Lecture 9: Regression Discontinuity Designs (Part I)

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

A Brief Recap of the IV Lectures

- IV estimators offer solutions for tackling noncompliance in randomized trials
  - The assigned treatment $\neq$ the delivered treatment
  - The effect of assigning the treatment is the **intention-to-treat effect**
  - The (randomly) assigned treatment may be used as an IV for actually delivered treatment

- Interpreting IV estimates with heterogeneous potential outcomes is tricky
  - Assuming independence, exclusion, relevance and **monotonicity,** the IV estimates can be interpreted as estimating a **Local Average Treatment Effect**
  - This parameter is estimated only over the **compliers**
  - It is not the ATT, the ATE or the ATU!

Is the LATE a Useful Parameter?

- Imagine, for instance, an insurance company randomly offers some farmers a discount on their insurance premium
  - Getting a discount makes some people more likely to buy crop insurance
  - Nobody is less likely to take up crop insurance by getting a discount (no defiers)
  - Some people take up the offer, others do not
  - Then, using the randomized discount offer as an IV will identify a LATE

- We will now investigate whether the LATE helps us answer various potential uses

Is the LATE Informative of the Effect of...

- Providing crop insurance to all farmers?
- Providing mandatory crop insurance to those farmers who do not have it already?
- Removing the crop insurance subsidy?

Is the LATE Informative of the Effect of...

- Providing crop insurance to all farmers?
  - Nope! That is the ATE, and we saw that LATE $\neq$ ATE
- Providing mandatory crop insurance to those farmers who do not have it already?
- Removing the crop insurance subsidy?

Is the LATE Informative of the Effect of...

- Providing crop insurance to all farmers?
  - Nope! That is the ATE, and we saw that LATE $\neq$ ATE
- Providing mandatory crop insurance to those farmers who do not have it already?
  - Nope! That is the ATU—and we saw LATE $\neq$ ATU
- Removing the crop insurance subsidy?

Is the LATE Informative of the Effect of...

- Providing crop insurance to all farmers?
  - Nope! That is the ATE, and we saw that LATE $\neq$ ATE
- Providing mandatory crop insurance to those farmers who do not have it already?
  - Nope! That is the ATU—and we saw LATE $\neq$ ATU
- Removing the crop insurance subsidy?
  - The causal effect of removing the subsidy is the ITT effect—you do not need LATE for that!

Is the LATE Informative of the Effect of...

- Providing crop insurance to all farmers?
  - Nope! That is the ATE, and we saw that LATE $\neq$ ATE
- Providing mandatory crop insurance to those farmers who do not have it already?
  - Nope! That is the ATU—and we saw LATE $\neq$ ATU
- Removing the crop insurance subsidy?
  - The causal effect of removing the subsidy is the ITT effect—you do not need LATE for that!
- So what does the LATE give us?
  - LATE tells us the effect of getting the compliers to buy insurance
  - Whether this is useful or not depends on what your goals are!

LATE with One-Sided Non-Compliance

- Interpreting LATE is easier when non-compliance is one-sided
- The common case is when nobody in the control group can access the intervention
  - E.g. trials of an experimental drug which is not otherwise available
  - I.e. there are no "always-takers"
  - The 2SLS estimate is still estimated over compliers but since there are no "always-takers", this is the full population of those who were treated!
  - Then $LATE = ATT$!
- The other case of one-sided non-compliance is where there are no "never-takers"
- Then the $LATE = ATU$

Plan for Today

1. A brief recap
2. Introduction to regression discontinuity designs
   - Sharp RDD
   - Fuzzy RDD
3. Analysis of RDDs
   - Checking for balance
   - Manipulation of the running variable (around the threshold)
   - Heaping
4. An empirical example
5. Conclusions and references

Regression Discontinuity Designs: An Introduction

- We started the course by talking about the problem caused by selection bias for the identification of causal effects

- Since then we've investigated multiple ways of getting past it:
  - Randomized trials—selection removed by design
  - OLS to deal with selection on observables—control for relevant differences
  - IV to deal with selection on unobservables—find exogenous and excludable variation affecting X

- We now turn to a further tool in our empirical toolkit: the regression discontinuity design (RDD)

Regression Discontinuity Designs

- Our problem is the same as before:
  - We want to identify the effect of some binary treatment W_i
  - Potential outcomes may be heterogeneous
  - W_i may be correlated with many unobservables

- RDDs aim to offer a solution to the selection problem in settings without randomization

- They do this by looking at a particular deterministic component of program assignment...
  - ...in a way that is not correlated with other characteristics
  - ...trying to mimic a randomized trial in an observational setting

Regression Discontinuity Designs

- The basic idea is simple
- There is some underlying variable (X_i—known as a running variable or an assignment variable) that determines whether you get access to a program based on a threshold
- On crossing the threshold, the probability of treatment jumps discontinuously
- If potential outcomes vary continuously over the threshold, we can get consistent estimates
- At the cutoff, it is as-good-as-random whether you end up to the left-hand or the right-hand side of the cutoff—i.e., whether a unit is treated or not

What kind of discontinuities can you think of?

Sharp RDD

- The clearest application of this approach is in cases where...
  - ...nobody gets the treatment up to some value c
  - ...and everyone gets it after

- The treatment is then defined as

$$W_i = 1\{X_i \geq c\}$$

- We can then look at the jump in the conditional expectation of the outcome (Y_i) given the covariate (X_i) at the threshold c

- NB. We usually want to have a continuous running variable

Sharp RDD

- We can then look at the jump in the conditional expectation of the outcome (Y_i) given the covariate (X_i) at the threshold c

$$\lim_{x \downarrow c} E(Y_i|X_i = x) - \lim_{x \uparrow c} E(Y_i|X_i = x)$$

which can be interpreted as the average causal effect of treatment at the discontinuity point

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c]$$

Sharp RDD: Identification

- In general, remember that we need the Conditional Independence Assumption to hold
  - In the sharp RD case, no treatment and control unit share the same X (no common support)
  - So we cannot directly match observations across treatment and control and get at the treatment effect

- What we will assume is that potential outcomes are continuous in X so we can look at observations infinitesimally close to c

- With this assumption,
  - $E[Y(1)|X = c] = \lim_{x \downarrow c} E[Y_i|X_i = x]$
  - $E[Y(0)|X = c] = \lim_{x \uparrow c} E[Y_i|X_i = x]$

Sharp RDD (Imbens and Lemieux 2008)

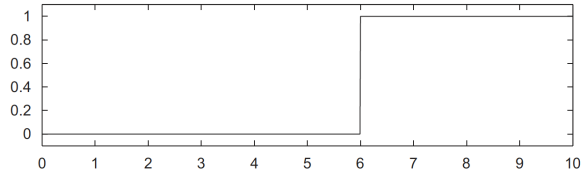


Fig. 1. Assignment probabilities (SRD).

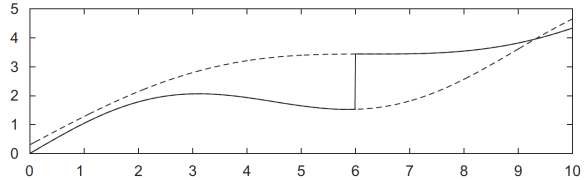
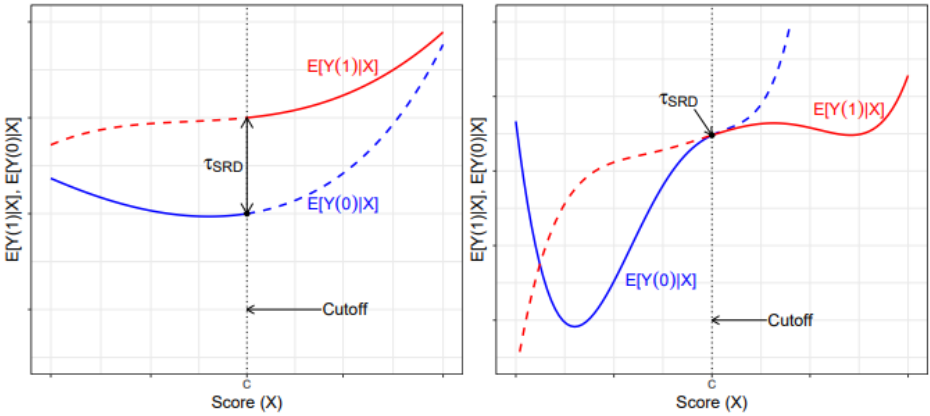


Fig. 2. Potential and observed outcome regression functions.

The "local" nature of RD identification



(a) Mild Heterogeneity

(b) Severe Heterogeneity

Figure 4: Local Nature of the RD Effect

Estimating a Sharp RDD

- We can express a sharp regression discontinuity design as

$$Y_i = \alpha + \beta 1(X_i \geq c) + f(X_i) + \epsilon_i$$

- Here α is a constant, $1(X_i \geq c)$ is an indicator for the running variable or assignment variable crossing the cutoff c, $f(X_i)$ is a control polynomial, and ϵ_i is the error term

- We will discuss the choice of f more next week

Fuzzy RDD

- To estimate a causal effect, we do not require the probability of treatment to go from zero to one at the discontinuity
- We just need it to be discontinuous ("jump")
- As long as that is the case, we can get consistent estimates from the Wald Estimator:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E(Y_i | X_i = x) - \lim_{x \uparrow c} E(Y_i | X_i = x)}{\lim_{x \downarrow c} E(W_i | X_i = x) - \lim_{x \uparrow c} E(W_i | X_i = x)}$$

- This should look familiar by now
- It is the reduced-form (ITT) estimate divided by the first-stage— fuzzy RDD is IV!

Fuzzy RDD

- We still need a strong first stage for the fuzzy RDD to work
- What does this mean? Typically, the first-stage F-statistic should be > 10
- That is, the jump in the probability of treatment should be "clear" enough in terms of statistical significance
- Exclusion restriction must hold at the cutoff
- The effect of the threshold only affects the outcome via actual treatment status
- This could be hard to defend when multiple treatments occur at the same cutoff... But ITT is still valid

Fuzzy RDD (Imbens and Lemieux 2008)

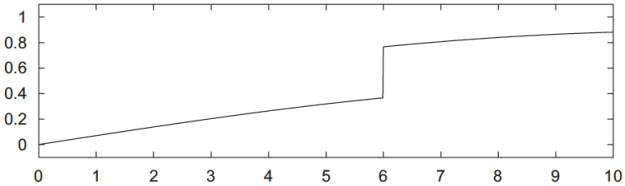


Fig. 3. Assignment probabilities (FRD).

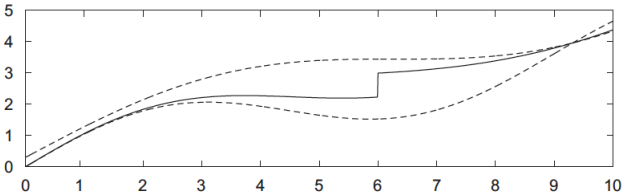


Fig. 4. Potential and observed outcome regression (FRD).

Estimating a Fuzzy RDD

- We can express a fuzzy regression discontinuity design as

$$Y_i = \gamma + \delta \hat{D}_i + g(X_i) + \mu_i$$

- Here γ is a constant, $\hat{D}_i$ comes from the first stage, $g(X_i)$ is a control polynomial (typically estimated separately on each side of the cutoff, and μ_i is the error term

- The first-stage regression is

$$D_i = \lambda + \kappa 1(X_i \geq c) + h(X_i) + \xi_i$$

Checking for Covariate Smoothness

- The identifying assumption of the RDD is that only the probability of treatment jumps discontinuously at the threshold $\Rightarrow$ there is no discontinuous jump in other potential determinants
- This is testable and speaks directly to the validity of the RDD
- A significant advantage over usual IV methods since it allows for checking validity of the instrument (here, crossing the threshold) which is not typically possible in general settings
- You could also interpret RDD as the treatment being randomly assigned within some narrow window around the cutoff and test whether covariates are balanced to the left and right of the cutoff

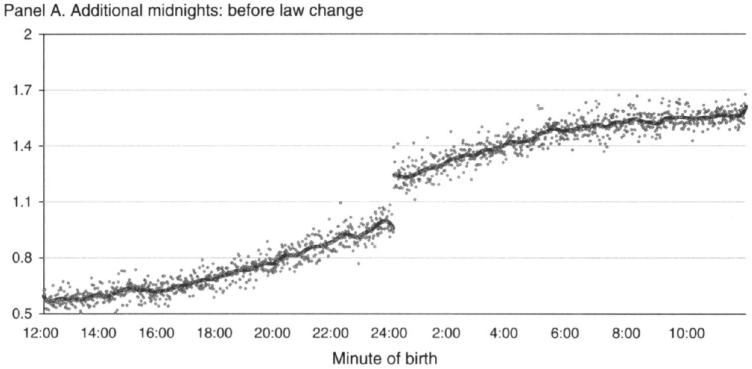
Example: Almond and Doyle (2011)

- Almond and Doyle (2011) study the effect of post-partum hospital stays on child and maternal health
- They utilize a feature of the medical insurance reimbursement in the US
  - Women are covered for x days (usually 2) after childbirth
  - This creates a discontinuity for children born just before/after midnight in the duration for which their mother can stay in the hospital after delivery
- The underlying assumption is that children born just before midnight and just after should be equal in all other respects—timing of birth cannot be determined precisely

Covariate Balance in Almond and Doyle (2011)

	All years		
	Before midnight	After midnight	p-value
Pregnancy characteristics			
At least one pregnancy complication	0.585	0.589	(0.264)
$<$ 9 prenatal visits	0.199	0.204	(0.052)
9–15 prenatal visits	0.695	0.689	(0.043)*
$>$ 15 prenatal visits	0.088	0.089	(0.655)
Prenatal visits missing	0.019	0.019	(0.850)
Mother's characteristics			
Born in California	0.390	0.391	(0.753)
Born outside US	0.472	0.475	(0.411)
First birth	0.400	0.394	(0.047)*
Age	26.82	26.79	(0.489)
High school dropout	0.355	0.356	(0.652)
High school	0.287	0.288	(0.684)
Some college	0.184	0.181	(0.250)
College +	0.164	0.164	(0.907)

First Stage in Almond and Doyle (2011)



- However... no effects on readmission or mortality!

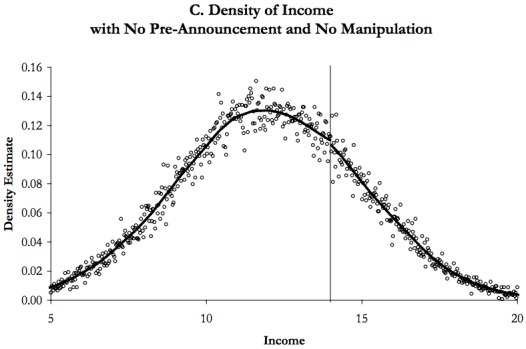
Graphical Analyses of Balance

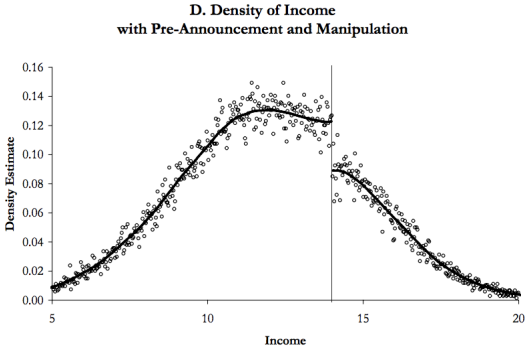
- Tables showing balance on observables have close analogy to RCTs and are familiar to interpret
- But the mere illustration of differences-in-means is not the best way to approach testing for balance in RDDs in general
  - In an RCT, the treatment and control groups are (in expectation) identical in all respects
  - In an RDD, we merely require that covariates not jump discontinuously at threshold
- In most applications, we do not have enough observations just at the threshold to be able to compare only individuals above/below (we take observations close to the threshold)
- A more general test would be to look for discontinuities in the covariate at the threshold

Manipulation of X Around the Threshold

- In an RDD, the assignment variable should, in the neighborhood of the discontinuity, be as good as randomly assigned
- One threat to this validity is if individuals can manipulate the value of the assignment variable
  - Suppose I have to provide remedial instruction to those of you who score below 50%
  - If you are getting only 48%, I might be tempted to give you an extra 2% to make you pass the threshold!
  - Can parents, for instance, misreport children's date of birth so that they enter next year instead of this year?
  - Suppose a child is born on December 31st; do I want to report January 1st instead?
- A consequence of this type of manipulation is that you would expect to see "bunching" in the assignment variable at the threshold
- With such manipulation, the RDD assumptions are unlikely to hold!

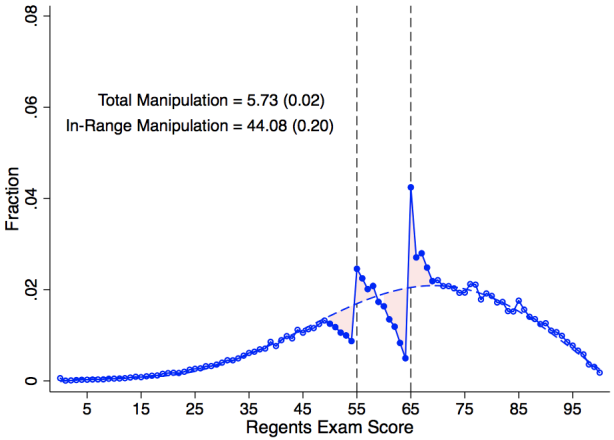
A Hypothetical Example: Misreporting Income (McCrary 2008)





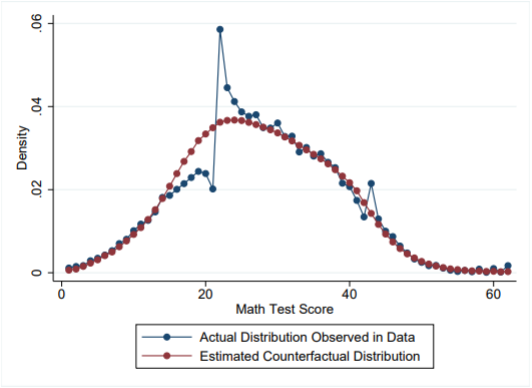
Manipulation in the New York Regents Exam (Dee et al. 2019)

Figure 1: Test Score Distributions for Core Regents Exams, 2004-2010



Manipulation in Swedish Grade 9 Exam (Diamond and Persson 2016)

Figure 3: National Test Score Distribution and Estimated Counterfactual, 2010

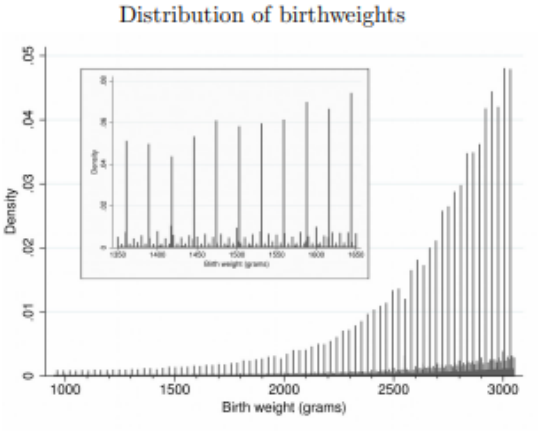


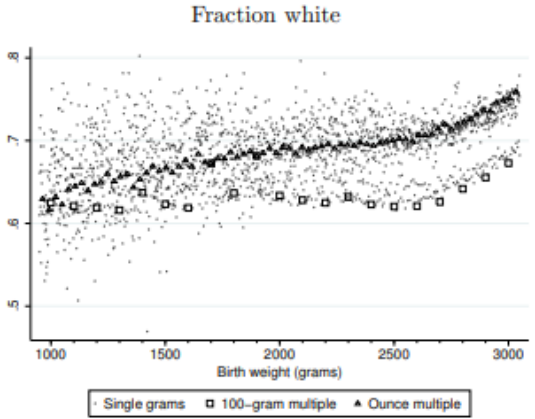
Note: The figure illustrates the national test score distribution and the estimated counterfactual (aggregated from the county*voucher estimated counterfactuals) in 2010. The estimation of the counterfactual density is described in Section 5. The blue connected line plots the actual distribution of test scores, and the red connected line shows the estimated counterfactual density in the absence of manipulation.

Heaping

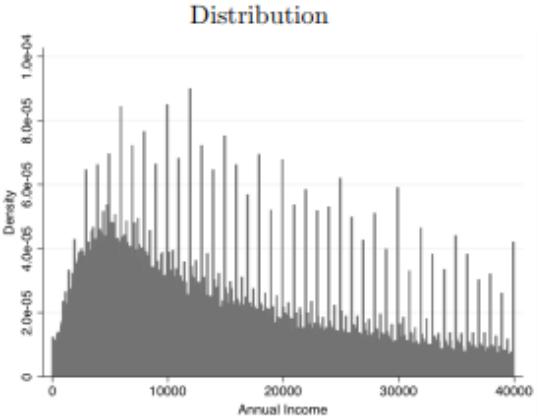
- A lot of policy thresholds used by RDDs center around round numbers
- Unfortunately, often so does (non-random) measurement error
  - E.g. recall errors mean that incomes are often reported in multiples of 50 or 100
  - E.g. age is not well-recorded in countries without high-quality birth records
- This leads to heaps in the data
  - Non-random heaping could cause bias in the RDD estimates (Barreca et al. 2016)
  - Important: this is even though heaping may not result from strategic misreporting or manipulation!
- Recommendation: "Donut RD" which drops sample at heap points

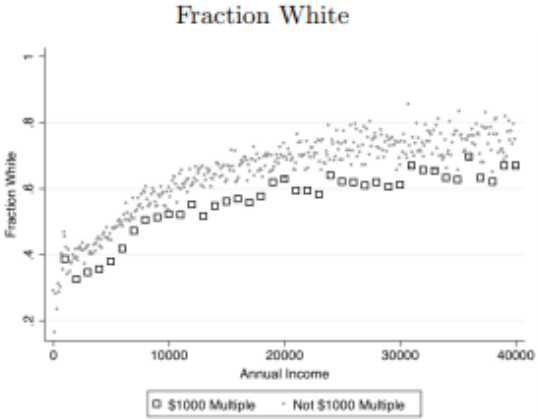
Heaping (Barreca et al. 2016)





Heaping (Barreca et al. 2016)





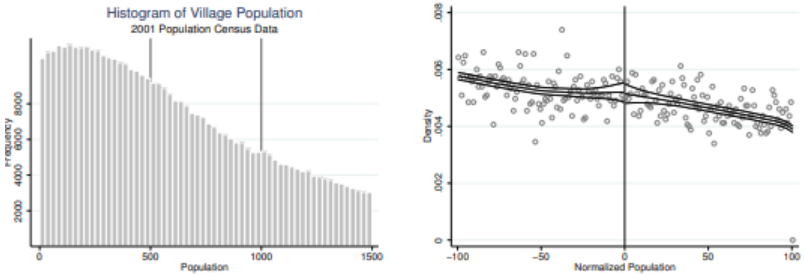
The Impact of Roads on Economic Development

- Access to markets is potentially an important constraint that limits the development of rural communities

- A major focus of governments in many countries has been to build roads and other transport infrastructure
  - Expensive, sometimes disruptive
  - Very hard to evaluate
  - Nobody goes placing roads at random (nor would it be sensible to!)

- Asher and Novosad (2020) look at this question in the Indian context
  - The rural roads initiative aimed to connect all unconnected villages with all-weather road access
  - Target: all villages $>$1000 pop by 2003, $>$500 by 2007, $>$250 after that
  - This gives a potential RDD to evaluate the effect of roads on local economic activity

No Manipulation of the Running Variable



Notes: The figure shows the distribution of village population around the population thresholds. The left panel is a histogram of village population as recorded in the 2001 Population Census. The vertical lines show the program eligibility thresholds used in this paper, at 500 and 1,000. The right panel uses the normalized village population (reported population minus the threshold, either 500 or 1,000). It plots a non-parametric regression to each half of the distribution following McCrary (2008), testing for a discontinuity at zero. The point estimate for the discontinuity is -0.01, with a standard error of 0.05.

No Discontinuities in Background Variables

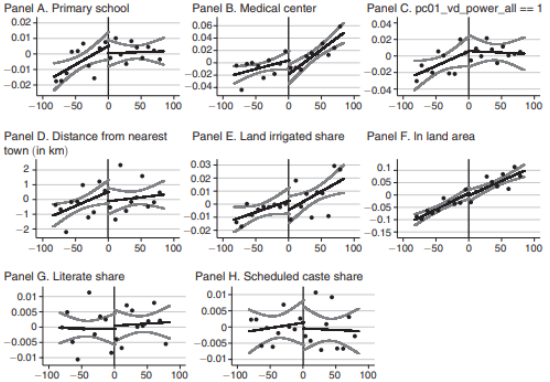


FIGURE 2. BALANCE OF BASELINE VILLAGE CHARACTERISTICS

Notes: The figure plots residualized baseline village characteristics (after controlling for all variables in the main specification other than population) over normalized village population in the 2001 Population Census. Points to the right of 0 are above treatment thresholds, while points to the left of 0 are below treatment thresholds. Each point represents approximately 570 observations. As in the main specification, a linear fit is generated separately for each side of 0, with 95 percent confidence intervals displayed. The sample consists of villages that did not have a paved road at baseline, with baseline population within an optimal bandwidth (84) of the threshold (see text for details).

First Stage (Based on Population Size)

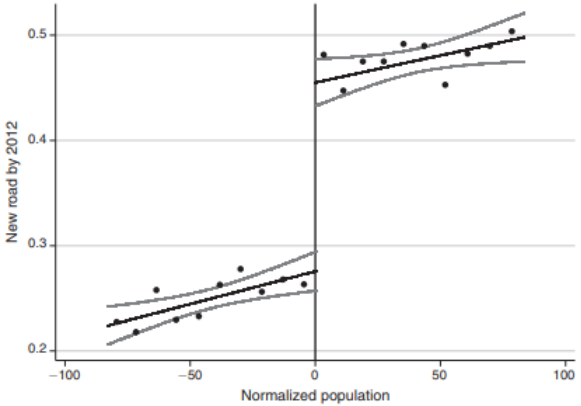


FIGURE 4. FIRST STAGE: EFFECT OF ROAD PRIORITIZATION ON PROBABILITY OF NEW ROAD BY 2012

Reduced Form (Based on Population Size)

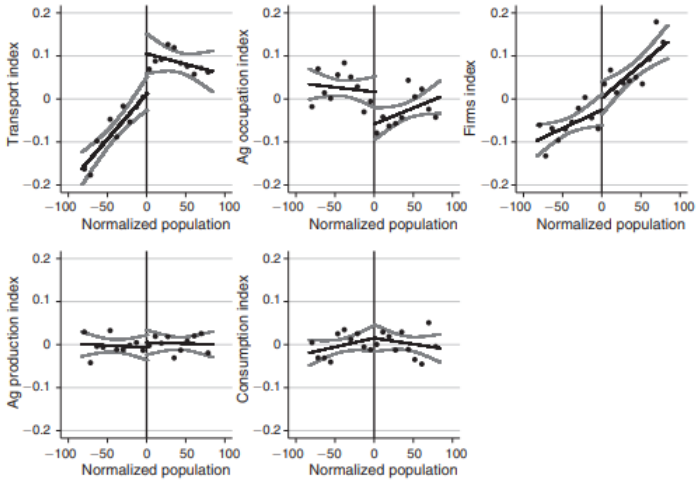


FIGURE 5. REDUCED FORM: EFFECT OF ROAD PRIORITIZATION ON INDICES OF MAJOR OUTCOMES

Summary Results

TABLE 5—IMPACT OF NEW ROAD ON OCCUPATION AND INCOME SOURCE

	Occupation		Household income source	
	Agriculture	Manual labor	Agriculture	Manual labor
New road	−0.092	0.072	−0.030	−0.011
	(0.043)	(0.043)	(0.044)	(0.044)
Control group mean	0.476	0.448	0.418	0.507
Observations	11,432	11,432	11,432	11,432
R^2	0.28	0.26	0.31	0.28

Summary and Next Lecture

- The basic idea behind RDD is very straightforward
- In some sense, it builds on the RCT and IV analyses that you have seen already
- But the implementation of RDD has several "best-practices" that distinguish it from more common IV type papers
- E.g., issues of functional form, bandwidth choice, and graphical analyses
- In the next lecture, we will look at these in some detail, along with examples

Readings

Recommended

- The chapter on RDD in Mixtape (p. 153-204)

Other suggested readings

- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142(2), 615-635.
- Chapter 6 in Angrist and Pischke (2009)
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48, 281-355.

Papers Mentioned in This Lecture

- Almond, D., & Doyle, J. J. (2011). After midnight: A regression discontinuity design in length of postpartum hospital stays. American Economic Journal: Economic Policy, 3(3), 1-34.
- Asher, S., & Novosad, P. (2020). Rural roads and local economic development. American Economic Review, 110(3), 797-823.
- Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. Economic inquiry, 54(1), 268-293.
- Dee, T. S., Dobbie, W., Jacob, B. A., & Rockoff, J. (2019). The causes and consequences of test score manipulation: Evidence from the New York regents examinations. American Economic Journal: Applied Economics, 11(3), 382-423.
- Diamond, R., & Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests (No. w22207). National Bureau of Economic Research.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics, 142(2), 698-714.