Lecture 20: Regression Discontinuity Design

Pierre Biscaye

Fall 2022

Agenda

1. Randomization review
2. Regression Discontinuity Design
3. Sharp vs. Fuzzy RDD
4. Potential concerns with RDD

Quasi-random impact evaluations

- For lots of questions we are interested in, randomization will be impossible, even via encouragement design
- What can we do to get a handle on MLR4 if we can't randomize?
- 4 broad classes of approaches:
  1. Controlling for observables (did in first half); includes matching estimators - will not cover in this course
  2. Regression Discontinuity Design: matching on eligibility for treatment
  3. Panel data techniques: control for broader set of potential omitted variables
  4. Instrumental variable techniques: use a third variable to isolate quasi-random variation in independent variable of interest
- Call these 'quasi-random' because we attempt to identify "as good as random" variation in the independent variable to generate a causal estimate.

Programs with rules

- Start by looking at interventions/programs with rules
  - Many programs have a threshold level for treatment.
  - E.g., poverty-related programs: may need to have income below some level.
  - True in PROGRESA too: needed to have an asset index below some level to be eligible.
- The use of thresholds means that there will be some similar people who get very different access to programs.
- Regression Discontinuity Designs (RDD) take advantage of this.

Regression Discontinuity Design (RDD)

- RDD is one means of estimating treatment effects when treatment is not random.
- Key idea: many programs have thresholds for eligibility.
  - EITC - eligible if income less than some level
  - PROGRESA - eligible if asset index less than some level
  - Scholarship competition - receive a scholarship if in top k% of performers
- The variable determining eligibilty is called the *running variable*.
- RDD uses eligibility thresholds to make progress on MLR 4.
  - Focus on groups with running variable values near the threshold.
  - For these, treatment is "as good as random" conditional on the threshold.

Key RDD insight and assumption

- Key insight: specific threshold is arbitrary, and individuals just below and just above should be similar.
- When threshold determines treatment eligibility, can thus compare these individuals to identify impact of treatment.
- Key assumption: relationship between outcome and running variable would not change at the threshold if not for the treatment.
- Can test for this by checking for
    1. Evidence of manipulation of running variable around threshold, e.g., bunching just inside eligibility region.
    2. Discontinuities across the threshold in variables other than the outcome and treatment.
    3. Either of these issues would mean differences in the outcome at the threshold no longer identify the causal impact of treatment.

Basic RDD setup

- A continuous running variable *Running* is used to determine eligibility for some program.
- Individuals above the eligibility threshold for *Running* receive the program, which we capture by a dummy variable *Treat*.
- The sharp change in *Treat* at the *Running* threshold leads to a discontinuity in outcomes (assuming *Treat* is effective).
- This discontinuity provides an estimate of the Local Average Treatment Effect (LATE) around the eligibility threshold.
  - It is "local" because it is estimated right at the *Running* threshold, rather than across all individuals.
- We estimate the RDD LATE by modeling

$$y_i = \beta_0 + \beta_1 Treat_i + \beta_2 Running_i + \beta_3 Running_i * Treat_i + u_i \qquad (1)$$

- What does each coefficient estimate?
  - Note: we sometimes re-center the running variable as $(Running_i - Threshold)$.

Example: How does Yelp.com affect restaurant profits?

- Many consumers in the US use Yelp to decide which restaurant to eat at.
- Presumably, restaurants with better reviews get more customers.
- Leads to a research question: (how much) do good Yelp reviews influence restaurant profits?

Yelp displays reviewer ratings in stars

- A 4 star average rating for a restaurant on Yelp is considered very good.
- We don't have data on restaurant profits, but suppose we had data on whether we could make reservations at a restaurant (as an indicator of popularity), and estimated

$$Res_i = \beta_0 + \beta_1 4star_i + u_i \tag{2}$$

- We might worry $E[u_i | 4star_i] \neq 0$
- Why?

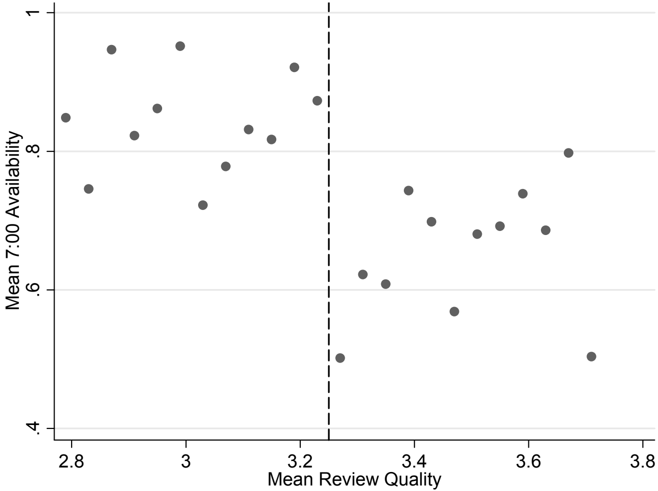
A rule we can exploit for our analysis

- Yelp displays ratings in half-star increments.
- But the true rating is continuous: it is the average number of stars from all of the reviewers.
- So a restaurant with an average score of 3.6 stars will be displayed as 3.5 stars, and so will a restaurant with an average score of 3.4 stars.
- Yelp's rule: $3.25 \leq \textit{rating} < 3.75 \Rightarrow 3.5$ stars.
  $3.75 \leq \textit{rating} < 4.25 \Rightarrow 4$ stars
  - (and so on)

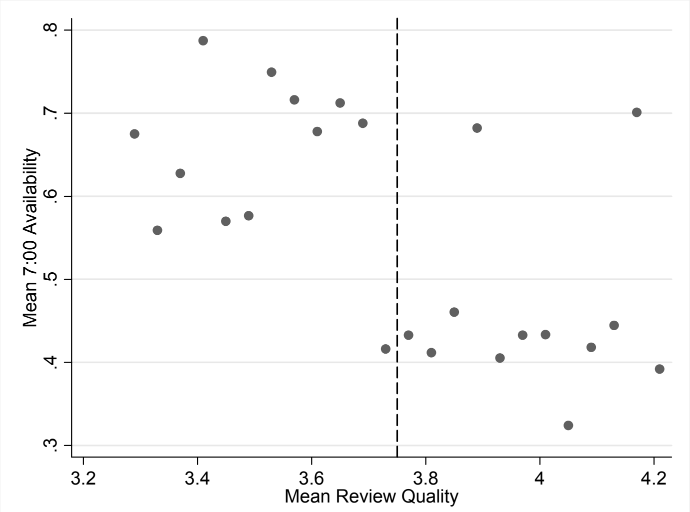
Using the rounded ratings to assess Yelp with an RDD

- We are worried that on average 4 star restaurants may be better than 3.5 star restaurants: omitted variable bias.
- But *some* 4 star restaurants *just barely* receive 4 stars - those with a rating just above 3.75 stars.
- And *some* 3.5 star restaurants *almost* receive 4 stars - those with a rating just below 3.75 stars.
- Reviewers evaluated these restaurants as *very* similar, but Yelp reports their quality as very different.
- An RDD analysis exploits the discontinuity in the treatment (in this case, 4 star display rating) at the threshold value of the running variable (in this case 3.75 star average rating)
  - Restaurants just above and just below this threshold should be similar in quality: eliminate concerns about OVB.
- Does reservation availability vary across display rating thresholds?

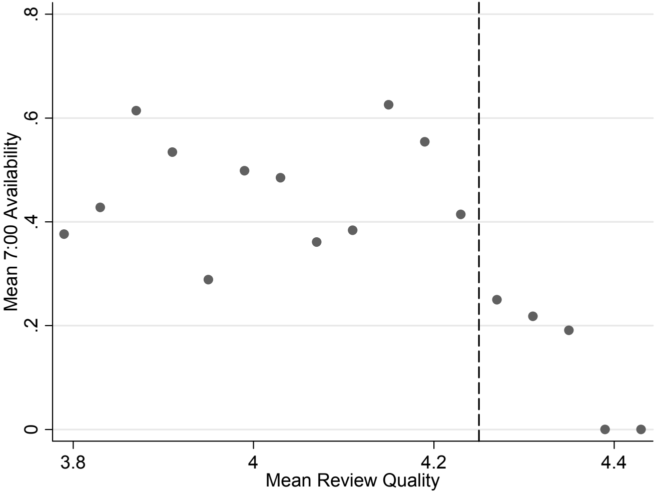
Discontinuity at 3.5 stars



Discontinuity at 4 stars



Discontinuity at 4.5 stars



RDD as a regression

- Suppose $4stars_i = 1$ for a restaurant with a rating between 3.75 and 4.25.
  - Note there are two thresholds here: on the low end compare against 3.5 star restaurants and on the high end compare to 4.5 start restaurants.
  - With an RDD want to focus analysis on observations near the threshold, otherwise introduce OVB.
- We can regress

$$res_i = \beta_0 + \beta_1 4stars_i + \beta_2 rating_i + \beta_3 rating_i * 4stars_i + u_i \tag{3}$$

- *rating* is the running variable, the average of reviewer ratings.
- How to interpret?

Effect of Yelp ratings on reservation availability

	6:00 Availability			7:00 Availability		
Yelp Display Rating	(1)	(2)	(3)	(4)	(5)	(6)
3.5 Yelp Stars	-0.079			-0.213 **		
	(0.086)			(0.096)		
4 Yelp stars		-0.101			-0.192 **	
		(0.075)			(0.093)	
4.5 Yelp stars			0.004			-0.113
			(0.185)			(0.127)
Yelp Rating	-0.228	0.145	-0.131	0.082	0.024	-0.022
	(0.201)	(0.203)	(0.230)	(0.216)	(0.255)	(0.271)
Yelp Rating*Yelp Star	0.372	-0.275	-2.934 **	-0.057	-0.048	-1.817 ***
	(0.287)	(0.309)	(1.342)	(0.335)	(0.375)	(0.674)
Observations	8,705	11,858	5,597	8,705	11,858	5,597

What happened to MLR4?

- Instead of $E[u_i|4star_i] = 0$
- We need $E[u_i|4star_i, rating_i] = 0$
- in other words, we treated the *true* rating as an omitted variable and controlled for it.
- For MLR4 to hold now, we have to assume that the true Yelp rating captures all relevant restaurant characteristics that would affect the outcome and be associated with the Yelp display rating.

Concerns that remain

- We might be worried that the true relationship between ratings and reservations are non-linear.
  - In this case, an apparent discontinuity at the threshold can be due to modeling decisions as opposed to a true effect.
  - Could test sensitivity to this by specifying non-linear functions of *Running*.
  - Should be focusing on observations relatively near the threshold, otherwise estimates can get thrown off.

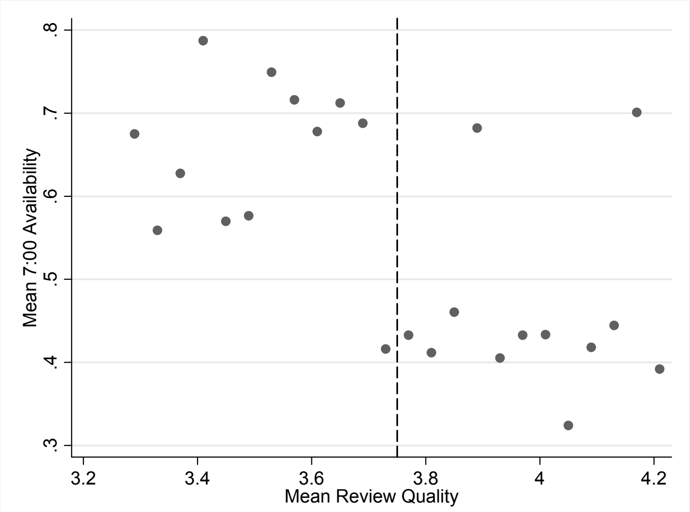
Concerns that remain

- We might be worried that the true relationship between ratings and reservations are non-linear.
  - In this case, an apparent discontinuity at the threshold can be due to modeling decisions as opposed to a true effect.
  - Could test sensitivity to this by specifying non-linear functions of *Running*.
  - Should be focusing on observations relatively near the threshold, otherwise estimates can get thrown off.
- But as long as *u does not* change really sharply around the threshold, RDD is still a valid way to recover causal treatment effects.
  - Can partly test this by running RDD with other characteristics to see if they vary sharply at the threshold.

Concerns that remain

- We might be worried that the true relationship between ratings and reservations are non-linear.
  - In this case, an apparent discontinuity at the threshold can be due to modeling decisions as opposed to a true effect.
  - Could test sensitivity to this by specifying non-linear functions of *Running*.
  - Should be focusing on observations relatively near the threshold, otherwise estimates can get thrown off.
- But as long as *u does not* change really sharply around the threshold, RDD is still a valid way to recover causal treatment effects.
  - Can partly test this by running RDD with other characteristics to see if they vary sharply at the threshold.
- Idea of RDD: treatment changes really sharply at the threshold, so should be able to detect sharp change in outcome as well.
  - This is why graphical evidence is critical with RDD.
  - The graphical evidence also helps easily identify the counterfactual.

Discontinuity at 4 stars



Example 2: Irrigation in Rwanda

- Hillside irrigation systems constructed by the government.
  - Each has a canal that brings water from a distant water source.
  - Gravity fed: pipes bring water from the canal down to the terraces below.
  - Everyone *below* the canal gets access to the water.
  - Access is difficult at best above the canal.
  - But, plots below the canal may be different from plots above the canal in other ways.

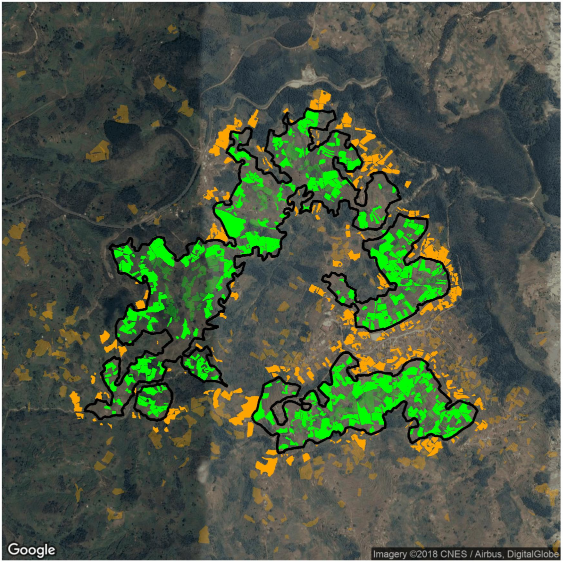
Hillside Irrigation in Rwanda

Estimating the effects of irrigation via RDD

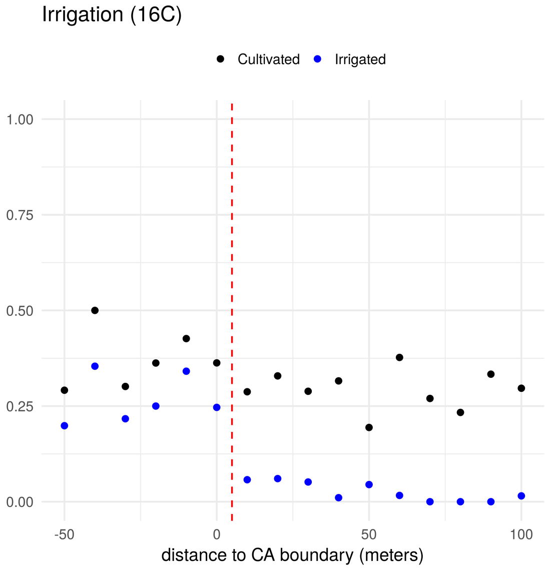
- One approach: treat space as the running variable that determines access to irrigation (treatment).
- Measure where every farm plot is, relative to the canal.
- Plots immediately adjacent to the canal are distance 0
- Plots below the canal have a “negative” distance
- Plots above the canal have a “positive” distance
- Draw comparisons just among plots that are very close to the canal.
- A study based on this was published in the American Economic Review (Jones, Kondylis, Loeser, & Magruder 2022).

Close comparisons

Data on nearby plots



Irrigation practices and distance to canal



Sharp vs. Fuzzy RDD

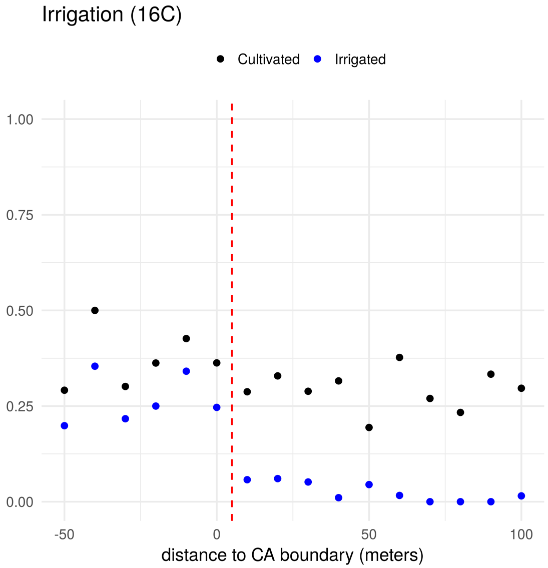
- Yelp display ratings was a *sharp* RDD.
  - Treatment changes exactly the threshold.
  - *All* restaurants with a $3.75 \leq \textit{rating} < 4.25$ show 4 stars.
  - *No* restaurantswith $\textit{rating} < 3.75$ show 4 stars.
- Irrigation may be considered a *fuzzy* RD.
  - Treatment changes at the threshold, but not perfectly.
  - Not all farmers below the canal use irrigation.
  - A few farmers above the canal report using irrigation.
- How does this affect analysis?

Evaluating a fuzzy RDD

$$IRR_p = \beta_0^I + \beta_1^I Access_p + \beta_2^I Dist_p + \beta_3^I Dist_p * Access_p + u_P^I \qquad (4)$$

- Here *Access* is a dummy for being downhill from the canal.
- With a fuzzy RDD, will want to first make sure that there is a discontinuous jump in irrigation at the canal.
- Otherwise, any discontinuity in *outcomes* will not be credible.

Irrigation practices

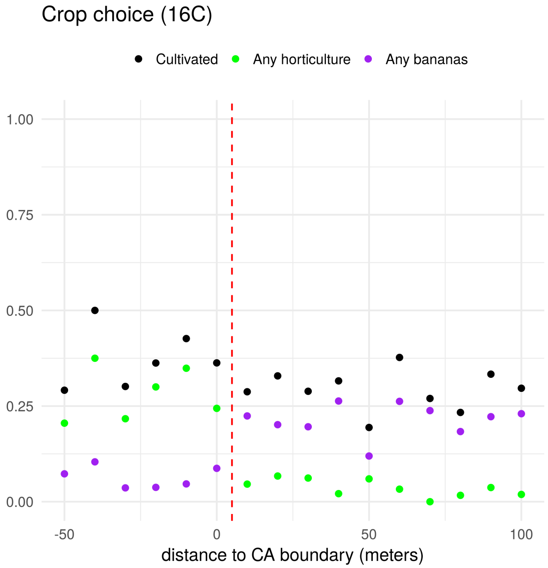


If there is a discontinuous change in treatment, we turn to outcomes

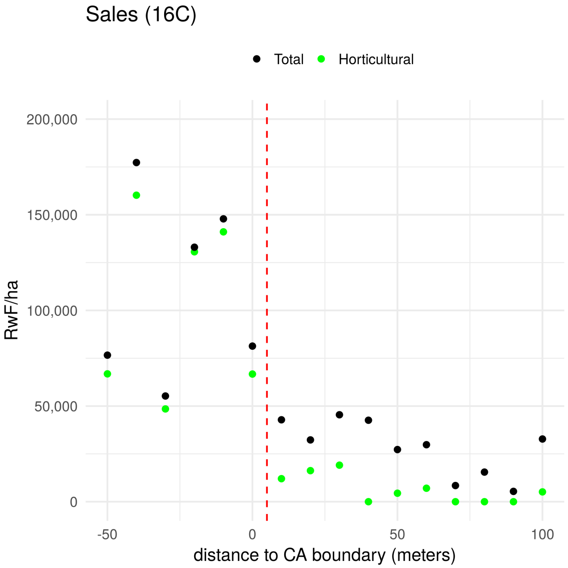
$$y_p = \beta_0 + \beta_1 Access_p + \beta_2 Dist_p + \beta_3 Dist_p * Access_p + u_p \qquad (5)$$

- 2 key outcomes considered in this project:
  1. Horticulture production (crop choice)
  2. Crop sales
- Since this is a *fuzzy* discontinuity, interpret as *ITT* impacts
  - Similar logic to RCT encouragement designs where being "assigned" to treatment does not perfectly predict actual treatment status.

Crop choice


Crop choice (16C)
Cultivated
Any horticulture
Any bananas
1.00
0.75
0.50
0.25
0.00
-50
0
50
100
distance to CA boundary (meters)

Crop sales



Sales (16C)
Total
Horticultural
200,000
150,000
100,000
50,000
0
RwF/ha
-50
0
50
100
distance to CA boundary (meters)

Fuzzy RDD: Getting the ToT estimate

- Since not everyone irrigates, the effect of irrigation on production is not given by the RDD estimator.

$$y_p = \beta_0 + \beta_1 Access_p + \beta_2 Dist_p + \beta_3 Dist_p * Access_p + u_p \tag{6}$$

$$IRR_p = \beta_0^I + \beta_1^I Access_p + \beta_2^I Dist_p + \beta_3^I Dist_p * Access_p + u_P^I \tag{7}$$

$$ToT = \frac{\beta_1}{\beta_1^I} = \frac{\Delta y}{\Delta Irr} \tag{8}$$

- Just as with PROGRESA, the ToT is the treatment effect for *compliers*.
- Here, compliers are those induced to irrigate by the infrastructure (canal).

RDD is a powerful tool

- We often know a reason *why* some people received treatment and some didn't.
- Often, that reason is threshold in an underlying continuous variable.
- Other examples:
  1. Microfinance with eligibility requirements
  2. Admissions into colleges or high schools (maybe esp. in countries with strict entrance exams)
  3. Pensions with age eligibility thresholds
  4. Anti-poverty programs with wealth/income thresholds
  5. Minimum wage laws
  6. Class sizes in Israel

What do we need to worry about in a RDD?

- Was the threshold enforced?
  - Should be able to see discontinuity in treatment in a graph.
  - If not perfectly enforced have a fuzzy RDD.
- Did outcomes really change?
  - Also verifiable graphically.
- Can the running variable be manipulated?
  - Not too worried with the canal (plots can't move).
  - In Yelp case, need to verify that restaurants cannot stay just above critical thresholds, e.g., by writing selected reviewers to ask them to slightly increase their ratings or by posting fake positive reviews.
  - Can check visually for "bunching" of observations around thresholds as a sign of manipulation.

What do we need to worry about in a RDD? (2)

- Does anything else change sharply at the cutoff?
  - E.g., are there other policies that use the *same* threshold for eligibility?
  - Potential concern when using administrative boundaries as thresholds.
- Who are we identifying the effect for?
  - Sharp RDD identifies a LATE around the threshold: would effects be similar in other ranges of the running variable distribution?
  - Fuzzy RDD identifies ToT: Who are the compliers? Do we think they should have the same treatment effects as others? As with randomization: is the ToT close to the ATE?