0 + \beta_1(X_1 - c) + \delta D + \beta_2(X_1 - c)D + \varepsilon$$

using Z as an IV for D, where D is actual treatment status, and $Z = 1$ if $X > c$ and $Z = 0$ otherwise.

More on identifying assumption

- Potential outcomes must vary smoothly with the running variable in the neighborhood of the cutoff. This is another way of saying *potential* outcomes must be balanced on either side of the cutoff.
- What happens if the running variable is manipulable?

Testing for smoothness in the neighborhood of the cutoff

This is untestable, strictly speaking.

- Analogous to random assignment.
- Can't test balance of potential outcomes, but can test balance of baseline characteristics.

Testing for balance

Run an "RD" regression, with baseline variable B as the dependent variable:

$$B = \beta_0 + \beta_1(X_1 - c) + \delta D + \beta_2(X_1 - c)D + \varepsilon$$

for sharp RD, and

$$B = \beta_0 + \beta_1(X_1 - c) + \delta Z + \beta_2(X_1 - c)Z + \varepsilon$$

for fuzzy RD.

Balance tests from Deshpande (2016)

Table 2—Covariate Balance Tests

	Linear		Quadratic			
	Pt. est.	SE	Pt. est.	SE	Mean	Pct. break (linear)
Demographics						
Male	0.0038	(0.0066)	0.0039	(0.0098)	0.6192	0.62
Age at entry	−0.3140	(0.0609)	−0.0480	(0.0898)	12.16	−2.58
Single mother	0.0066	(0.0068)	−2.37e-05	(0.0101)	0.5054	1.30
No parents	−0.0062	(0.0049)	0.0015	(0.0072)	0.1660	−3.73
Latest record date	−87.7	(18.4)	0.9350	(27.2)	11,480	−0.76
Diagnosis						
Mental	0.0022	(0.0060)	0.0064	(0.0089)	0.7041	−0.54
None	−0.0035	(0.0020)	−0.0015	(0.0029)	0.0926	−7.37
Nervous	0.0066	(0.0031)	0.0069	(0.0045)	0.0418	12.00
Endocrine	0.0013	(0.0026)	0.00075	(0.0038)	0.0344	−1.78
Sensory	0.0064	(0.0025)	−0.0023	(0.0036)	0.0278	15.45
Infection	−0.0135	(0.0026)	−0.0118	(0.0040)	0.0189	−8.27
Musculoskeletal	0.00066	(0.0015)	0.0024	(0.0022)	0.0115	2.37
Neoplasm	0.00042	(0.0013)	0.00085	(0.0020)	0.0111	3.81
Respiratory	0.00074	(0.0014)	0.0003	(0.0022)	0.0098	5.21
Pretreatment outcomes						
Child SSI payment	152.1	(28.8)	11.4	(42.5)	\$3,096	4.91
Child earnings	−34.7	(9.9)	−0.709	(13.2)	\$360	−9.64
Family disability applications	−0.0055	(0.0042)	−0.0088	(0.0062)	0.1621	−3.36
Family disability income	39.7	(73.2)	−45.9	(107.6)	\$2,730	1.45
Parent earnings	−574.3	(182.0)	−421.5	(268.8)	\$9,650	−5.95
Observations	81,800		81,800			
Joint F-test	109.07		31.79			
p-value	0.0000		0.2833			

Manipulability

We need the running variable X to be *non-manipulable*.

If people can affect their value of the running variable, they can choose their treatment status. Back to usual self-selection issues. This is probably not an issue for Deshpande, because people can't choose their birthdays and birthday fraud is difficult.

Might be more of an issue for voting studies. Depends on the nature of the vote.

McCrary's density test

McCrary (2008) proposes a test for manipulation based on the density of the running variable on either side of the cutoff.

Intuition: If there are too many people on the favorable side of the cutoff, and too few on the unfavorable side, there is probably manipulation.

He illustrates test in two voting studies.

Study 1: Incumbency advantage in Congressional elections

Idea: Sharp design (originally due to Lee): 50 percent +1 Democratic vote elects the Democrat; less elects the Republican

Running variable is "democratic margin" = share of the popular vote for the D candidate - .5; cutoff is 0.

Question: is the running variable manipulable?

Density of D share

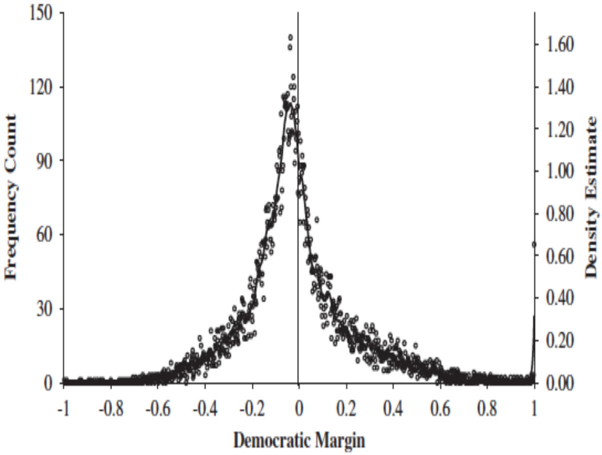


Fig. 4. Democratic vote share relative to cutoff: popular elections to the House of Representatives, 1900–1990.

Study 2: Congressional roll call votes

Idea: Sharp design: 50 percent $+1$ in favor and the measure passes; otherwise it fails

Running variable is share voting in favor; cutoff is .5.

Question: is the running variable manipulable?

Density of share voting in favor

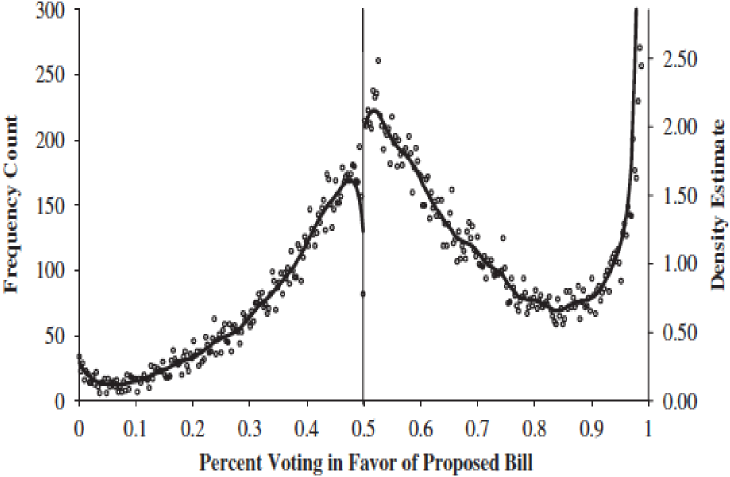


Fig. 5. Percent voting yeay: roll call votes, U.S. House of Representatives, 1857–2004.

Interpreting the RD estimate

RDD identifies a LATE

When the required assumptions are satisfied, RDD yields internally valid estimates for a very special population.

We get treatment impacts for people who are close to the cutoff

External validity of these estimates may be lacking. Treatment effect at other values of the running variable may be different.

RDD gives us a LATE. The treatment impact found is that for compliers. In the case of a sharp RDD, the proportion of compliers $= 1$