Regression Discontinuity

Mauricio Romero

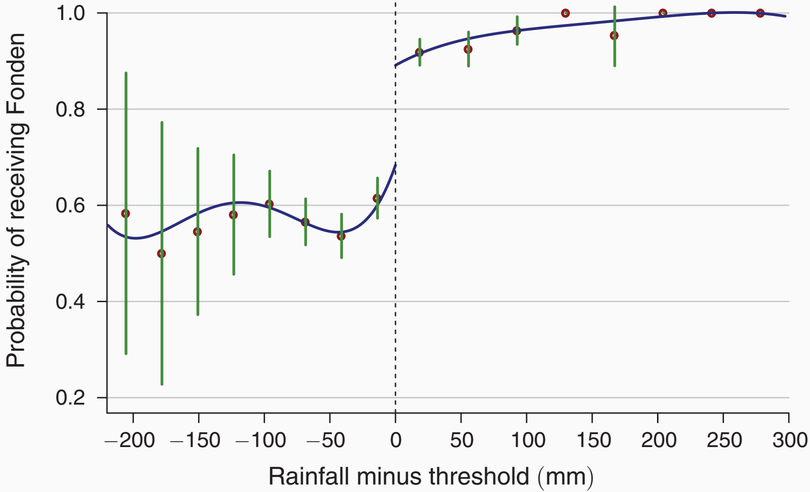
Introduction

What is regression discontinuity design (RDD)?

- Donald Campbell, educational psychologist, invented regression discontinuity design but then it went dormant for decades
- Angrist and Lavy (1999) and Black (1999) independently rediscover it
- It has become incredibly popular in economics

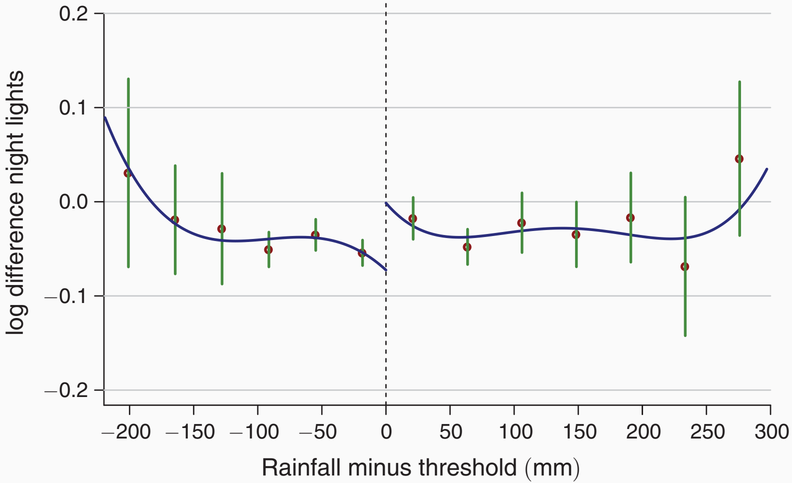
Tell me what you think is happening

Panel A. First stage



Tell me what you think is happening

Panel B. Intention-to-Treat



Running example from Mexico 1

American Economic Journal: Applied Economics 2020, 12(4): 164–195
https://doi.org/10.1257/app.20190002

Rules for Recovery: Impact of Indexed Disaster Funds on Shock Coping in Mexico†

By Alejandro del Valle, Alain de Janvry, and Elisabeth Sadoulet*

Government provision of disaster transfers is typically hampered by liquidity constraints and by weak rules and administrative capacity to disburse reconstruction resources. We show that by easing these hurdles, Mexico's indexed disaster fund (Fonden) considerably accelerates economic recovery after a disaster. To estimate Fonden impact on recovery, as measured by night lights, we exploit the heavy rainfall index that determines program eligibility. We find that for one year after a disaster, eligible municipalities are 6 percent brighter than those ineligible, with gains likely concentrated among less resilient municipalities. We additionally document how Fonden rules shield resources from political abuse. (*JEL* G22, H12, H84, O13, O18, Q54, R38)

What is a regression discontinuity design?

- Goal: estimate some causal effect of a treatment on some outcome
- Problem: Selection bias (i.e., $E[Y^0|D=1] \neq E[Y^0|D=0]]$)
- RDD basic idea: if treatment assignment occurs abruptly when some underlying variable X (the "running variable") passes a cutoff c_0, then we can use that arbitrary rule to estimate the causal effect *even of a self-selected treatment*

Arbitrary rules

- Firms, schools and govt agencies assign “things” based on arbitrary thresholds of continuous variables
- Consequently, probabilities of treatment will “jump” when that running variable exceeds a known threshold
    - Academic test scores: scholarships or prizes, higher education admission, certificates of merit
    - Poverty scores: (proxy-)means-tested anti-poverty programs (generally: any program targeting that features rounding or cutoffs)
    - Land area: fertilizer program or debt relief initiative for owners of plots below a certain area
    - Date: age cutoffs for pensions; dates of birth for starting school with different cohorts; date of loan to determine eligibility for debt relief
    - Elections: fraction that voted for a candidate of a particular party

Selection examples and solutions from the literature

Think of these in light of a treatment where $E[Y^0|D=1] \neq E[Y^0|D=0]$

- Yelp rounded a continuous score of ratings to generate stars
- US targeted air strikes in Vietnam using rounded risk scores
- Universal healthcare after age 65
- When a newborn's birthweight is below 1,500 grams it gets intensive medical care

Sharp vs. Fuzzy RDD

- There's traditionally thought to be two kinds of RD designs:

    1. Sharp RDD: Treatment is a deterministic function of running variable, X

    2. Fuzzy RDD: Discontinuous "jump" in the *probability* of treatment when $X > c_0$. Cutoff is used as an instrumental variable for treatment

- Fuzzy is a type of IV strategy and requires explicit IV estimators like 2SLS

Sharp Design

Treatment assignment in the sharp RDD

Deterministic treatment assignment ("sharp RDD")
In Sharp RDD, treatment status is a deterministic and discontinuous function of a covariate, X_i:

$$T_i = \begin{cases} 1 \text{ if} & X_i \geq c_0 \\ 0 \text{ if} & X_i < c_0 \end{cases}$$

where c_0 is a known threshold or cutoff. In other words, if you know the value of X_i for a unit i, you know treatment assignment for unit i with certainty.

Treatment effect definition and estimation

Definition of treatment effect
The treatment effect δ, is the discontinuity in the conditional expectation function:

$$\begin{aligned} \delta &= lim_{X_i \rightarrow c_0} E[Y_i^1 | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i^0 | X_i = c_0] \\ &= lim_{X_i \rightarrow c_0} E[Y_i | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i | X_i = c_0] \end{aligned}$$

Average causal effect of the treatment at the discontinuity

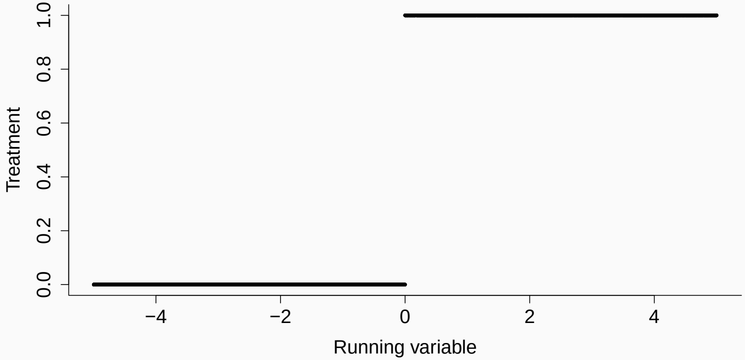
$$\delta_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c_0]$$

T is correlated with X and deterministic function of X; overlap only occurs in the limit and thus the treatment effect is in the limit as X approaches c_0

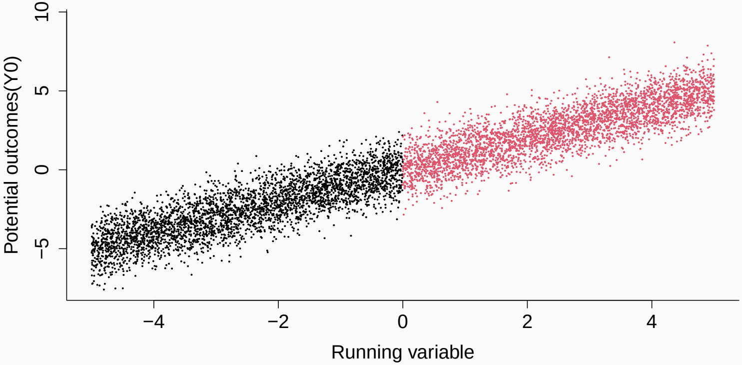
Simulations!

```
## Basic RD Model
N=1000 #number of observations
X=runif(N,-5,5)
Y0 <-  rnorm(n=N, mean=X, sd=1)      # control potential outcome
Y1 <-  rnorm(n=N, mean=X+2, sd=1)    # treatment potential outcome
#You only get treatment if X>0
Treatment=(X>=0)
#What we observe
Y=Y1*Treatment+Y0*(1-Treatment)
```

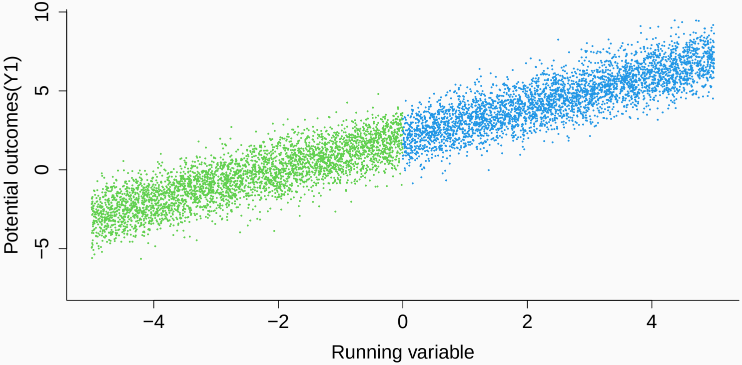
Treatment assignment



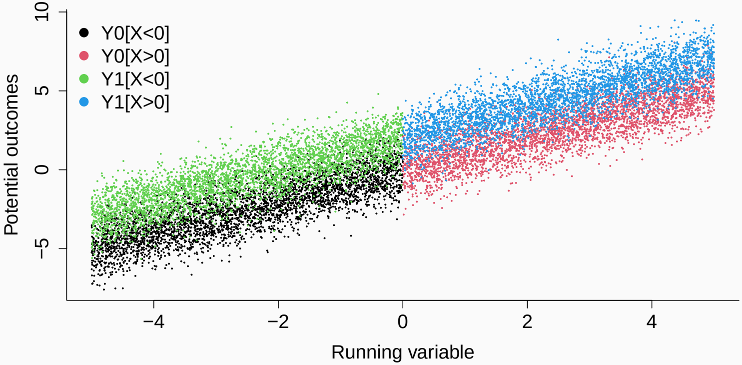
Potential outcomes (Y_0)



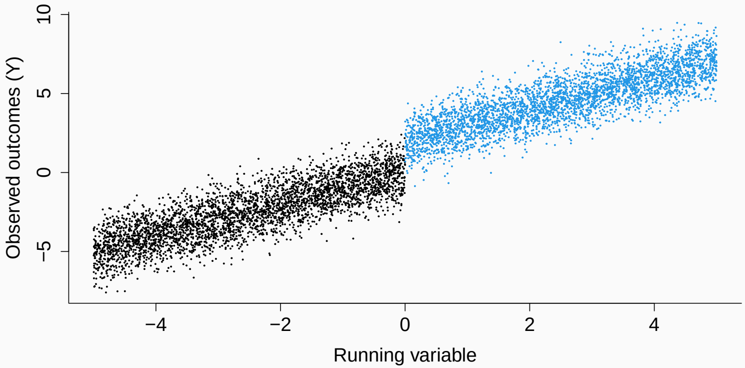
Potential outcomes (Y_1)



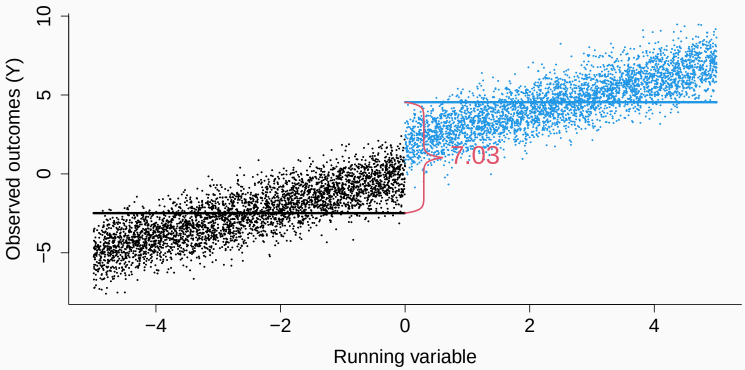
Potential outcomes (Y_0 and Y_1)



Observed outcomes

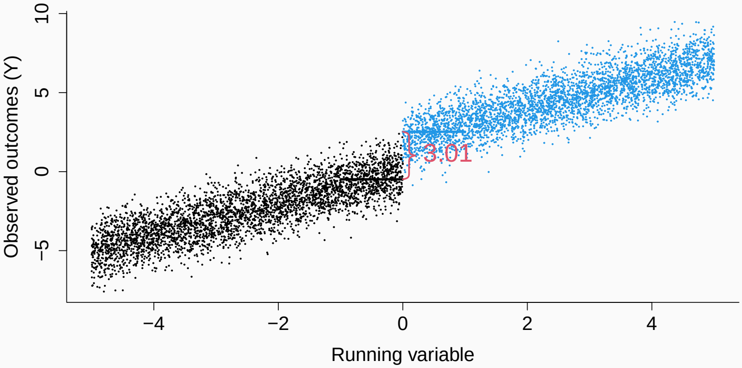


RDD estimator



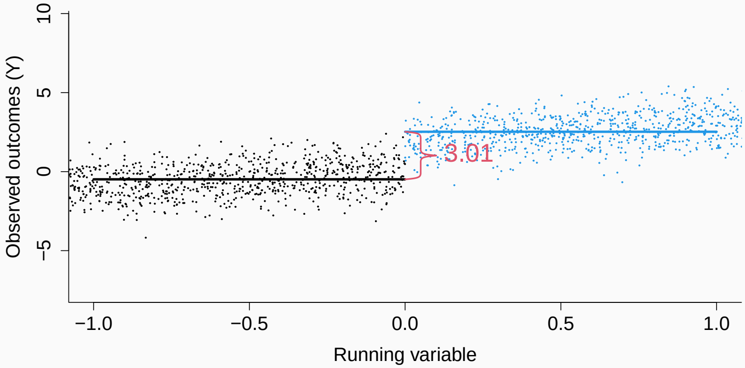
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

RDD estimator



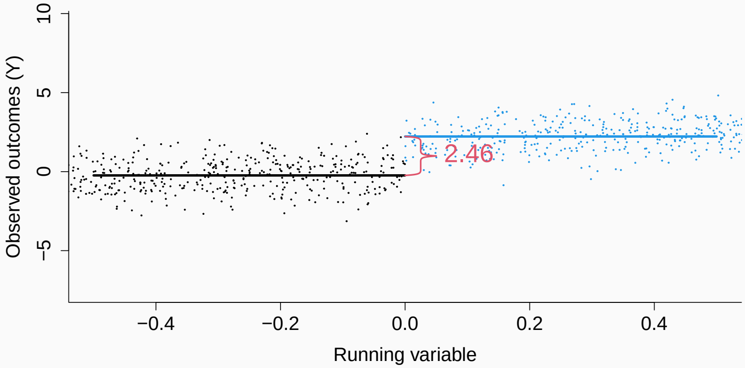
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-1 \leq X_i \leq 1$ via OLS

RDD estimator



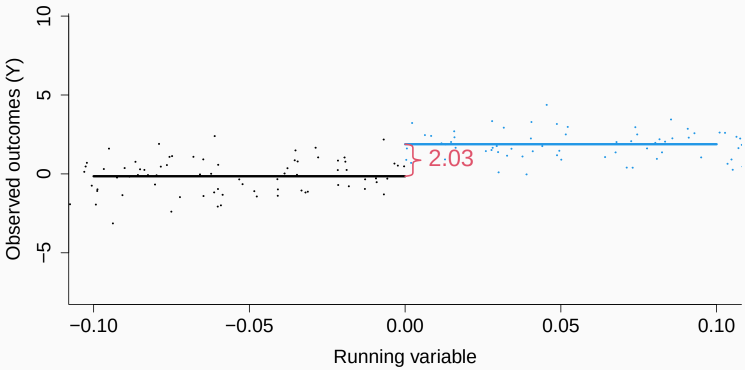
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-1 \leq X_i \leq 1$ via OLS

RDD estimator



Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-0.5 \leq X_i \leq 0.5$ via OLS

RDD estimator



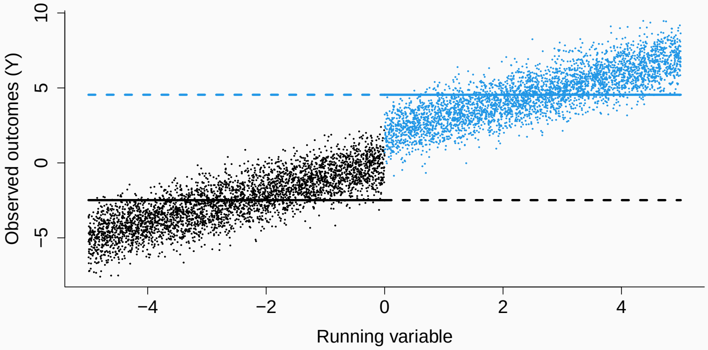
Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-0.1 \leq X_i \leq 0.1$ via OLS

Extrapolation

Extrapolation

- In RDD, the counterfactuals are conditional on X
- We use *extrapolation* in estimating treatment effects with the sharp RDD because we do not have overlap
  - Left of cutoff, only non-treated observations, $T_i = 0$ for $X < c_0$
  - Right of cutoff, only treated observations, $T_i = 1$ for $X \geq c_0$
- The extrapolation is to a counterfactual

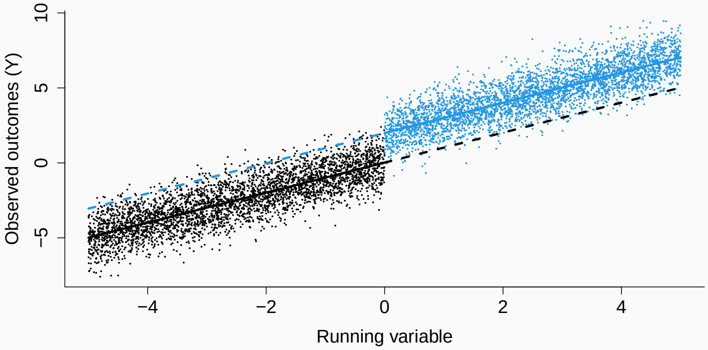
Approximate the limiting parameter using units left and right of the cutoff



Dashed lines are extrapolations

Equivalent to estimating $Y_i = \alpha + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

Approximate the limiting parameter using units left and right of the cutoff



Dashed lines are extrapolations

Equivalent to estimating $Y_i = \alpha + \beta X_i + \lambda X_i * T_i + \delta T_i + \varepsilon_i$ for $-5 \leq X_i \leq 5$ via OLS

Smoothness assumption

Key identifying assumption

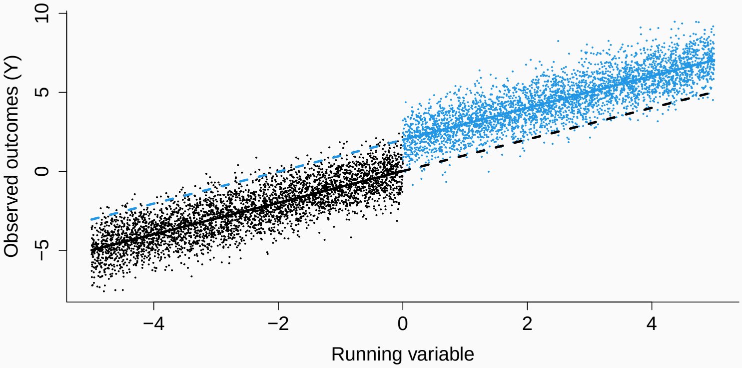
Smoothness (or continuity) of conditional expectation functions (Hahn, Todd and Van der Klaauw 2001)
$E[Y_i^0|X = c_0]$ and $E[Y_i^1|X = c_0]$ are continuous (smooth) in X at c_0

- Potential outcomes not actual outcomes
- If population average *potential outcomes*, Y^1 and Y^0, are smooth functions of X through the cutoff, c_0, then potential average outcomes *won't* jump at c_0.
- Implies the cutoff is exogenous – i.e., nothing else changes related to potential outcomes at c_0
- Unobservables are evolving smoothly, too, through the cutoff

Smoothness is the identifying assumption and untestable

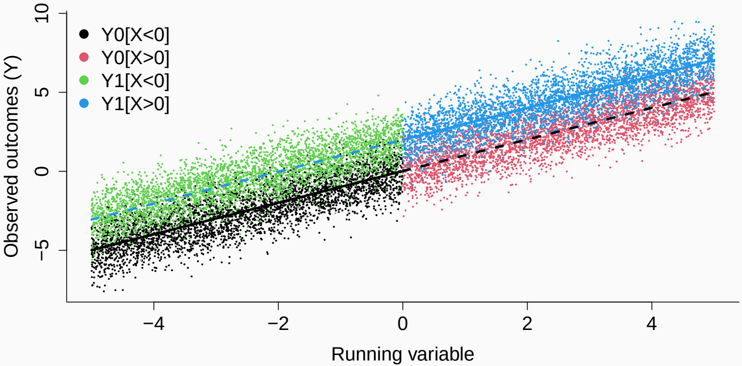
- The smoothness assumption allows us to use average outcome of units right below the cutoff as a valid counterfactual for units right above the cutoff
- Extrapolation is allowed if smoothness is credible, and extrapolation is nonsensical if smoothing isn't credible
- Why not directly testable? Because potential outcomes are not observable

Approximate the limiting parameter using units left and right of the cutoff



Dashed lines are extrapolations

Approximate the limiting parameter using units left and right of the cutoff



Dashed lines are extrapolations

Estimation

Re-centering the data

- It is common for authors to transform X by "centering" at c_0:

$$Y_i = \alpha + \beta(X_i - c_0) + \lambda(X_i - c_0) * T_i + \delta T_i + \varepsilon_i$$

- This doesn't change the interpretation of the treatment effect – only the interpretation of the intercept.

Nonlinearities

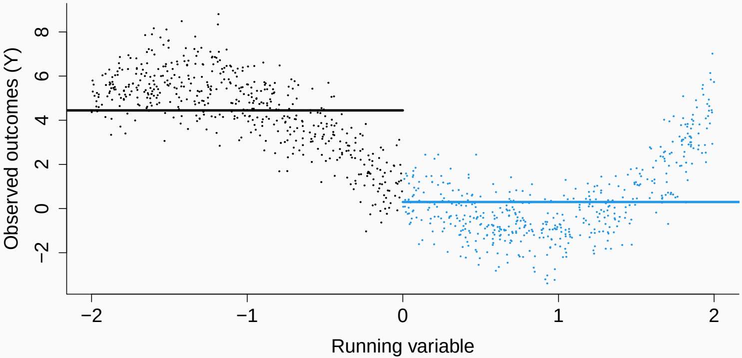
Nonlinearity bias

- Smoothness and *linearity* are different things.
- What if the trend relation $E[Y_i^0|X_i]$ does not jump at c_0 but rather is simply nonlinear?
- Then your linear model will identify a treatment effect when there isn't because the functional form had poor predictive properties beyond the cutoff
- Let's look at a simulation

Simulations!

```
## Non linear RD
N=1000 #number of observations
X=runif(N,-2,2)
X2=X*X
X3=X*X*X
#You only get treatment if X>0
Treatment=(X>=0)
#DGP (noticce there is no treatment effect)
Y=1+0*Treatment-4*X+X2+X3+rnorm(N)
#Constant Models
Const1=lm(Y~Treatment)
Const2=lm(Y~Treatment,subset=abs(X)<1)
Const3=lm(Y~Treatment,subset=abs(X)<0.5)
Const4=lm(Y~Treatment,subset=abs(X)<0.1)
```

Non-Linear



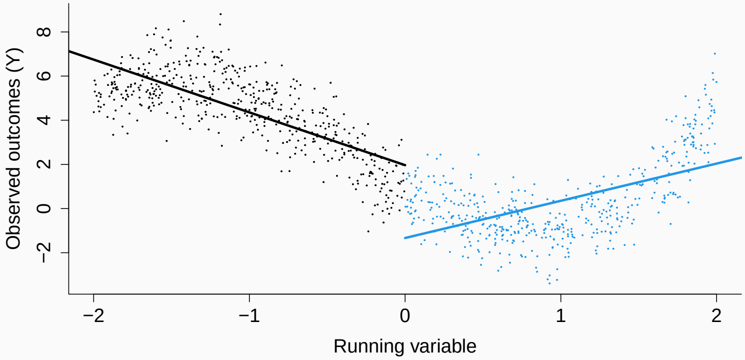
Dashed lines are extrapolations

	(1)	(2)	(3)	(4)
Treatment	-4.15^{***}	-3.64^{***}	-2.03^{***}	-0.74^{***}
	(0.11)	(0.12)	(0.15)	(0.24)
Constant	4.45^{***}	3.18^{***}	2.02^{***}	1.24^{***}
	(0.08)	(0.09)	(0.11)	(0.19)
Sample	*Full*	$\|X\| < 1$	$\|X\| < 0.5$	$\|X\| < 0.1$
Observations	1,000	508	243	48
Note:			*p<0.1; **p<0.05; ***p<0.01	

Simulations!

```
#Linear Models
Linear1=lm(Y~Treatment+X+X*Treatment)
Linear2=lm(Y~Treatment+X+X*Treatment,subset=abs(X)<1)
Linear3=lm(Y~Treatment+X+X*Treatment,subset=abs(X)<0.5)
Linear4=lm(Y~Treatment+X+X*Treatment,subset=abs(X)<0.1)
```

Non-Linear



Dashed lines are extrapolations

	(1)	(2)	(3)	(4)
Treatment	−3.30***	−0.45**	0.10	−1.35***
	(0.17)	(0.18)	(0.25)	(0.48)
X	−2.39***	−4.21***	−5.82***	8.73
	(0.11)	(0.21)	(0.64)	(6.57)
X*Treatment	4.08***	2.33***	3.27***	−4.34
	(0.15)	(0.30)	(0.85)	(8.86)
Constant	1.97***	0.95***	0.56***	1.63***
	(0.12)	(0.13)	(0.19)	(0.35)
Sample	*Full*	$\|X\| < 1$	$\|X\| < 0.5$	$\|X\| < 0.1$
Observations	1,000	508	243	48
Note:			*p<0.1; **p<0.05; ***p<0.01	

Sharp RDD: Nonlinear Case

- Suppose the nonlinear relationship is $E[Y_i^0|X_i] = f(X_i)$ for some reasonably smooth function $f(X_i)$
- In that case we'd fit the regression model:

$$Y_i = f(X_i) + \delta T_i + \eta_i$$

- There are 2 common ways of approximating $f(X_i)$

Nonlinearities

"higher order polynomials" but problematic due to overfitting. Gelman and Imbens 2018 recommend at best a quadratic

1. Use global and local regressions with $f(X_i)$ equalling a p^{th} order polynomial

$$Y_i = \alpha + \delta T_i + \beta_1 x_i + \beta_2 x_i^2 + \lambda_1 x_i * T_i + \lambda_2 x_i^2 * T_i + \eta_i$$

2. Or use some nonparametric kernel method (we won't cover that)

General case

Different polynomials on the 2 sides of the discontinuity

- We can generalize the function, $f(x_i)$, by allowing it to differ on both sides of the cutoff by including them both individually and interacting them with T_i.

- In that case we have:

$$
\begin{aligned}
E[Y_i^0|X_i] &= \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \cdots + \beta_{0p}\tilde{X}_i^p \\
E[Y_i^1|X_i] &= \alpha + \delta + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + \cdots + \beta_{1p}\tilde{X}_i^p
\end{aligned}
$$

where $\tilde{X}_i$ is the centered running variable (i.e., $X_i - c_0$).

- Re-centering at c_0 ensures that the treatment effect at $X_i = c_0$ is the coefficient on T_i in a regression model with interaction terms

Different polynomials on the 2 sides of the discontinuity

- To derive a regression model, first note that the observed values must be used in place of the potential outcomes:

$$E[Y|X] = E[Y^0|X] + \left(E[Y^1|X] - E[Y^0|X]\right) T$$

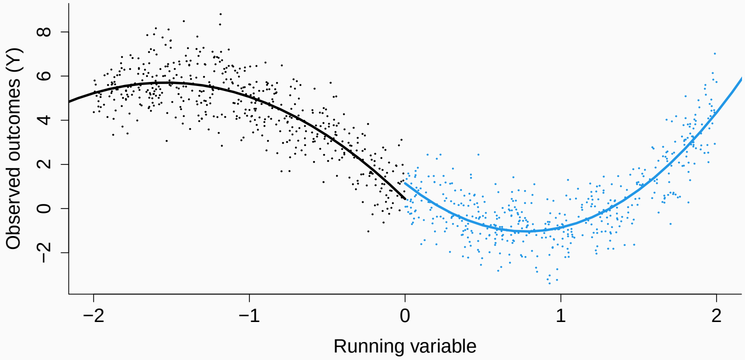
- Regression model you estimate is:

$$\begin{aligned} Y_i &= \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \cdots + \beta_{0p}\tilde{x}_i^p \\ &\quad + \delta T_i + \beta_1^* T_i\tilde{x}_i + \beta_2^* T_i\tilde{x}_i^2 + \cdots + \beta_p^* T_i\tilde{x}_i^p + \varepsilon_i \end{aligned}$$

  where $\beta_1^* = \beta_{11} - \beta_{01}$, $\beta_2^* = \beta_{21} - \beta_{21}$ and $\beta_p^* = \beta_{1p} - \beta_{0p}$

- The treatment effect at c_0 is δ

Non-Linear



Dashed lines are extrapolations

	(1)	(2)	(3)	(4)
Treatment	0.71***	0.05	−0.67*	−0.49
	(0.19)	(0.26)	(0.37)	(0.79)
X	−6.85***	−6.24***	1.02	−20.48
	(0.31)	(0.87)	(2.54)	(28.64)
X2	−2.23***	−1.96**	13.40***	−289.92
	(0.15)	(0.82)	(4.82)	(276.63)
X*Treatment	1.24***	3.43***	−1.36	4.82
	(0.44)	(1.18)	(3.45)	(36.76)
X2*Treatment	5.84***	2.88**	−17.78***	493.05
	(0.21)	(1.12)	(6.57)	(357.01)
Constant	0.44***	0.60***	1.16***	1.14*
	(0.14)	(0.19)	(0.28)	(0.59)
Sample	*Full*	$\|X\| < 1$	$\|X\| < 0.5$	$\|X\| < 0.1$
Observations	1,000	508	243	48
Note:				*p<0.1; **p<0.05; ***p<0.01

Testing for violations

Robustness against what?

- Are you done now that you have your main results? No
- You main results are only causal insofar as smoothness is a credible belief, so you need to convince the reader this is true
- You must now scrutinize alternative hypotheses that are consistent with your main results through sensitivity checks, placebos and alternative approaches

Main Challenges

- Classify your concern regarding smoothness violations into two categories:
    - Manipulation on the running variable
    - Endogeneity of the cutoff
- Most robustness is aimed at building credibility around these

Manipulation of your running variable score

- Treatment is not as good as randomly assigned around the cutoff, c_0, when agents can "perfectly" manipulate their running variable. This happens when:

  1. The assignment rule is known in advance
  2. Agents are interested in adjusting
  3. Agents have the time/ability to adjust

- Since necessarily treatment assignment is no longer independent of potential outcomes, it's likely this implies smoothness has been violated

A badly designed RCT

- Suppose a doctor randomly assigns heart patients to statin and placebo to study the effect of the statin on heart attacks within 10 years
- Patients are placed in two different waiting rooms, A and B, and plans to give those in A the statin and those in B the placebo
- The doors are unlocked and movement between the two can happen

McCrary Density Test

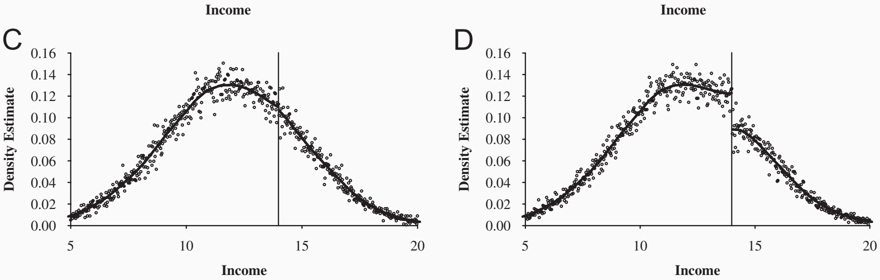
We would expect waiting room *A* to become *crowded*. In the RDD context, sorting on the running variable implies heaping on the "good side" of c_0

- McCrary (2008) test: under the null the *density* should be continuous at the cutoff
- Under the alternative hypothesis, the density should increase at the "good side" of c_0
    1. Partition the running variable into bins and calculate frequencies in each bin
    2. Treat those frequency counts as dependent variable in an RD regression
- You need no jump to "pass" this test

McCrary density test

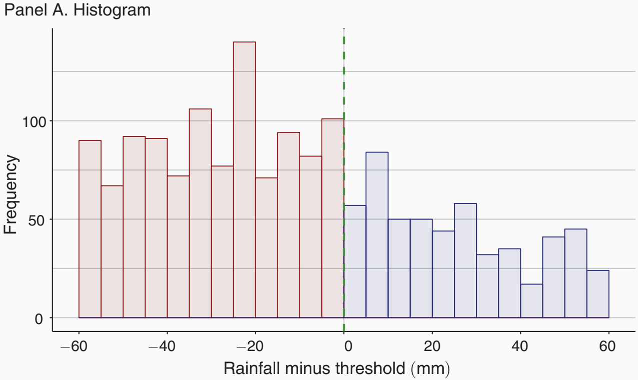
- The McCrary Density Test has become **mandatory** for every analysis using RDD.
  - You can install `rdrobust` for Stata/R, and it will implement the test

McCrary density test

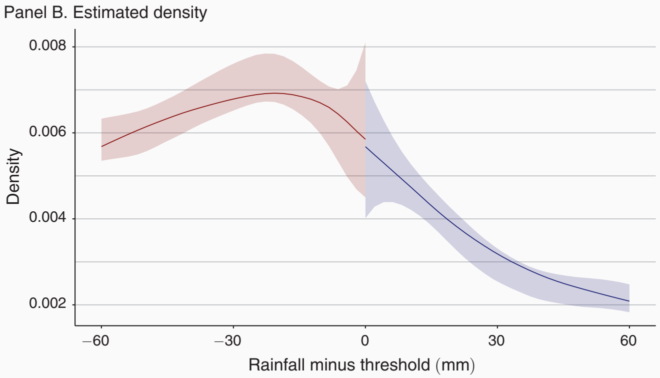


Panel C is density of income when there is no pre-announcement and no manipulation. Panel D is the density of income when there is pre-announcement and manipulation. From McCrary (2008).

McCrary density test – FONDEN running example



McCrary density test – FONDEN running example



Caveats about McCrary Density Test

- For RDD to be useful, you need to know something about the mechanism generating the running variable and how susceptible it could be to manipulation
- A discontinuity in the density is "suspicious" – it *suggests* manipulation of X around the cutoff is probably going on. In principle one doesn't need continuity.
- This is a data-hungry test. You need a lot of observations at c_0 to distinguish a discontinuity from noise

Visualizing manipulation — Proxy means test in Colombia

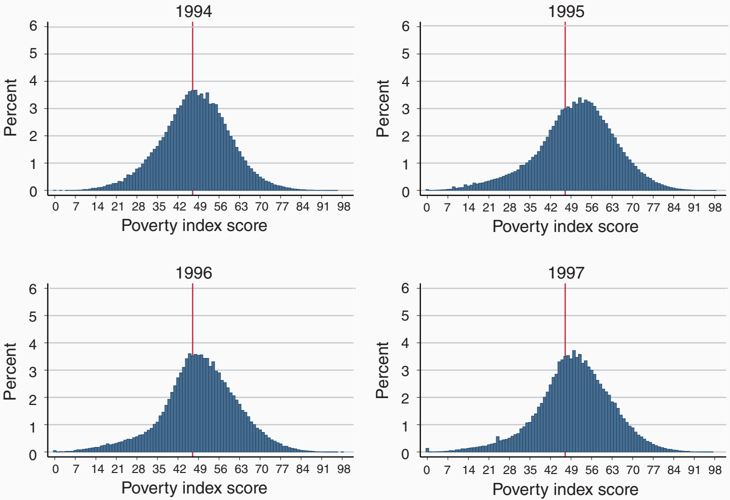
American Economic Journal: Economic Policy 3 (May 2011): 41–65
http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.2.41

Manipulation of Social Program Eligibility†

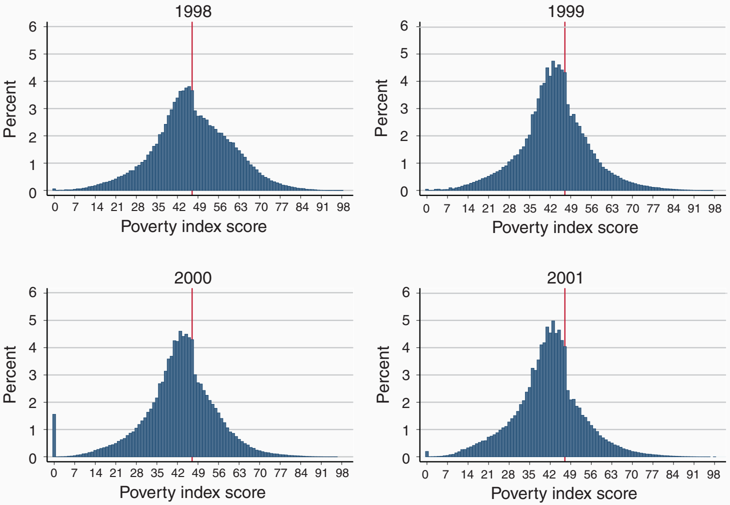
By Adriana Camacho and Emily Conover*

We document how manipulation of a targeting system for social welfare programs evolves over time. First, there was strategic behavior of some local politicians in the timing of the household interviews around local elections. Then, there was corrupt behavior with the sudden emergence of a sharp discontinuity in the score density, exactly at the eligibility threshold, which coincided with the release of the score algorithm to local officials. The discontinuity at the threshold is larger where mayoral elections are more competitive. While cultural forces are surely relevant for corruption, our results also highlight the importance of information and incentives. (*JEL* D72, I32, I38, O15, O17).

Visualizing manipulation — Proxy means test in Colombia



Visualizing manipulation — Proxy means test in Colombia



Visualizing manipulation — Proxy means test in Colombia

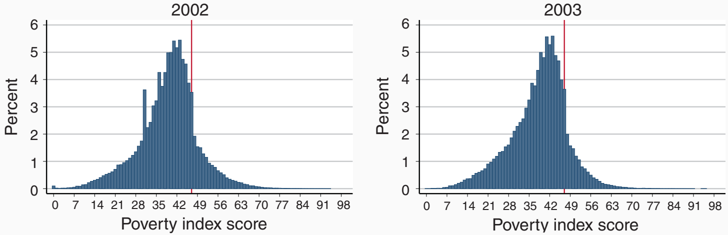


FIGURE 1. POVERTY INDEX SCORE DISTRIBUTION 1994–2003, ALGORITHM DISCLOSED IN 1997

Notes: Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.

Endogeneous cutoffs: Evaluating smoothness through balance

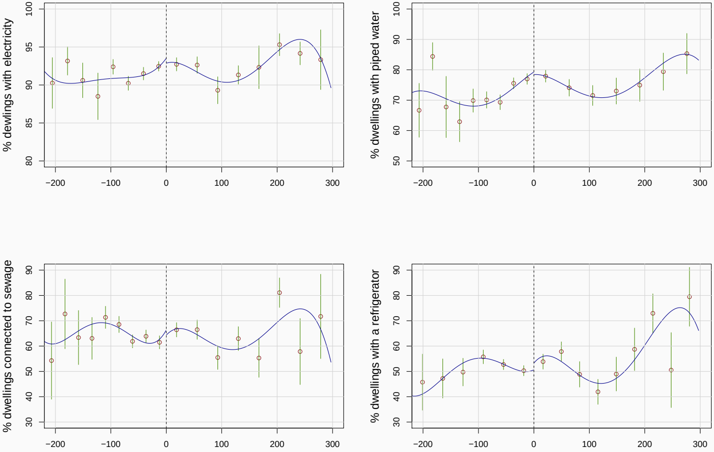
- Balance tests and placebo tests are related but distinct
- We can't directly test smoothness because we don't observe potential outcomes
- RD is like a "local RCT": Average values of exogenous covariates shouldn't jump around the cutoff
- Balance tests are indirect searching for evidence supporting smoothness

Balance implementation

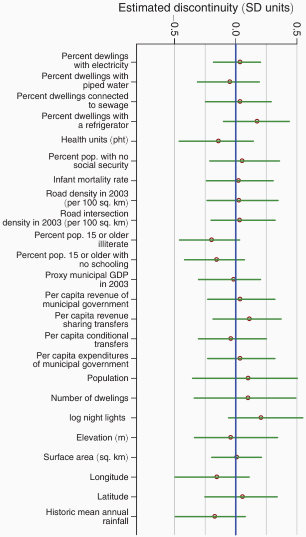
Don't make it hard – do what you did to Y, only to Z

- Choose other noncolliders associated with potential outcomes, Z
- Create similar graphical plots as you did for Y
- Could also conduct the parametric and nonparametric estimation on Z
- You do **not** want to see a jump around the cutoff, c_0

Balance – FONDEN running example



Balance – FONDEN running example



Placebos at non-discontinuous points

- Placebos in time are common with panels; placebo in running variables are their equivalent in RDD

- Imbens and Lemieux (2010) suggest we look at one side of the discontinuity (e.g., $X < c_0$), take the median value of the running variable in that section, and pretend it was a discontinuity, c_0'

- Then test whether in reality there is a discontinuity at c_0'. You do **not** want to find anything.

- Remember: smoothness at placebo points is neither necessary nor sufficient for smoothness in the potential outcomes at the cutoff

Balance – FONDEN running example

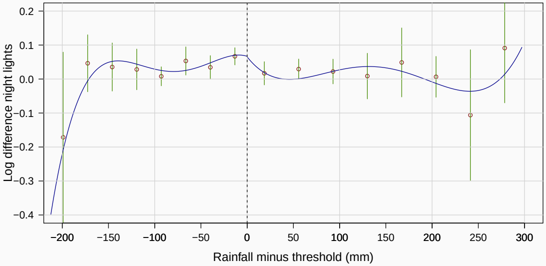


Figure A12: Intention-to-treat (placebo)

Note: The figure plots the log difference night lights, between two years before an event (months -24 to -13) and the year before (months -12 to -1), as a function of the running variable (rainfall minus threshold). The support of the running variable has been partitioned into disjoint bins. The number of bins is selected to minimize the integrated mean square error of the underlying regression function, as described in Calonico, Cattaneo and Titiunik (2015). The circles plot the local mean of the outcome at the mid-point of each bin. The error bars are the 95% confidence intervals for the local means. The solid lines are fourth-order global polynomials fits (estimated separately on each side of the threshold). Observations to the right of the vertical dashed line are eligible for Fonden under the heavy rainfall criteria.

Fuzzy design

Fuzzy RDD, IV and ITT

- Fuzzy RDD is an IV estimator, and requires those assumptions
- You may be more comfortable with presenting the intent-to-treat (ITT) parameter which is just the reduced form regression of Y on Z, therefore
- Many papers will not present an IV-style parameter, but rather a blizzard of ITT parameters, out of a "fear" that the exclusion restrictions may not hold
- But let's review the IV approach anyway for completeness (more IV to come!)

Probability of treatment jumps at discontinuity

Probabilistic treatment assignment (i.e. "fuzzy RDD")
The probability of receiving treatment changes discontinuously at the cutoff, c_0, but need not go from 0 to 1

$$lim_{X_i \rightarrow c_0} Pr(T_i = 1 | X_i = c_0) \neq lim_{c_0 \leftarrow X_i} Pr(T_i = 1 | X_i = c_0)$$

Examples: Incentives to participate in some program may change discontinuously at the cutoff but are not powerful enough to move everyone from non participation to participation.

Deterministic (sharp) vs. probabilistic (fuzzy)

- In the sharp RDD, T_i was *determined* by $X_i \geq c_0$
- In the fuzzy RDD, the *conditional probability* of treatment *jumps* at c_0.
- The relationship between the conditional probability of treatment and X_i can be written as:
$$P[T_i = 1|X_i] = g_0(X_i) + [g_1(X_i) - g_0(X_i)]Z_i$$
where $Z_i = 1$ if $(X_i \geq c_0)$ and 0 otherwise.

Instrumental variables

- As said, fuzzy designs are numerically equivalent and conceptually similar to IV (Instrument T with X and $X > c_0$)
  - "Reduced form" Numerator: "jump" in the regression of the outcome on the running variable, X.
  - "First stage" Denominator: "jump" in the regression of the treatment indicator on the running variable X.
- Same IV assumptions, caveats about compliers vs. defiers, and statistical tests that we discussed with instrumental variables apply here

Wald estimator

Wald estimator of treatment effect under Fuzzy RDD
Average causal effect of the treatment is the Wald IV parameter

$$\delta_{\text{Fuzzy RDD}} = \frac{lim_{X \rightarrow c_0} E[Y|X = c_0] - lim_{c_0 \leftarrow X} E[Y|X = c_0]}{lim_{X \rightarrow c_0} E[T|X = c_0] - lim_{c_0 \leftarrow X} E[T|X = c_0]}$$

Limitations of the LATE

- Fuzzy RDD has assumptions of all standard IV framework (exclusion, independence, nonzero first stage, and monotonicity)
- As with other binary IVs, the fuzzy RDD is estimating LATE: the local average treatment effect for the group of *compliers*
- In RDD, the compliers are those whose treatment status changed as we moved the value of x_i from just to the left of c_0 to just to the right of c_0

Balance – FONDEN running example

Next, we use local polynomial methods to estimate the first stage, the ITT, and the LATE. The specific estimating equations are as follows:

$$F_{mt} = \alpha_0 + \alpha_1 ABOVE_{mt} + g(R_{mt}) + v_{mt}, \tag{1}$$

$$Y_{mt} = \beta_0 + \beta_1 ABOVE_{mt} + g(R_{mt}) + \varepsilon_{mt}, \tag{2}$$

where F_{mt} is a binary variable that takes the value of one when a municipality is eligible for Fonden. The variable Y_{mt} represents our measure of the change in local economic activity (log difference night lights) for municipality m affected by a hydrometeorological event in year t. The variable $g(R_{mt})$ captures the relationship between the outcome and the running variable R_{mt}. The variable ABOVE is an indicator variable for observed rainfall exceeding the heavy rainfall threshold. Finally, ε_{mt} and v_{mt} are error terms. The parameters of interest are the first-stage estimate $\hat{\alpha}_1$ in equation (1), the ITT estimate $\hat{\beta}_1$ in equation (2), and the ratio $\tau_{FRD} = \hat{\beta}_1/\hat{\alpha}_1$ which can be interpreted as the LATE under some additional assumptions.[19]

Balance – FONDEN running example

TABLE 2—IMPACT OF FONDEN ON NIGHT LIGHTS

	(1)	(2)
Panel A. First stage (α_1)	0.227	0.230
p-value	<0.001	<0.001
CI 95 percent	[0.12, 0.28]	[0.13, 0.31]
Panel B. Intention-to-Treat (β_1)	0.059	0.072
p-value	0.010	0.006
CI 95 percent	[0.02, 0.12]	[0.02, 0.13]
Panel C. LATE (τ_{FRD})	0.260	0.313
p-value	0.009	0.011
CI 95 percent	[0.08, 0.56]	[0.08, 0.61]
Bandwidth (mm)	57.9	40.0
Observations (left | right)	1,038 | 525	741 | 410

Notes: Panel A presents estimates of equation (1), where the dependent variable is eligibility for Fonden resources. Panel B presents estimates of equation (2), where the dependent variable is the log difference in night lights between the 12 months before and after a disaster. Panel C reports the LATE estimate of eligibility for Fonden resources on night lights computed as the ratio of the ITT estimate to the first-stage coefficient. Estimates in panels A and B are derived using a triangular kernel and local linear polynomial. The bandwidth selection algorithm used in column 1 is optimal for point estimation; the selection algorithm in column 2 is optimal for inference of confidence intervals. The *p*-values and 95 percent confidence intervals reported are constructed using robust bias correction and clustering at the municipal level.

Visualization

Pictures, pictures and more pictures

- RDD is visually intense
- Eyeball tests are rampant (and deservedly) in RDD studies
- Let's review some of the graphs you have to include

Outcomes

1. **Outcome by running variable, (X_i):**

    - Construct bins and average the outcome within bins on both sides of the cutoff
    - Look at different bin sizes when constructing these graphs
    - Plot the running variables, X_i, on the horizontal axis and the average of Y_i for each bin on the vertical axis
    - Consider plotting a relatively flexible regression line on top of the bin means, but some readers prefer an eyeball test without the regression line to avoid "priming"

Probability of treatment

2. **Probability of treatment by running variable if fuzzy RDD**

   - In a fuzzy RDD, you also want to see that the treatment variable jumps at c_0
   - This tells you whether you have a first stage ("bite")
   - Let's look at that again from earlier Hoekstra (2008) and enrollment at the flagship

McCrary Density

3. **Density of the running variable**

- One should plot the number of observations in each bin.
- This plot allows to investigate whether there is a discontinuity or heaping in the distribution of the running variable at the threshold
- Heaping or discontinuities in the density suggest that people can manipulate their running variable score
- This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable, which may violate smoothness

Balance pictures

4. **Covariates by a running variable**

- Construct a similar graph to the outcomes graph but use a noncollider covariate as the "outcome"
- Balance implies smoothness through the cutoff, c_0.
- If noncollider covariates jump at the cutoff, one is probably justified to reject that potential outcomes aren't also probably jumping there