# ECON42720 Causal Inference and Policy Evaluation
## 6 Regression Discontinuity Designs

Ben Elsner (UCD)

# Resources

## Textbook chapters

- ▶ Cunningham, Causal Inference: The Mixtape, Ch. 6
- ▶ Huntington-Klein, The Effect: Ch. 20
- ▶ Huber, Causal Analysis Impact Evaluation and Causal Machine Learning with Applications in R, Ch.9

## Review Articles

- ▶ Cattaneo, Idrobo, Titiunik; A Practical Introduction to Regression Discontinuity Designs: Foundations ; In: Cambridge Elements: Quantitative and Computational Methods for Social Science.
- ▶ Cattaneo, Idrobo, Titiunik; A Practical Introduction to Regression Discontinuity Designs: Extensions ; In: Cambridge Elements: Quantitative and Computational Methods for Social Science.

# Regression Discontinuity: Starting Point

We want to **estimate a treatment effect**, but there is likely **selection bias**.

The required assumption

$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = 0$$

does not hold.

**RD exploits settings** where this assumption often holds

- ▶ arbitrary thresholds that determine treatment assignment
- ▶ typically regulatory thresholds
- ▶ Probability of treatment "jumps'' at the discontinuity

# Regression Discontinuity

## Examples for discontinuities

- Income thresholds for social benefits
- Cutoff rules for class sizes
- GPA thresholds for getting into college
- Special treatment for babies with $<1500$g birth weight
- ...

## Basic idea

- At the threshold, the **probability of treatment changes** sharply
- But **nothing else changes**
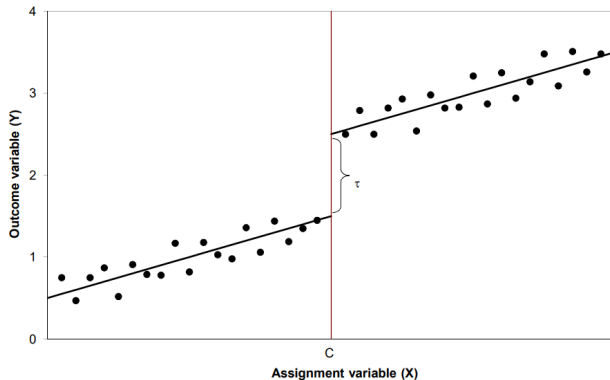- Being **above or below** the threshold is **as good as random**

# RD lingo

## The forcing variable $X$

▶ The **variable that determines treatment assignment**
▶ Also called assignment or running variable

## The discontinuity $X_0$

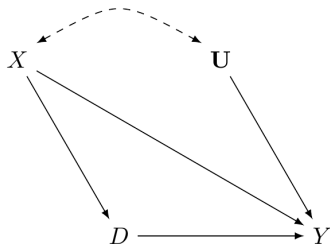▶ threshold value of the running variable at which treatment assignment jumps
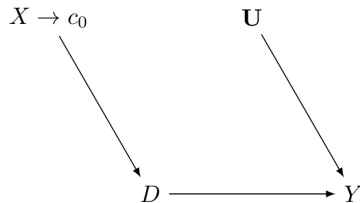
# Example of a linear RD



The aim is to estimate the **treatment effect** (here $\tau$) **at the discontinuity**

# RDD in a DAG



(A) Data generating graph    (B) Limiting graph

The **running variable $X$ determines the treatment $D$**
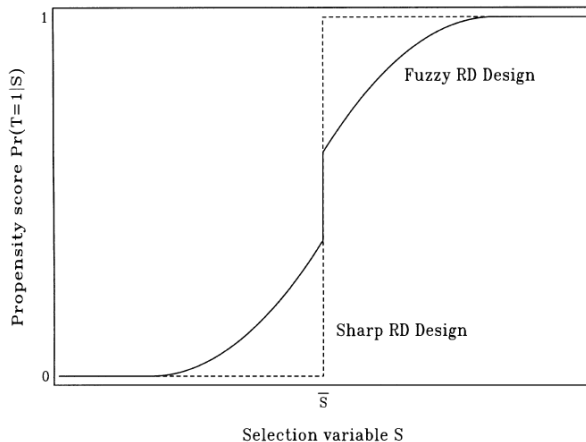
**Left:** $X$ may also affect the outcome through $U$

**Right:** Being above or below the threshold $c_0$ is as good as random

# Sharp and Fuzzy Regression Discontinuity

**Sharp RDD**: treatment probability jumps at $X_0$ from 0 to 1
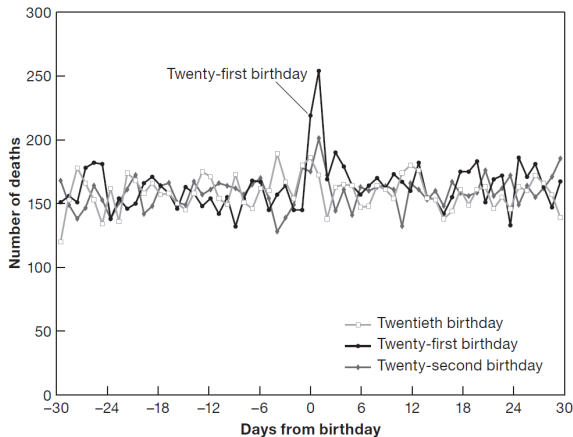
**Fuzzy RDD**: treatment probability jumps at $X_0$

# Example of a Sharp RDD: Carpenter & Dobkin (2009)

Observation: there is a **spike in deaths around the 21st birthday**



... but no difference around the 20th or 22nd birthday

# Example of a Sharp RDD: Carpenter & Dobkin (2009)

Carpenter & Dobkin (2009) investigate if this spike is **due to the legal drinking age** (21 in US)

Idea: at 21, nothing changes except that people can drink legally

Sharp RD: age is the running variable

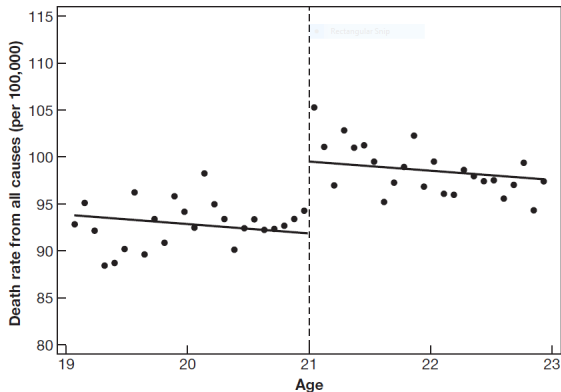$$D_a = \begin{cases} 1, & \text{if } a \geq 21. \\ 0, & \text{if } a < 21. \end{cases}$$

**Treatment status** is a **deterministic function of the running variable**

▶ if we know $a$, we know $D_a$

# Example of a Sharp RDD: Carpenter & Dobkin (2009)
## Simple RD analysis in a regression framework

$$\text{death rate}_a = \alpha + \rho D_a + \gamma a + e_a$$

# Example of a Sharp RDD: Carpenter & Dobkin (2009)

Does the **jump in the death rate $\rho$ represent a causal effect**?

**Yes if $D_a$ is solely determined by $a$**

- ▶ This is plausible in the given setting
- ▶ in this case there is no omitted variable bias
- ▶ no need to control for anything

**Advantage of RDs**: they are credible and transparent

**Downside of RDs**: they estimate local effects; difficult to extrapolate

# Example of a Fuzzy RDD: Hoekstra (2009)

**Question**: what is the impact of attending a State flagship university on earnings?

**Background:** many U.S. states have flagship universities

- ▶ Examples: University of Virginia, University of Florida, University of Michigan
- ▶ Is **going to University of Florida better** than going to **University of Southern Florida?**

**Challenge:** it is not random who goes to a particular university

# Example of a Fuzzy RDD: Hoekstra (2009)

**Identification strategy:** Hoekstra exploits admission cutoffs in a fuzzy RD design

- ▶ Applicants need a certain SAT score to be admitted
- ▶ The design is fuzzy because not everyone with a high SAT score applies and enrols

**Identification assumption:** students cannot manipulate whether they are above or below the cutoff

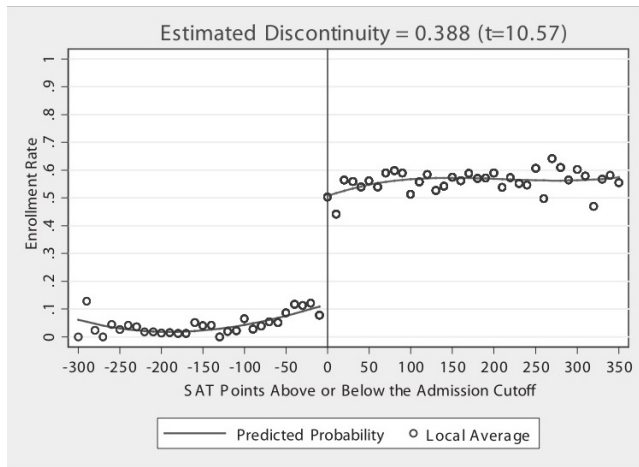# Data used in Hoekstra (2009)

**Administrative records** from a large **flagship university** (not disclosed which one)

- ▶ Includes data on all applicants, including their SAT scores

**Social Security Administration (SSA) data** on earnings

# First stage: SAT vs enrolment

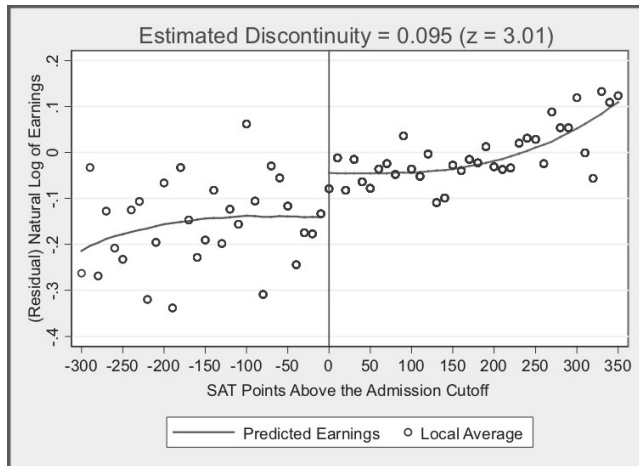Binscatter of SAT scores and enrolment rates

# Reduced form: SAT vs log earnings

Binscatter of SAT scores and enrolment rates

# Lessons from Hoekstra (2009)

Being **above the cutoff increases**

- ▶ the probability of enrolling by close to 40 percentage points
- ▶ earnings by around 10 percent

We can **calculate the LATE**: $\frac{0.1}{0.4} \times 100\% = 25\%$

# Sharp RD: Formal Derivation Angrist & Pischke (2009, ch. 6)

**Treatment status** $D_i$ is a deterministic function of $x_i$ with a **discontinuity** at $x_0$

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \\ 0 & \text{if } x_i < x_0 \end{cases}$$

Assume a **constant effects model**

$$E[Y_{0i}|x_i] = \alpha + \beta x_i$$
$$Y_{1i} = Y_{0i} + \rho$$

# Sharp RD: Formal Derivation (Angrist & Pischke 2009, Ch.6)

The **corresponding regression** is

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i$$

Of if the trend relation $E[Y_{0i}|x_i]$ is **non-linear**:

$$Y_i = \alpha + f(x_i) + \rho D_i + \eta_i$$

$f(x_i)$ modeled as a **p-th order polynomial**

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \rho D_i + \eta_i$$

# Sharp RD: Formal Derivation (Angrist & Pischke 2009, Ch.6)

Or allowing for **separate trend functions** for treated and untreated observations

$$E\left[Y_{0i}|x_i\right] = f_0\left(x_i\right) = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p$$
$$E\left[Y_{1i}|x_i\right] = f_1\left(x_i\right) = \alpha + \rho + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 + \ldots + \beta_{1p}\tilde{x}_i^p$$

with $\tilde{x}_i \equiv x_i - x_0$

Use the fact that $D_i$ is a deterministic function of $x_i$

$$E\left[Y_i|x_i\right] = E\left[Y_{0i}|x_i\right] + E\left[Y_{1i} - Y_{0i}|x_i\right]D_i$$

# Sharp RD: Formal Derivation (Angrist & Pischke 2009, Ch.6)

Substituting polynomials for conditional expectations yields the regression

$$Y_i = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p$$
$$+ \rho D_i + \beta_1^* D_i\tilde{x}_i + \beta_2^* D_i\tilde{x}_i^2 + \ldots + \beta_p^* D_i\tilde{x}_i^p + \eta_i$$

If we want to restrict the sample to a **bandwidth** $\delta$

$$E\left[Y_i | x_0 - \delta < x_i < x_0\right] \simeq E\left[Y_{0i} | x_i = x_0\right]$$
$$E\left[Y_i | x_0 < x_i < x_0 + \delta\right] \simeq E\left[Y_{1i} | x_i = x_0\right]$$

. . . the **estimand becomes**

$$\lim_{\delta \to 0} E\left[Y_i | x_0 < x_i < x_0 + \delta\right] - E\left[Y_i | x_0 - \delta < x_i < x_0\right] = E\left[Y_{1i} - Y_{0i} | x_i = x_0\right]$$

# Extrapolation in a Sharp RD
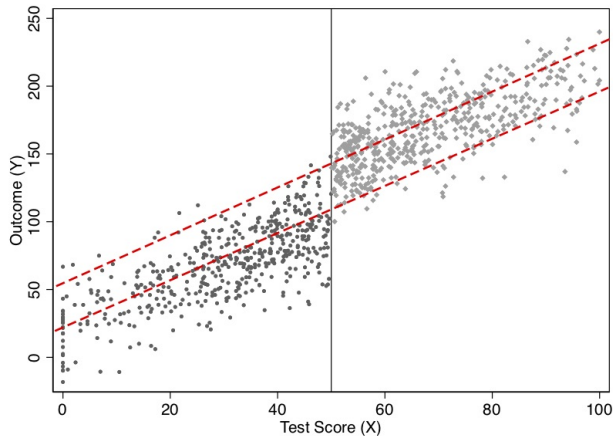


The RDD does not have common support

- ▶ Units above the cutoff are treated $D_i = 1$
- ▶ Units below the cutoff are untreated $D_i = 0$

# The Continuity Assumption

The **continuity assumption** is the key identifying assumption in the RD design

- It states that the potential outcomes are continuous functions of the running variable $X$
- $E[Y_i^0 \mid X = x_0] = E[Y_i^1 \mid X = x_0]$

**What does continuity mean?**

- potential outcomes don't jump at the cutoff
- there are no competing interventions
- people don't select into being above or below the cutoff

# RD: Importance of Functional Form

**RDs don't guarantee** the estimation of a **causal effect**
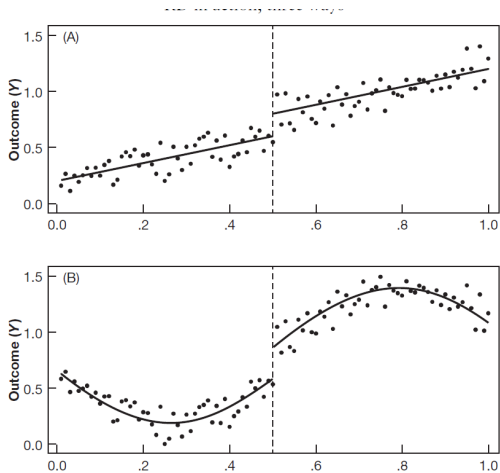
Problem: what looks like a discontinuous jump may actually be an **increase in a non-linear function**

It is important to

- **distinguish** between a **true causal effect** and an increase in a **non-linear trend**
- assess (and model) the **functional form** between the running variable and the outcome

# RD: Importance of Functional Form

Linear vs. quadratic function

# RD: Importance of Functional Form

Example of a spurious jump
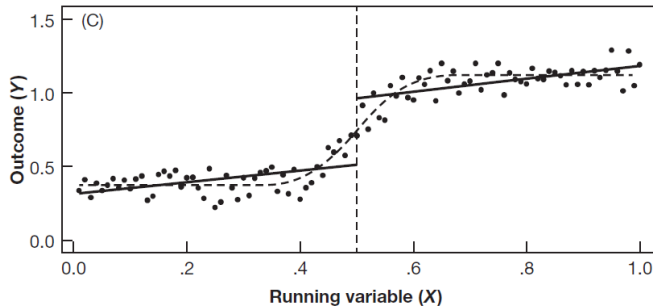


In an RD paper, it is important to show the **robustness of the results** to the choice of **functional form**

# RD: Importance of Functional Form

We can capture the curvature by including a quadratic in age

$$\text{death rate}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

Problem: often the **slope or curvature differs above and below the cutoff**

For example, below 21-year-olds are subject to minimum drinking age laws

# RD: Importance of Functional Form

Two measures solve this problem

- center the running variable around the cutoff (i.e. use $a - a_0$)
- add an interaction term $(a - a_0)D_a$

$$\text{death rate}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + e_a$$

This equation still identifies the effect at the cutoff ($a = a_0$)

# RD: Importance of Functional Form

It is also possible to fit a **polynomial on either side of the cut-off**

$$\text{death rate}_a = \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2$$
$$+ \delta_1[(a - a_0)D_a] + \delta_2[(a - a_0)^2 D_a] + e_a$$

# RD: Importance of Functional Form

Linear vs quadratic functional form in Carpenter & Dobkin (2009)
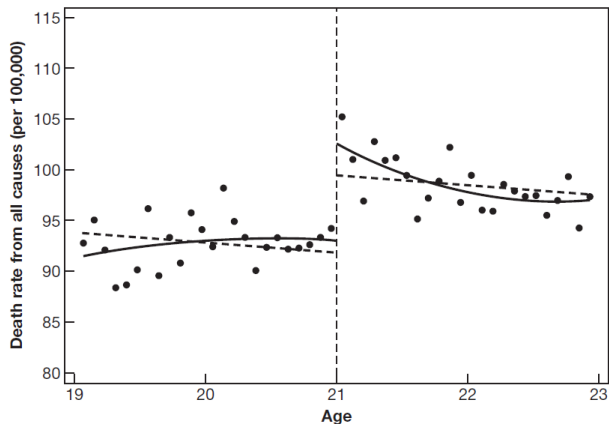


Treatment effect is larger with quadratic controls

# RD: Importance of Functional Form



(a) Order-2 Polynomial RDD

(b) Order-6 Polynomial RDD

(c) Order-25 Polynomial RDD

(d) Linear with Bandwidth
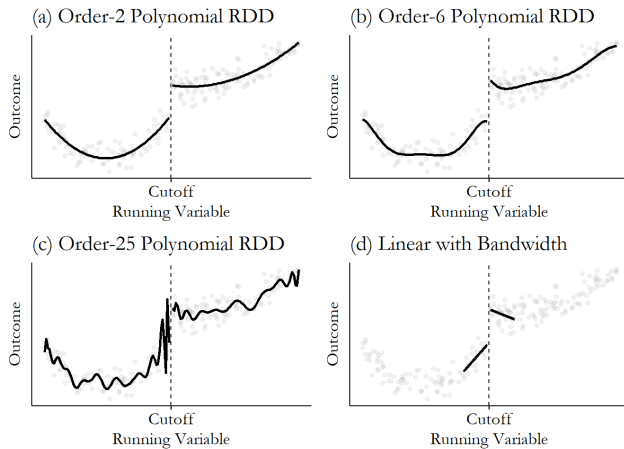
Source: Huntington-Klein, The Effect

# Overfitting? Crimes against Data?

**Gelman & Imbens (2019)**: polynomials can lead to **overfitting**



The estimated change in life expectancy (and height of the brace) just north of the Huai River is -5.04 years and is statistically significant (95% CI: -8.81, -1.27).
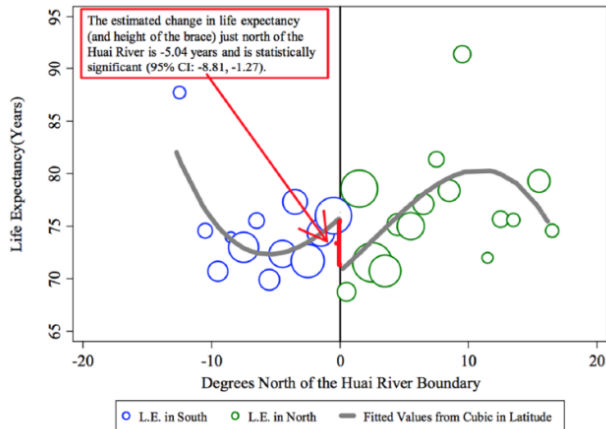
The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

# Overfitting? Crimes against Data?

**Overfitting**

▶ There is often **no scientific reason** to have high-order **polynomials**
▶ Overfitting: parameter estimates rely on **too few data points**
▶ **Large weights** are given to observations **far away from the discontinuity**
▶ Genuine **uncertainty** from model dependence is **not reflected in standard errors**

More on overfitting in RDs: Green *et al.* (2009), Gelman & Imbens (2019)

# Bandwidth Selection

One method to reduce the likelihood of spurious effects is to **narrow the bandwidth**

The bandwidth is the **"window" below and above the cutoff**

**Idea:**
- ▶ The closer we "zoom in'' on the cutoff
- ▶ ...the lower is the chance of picking up a trend

# Bandwidth Selection

**Trade-off in bandwidth selection**

- ▶ smaller bandwidth ⇒ **smaller bias**
- ▶ smaller bandwidth ⇒ **less precision**

The graph on the following page illustrates **two common methods**

- ▶ (non-parametric) kernel density estimation
- ▶ local linear regressions

**Optimal bandwidth selection** is a very active area of research in econometrics. See, for example, Imbens & Kalyanaraman (2012).

# Bandwidth Selection
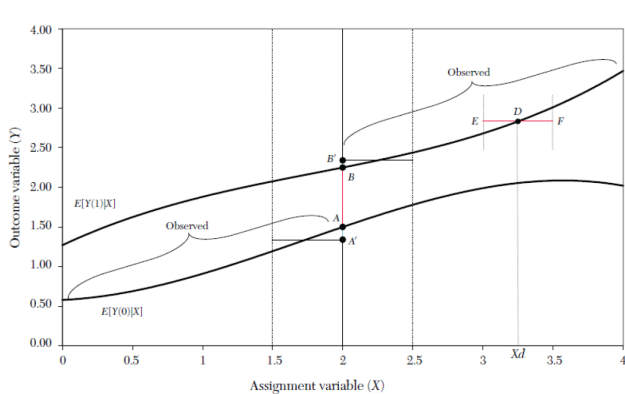
# Bandwidth Selection: Practical Advice

In a typical RDD paper you will see a **variety of bandwidths**

The preferred bandwidth in many studies is the **optimal bandwidth from Calonico et al. (2014)**

- ▶ It trades off bias and variance
- ▶ the `rdrobust` package in R implements this method

# Local Regression

Simplest form: **compare means above and below** within bandwidth

▶ Problem: differential trends in $Y$ above and below can lead to bias

**Alternatives:**

▶ Local linear regression
▶ Kernel regression

# Local Linear Regression

Simple **local linear regression restricts the sample to** $a_0 - b \leq a \leq a_0 + b$

And estimates a linear regression in this window:

$$\text{death rate}_a = \alpha + \rho D_a + \gamma a + e_a$$

Or, more commonly, we can **allow for different slopes above and below the cut-off**

# Kernel Regression

**Assign weights** that

- ▶ are maximized at the cutoff
- ▶ are zero outside the bandwidth

Example for **triangular kernel**



Source: Huntington-Klein, The Effect
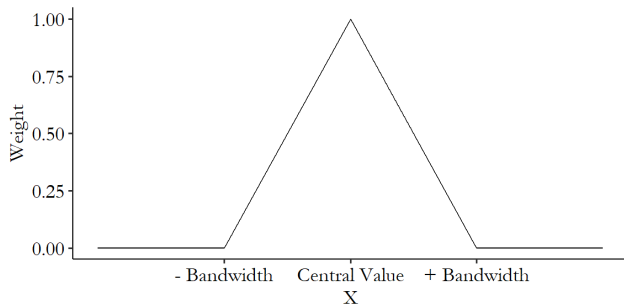
# Fuzzy RDD

In a fuzzy RDD, the **probability of treatment jumps at the cutoff**

$$P[D_i = 1|x_i] = \begin{cases} g_1(x_i), & \text{if } x_i \geq x_0. \\ g_0(x_i), & \text{if } x_i < x_0. \end{cases}$$

where $g_1(x_0) \neq g_0(x_i)$

This set-up is **equivalent to an IV estimator**

- The discontinuity is the instrument for the treatment
- If we control for the forcing variable, the assignment of the IV is as good as random

# Fuzzy RDD

**Simplest case:** let $T_i$ be the discontinuity and $D_i$ be the treatment

The (hypothetical) **first stage** is

$$D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 T_i + e_i$$

But because we often don't observe the treatment, we **estimate the reduced form**

$$y_i = \alpha + \beta T_i + \delta X_i + u_i$$

# Fuzzy RDD

**Wald Estimator** with a bandwidth of $\delta$

$$\lim_{\delta \to 0} \frac{E\left[Y_i | x_0 < x_i < x_0 + \delta\right] - E\left[Y_i | x_0 - \delta < x_i < x_0\right]}{E\left[D_i | x_0 < x_i < x_0 + \delta\right] - E\left[D_i | x_0 - \delta < x_i < x_0\right]} = \rho$$

# Fuzzy RDD

Fuzzy RD becomes **more tricky with interactions** (treatment with forcing variable)

$\Rightarrow$ need a **separate IV and first stage** for each term including $D_i$

This means that one has as **many instruments** as there are terms including $D_i$

Rather than deriving this, we will look at an example: Angrist & Lavy (1999)

# Fuzzy RD: Angrist & Lavy (1999)

**Angrist & Lavy (1999)** study the **impact of class sizes on student achievement**

They exploit that **class sizes in Israeli schools follow Maimonides' rule**

- ▶ Class size is capped at 40
- ▶ If enrollment reaches 41, two classes are formed
- ▶ Three classes are formed if the enrollment reaches 80, etc

# Fuzzy RD: Angrist & Lavy (1999)



a. Fifth Grade

b. Fourth Grade
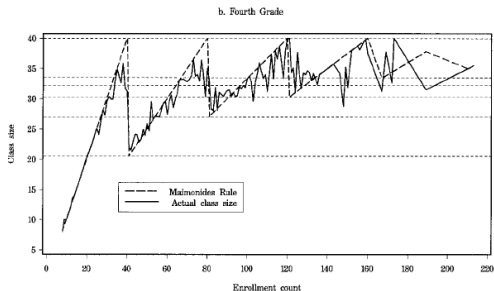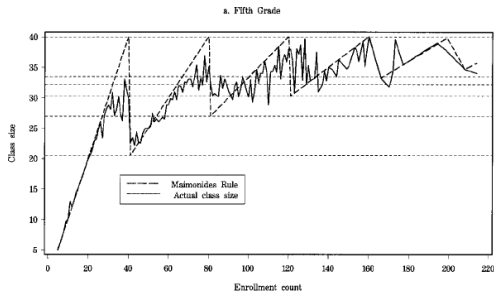
# Fuzzy RD: Angrist & Lavy (1999)

Angrist & Lavy (1999) use the **predicted class size as instrument for the actual class size**
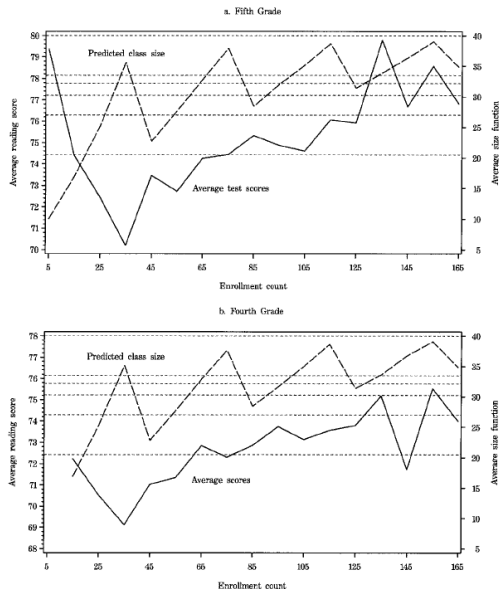
Prediction is based on a mathematical formula (namely Maimonides' rule)

Not all **schools fully comply, but most do**

This is a **classic example of a fuzzy RD**

# Fuzzy RD: Angrist & Lavy (1999)



a. Fifth Grade

b. Fourth Grade

# Fuzzy RD: Angrist & Lavy (1999)

TABLE II
OLS Estimates for 1991

| | 5th Grade | | | | | | 4th Grade | | | | | |
| | Reading comprehension | | | Math | | | Reading comprehension | | | Math | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Mean score* | | 74.3 | | | 67.3 | | | 72.5 | | | 69.9 | |
| *(s.d.)* | | (8.1) | | | (9.9) | | | (8.0) | | | (8.8) | |
| *Regressors* | | | | | | | | | | | | |
| Class size | .221 | −.031 | −.025 | .322 | .076 | .019 | 0.141 | −.053 | −.040 | .221 | .055 | .009 |
| | (.031) | (.026) | (.031) | (.039) | (.036) | (.044) | (.033) | (.028) | (.033) | (.036) | (.033) | (.039) |
| Percent disadvantaged | | −.350 | −.351 | | −.340 | −.332 | | −.339 | −.341 | | −.289 | −.281 |
| | | (.012) | (.013) | | (.018) | (.018) | | (.013) | (.014) | | (.016) | (.016) |
| Enrollment | | | −.002 | | | .017 | | | −.004 | | | .014 |
| | | | (.006) | | | (.009) | | | (.007) | | | (.008) |
| Root MSE | 7.54 | 6.10 | 6.10 | 9.36 | 8.32 | 8.30 | 7.94 | 6.65 | 6.65 | 8.66 | 7.82 | 7.81 |
| $R^2$ | .036 | .369 | .369 | .048 | .249 | .252 | .013 | .309 | .309 | .025 | .204 | .207 |
| N | | 2,019 | | | 2,018 | | | 2,049 | | | 2,049 | |

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes.

# Fuzzy RD: Angrist & Lavy (1999)

TABLE IV
2SLS ESTIMATES FOR 1991 (FIFTH GRADERS)

| | Reading comprehension | | | | | | Math | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Full sample | | | | +/− 5 Discontinuity sample | | Full sample | | | | +/− 5 Discontinuity sample | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| *Mean score* | | 74.4 | | | 74.5 | | | 67.3 | | | 67.0 | |
| *(s.d.)* | | (7.7) | | | (8.2) | | | (9.6) | | | (10.2) | |
| *Regressors* | | | | | | | | | | | | |
| Class size | −.158 | −.275 | −.260 | −.186 | −.410 | −.582 | −.013 | −.230 | −.261 | −.202 | −.185 | −.443 |
| | (.040) | (.066) | (.081) | (.104) | (.113) | (.181) | (.056) | (.092) | (.113) | (.131) | (.151) | (.236) |
| Percent disadvantaged | −.372 | −.369 | −.369 | | −.477 | −.461 | −.355 | −.350 | −.350 | | −.459 | −.435 |
| | (.014) | (.014) | (.013) | | (.037) | (.037) | (.019) | (.019) | (.019) | | (.049) | (.049) |
| Enrollment | | .022 | .012 | | | .053 | | .041 | .062 | | | .079 |
| | | (.009) | (.026) | | | (.028) | | (.012) | (.037) | | | (.036) |
| Enrollment squared/100 | | | .005 | | | | | | −.010 | | | |
| | | | (.011) | | | | | | (.016) | | | |
| Piecewise linear trend | | | | .136 | | | | | | .193 | | |
| | | | | (.032) | | | | | | (.040) | | |
| Root MSE | 6.15 | 6.23 | 6.22 | 7.71 | 6.79 | 7.15 | 8.34 | 8.40 | 8.42 | 9.49 | 8.79 | 9.10 |
| N | | 2019 | | 1961 | 471 | | | 2018 | | 1960 | 471 | |

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use $f_{sc}$ as an instrument for class size.

# Fuzzy RD: Angrist & Lavy (1999)

- Table II shows **OLS estimates**: There is a positive correlation between class size and test scores in the raw data. This correlation vanishes when the fraction of disadvantaged students is controlled for.
- Table IV shows the **IV results**, exploiting the regression discontinuities created by Maimonides' rule. The table displays various specifications with no, linear, quadratic, and piecewise linear controls for enrollment, as well as estimates in subsamples around the discontinuity points.

# Fuzzy RD: Angrist & Lavy (1999)

- **Controlling for enrollment is important**, particularly for the math test scores. The form of the control matters less.
- On the other hand, the **discontinuity samples give larger effects** (in absolute values) than the full sample, which is less comforting.
- Overall, they find that the IV estimates are larger than the OLS estimates.
- The **downward bias of OLS** is plausible as it may be the case that poorer performing students are placed in smaller classes.

# Regression Discontinuity: Comments

**Regression Discontinuity** has become one of **the most popular methods of causal inference**

Some reasons:

- ▶ It's easy to explain to non-economists
- ▶ The researcher is forced to show patterns in the data
- ▶ Identification assumptions can be inspected graphically
- ▶ It is often clear what drives the variation in the treatment

For useful practitioners' guides, see Matias Cattaneo (Michigan) and David Lee (Princeton)
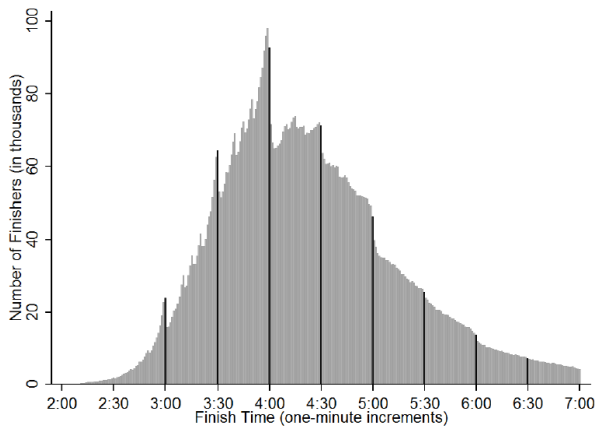
# Challenge to Identification

As with any method of causal inference, RD rests on an **untestable identification assumption**

- ▶ being **above or below the cutoff is as good as random**

This may not be true if there is **manipulation**

- ▶ people may be able to choose whether they are above or below the cutoff
- ▶ teachers may grade people up, etc

# Example for Manipulation

# What Can/Should You Do?

There are **non-parametric tests** for heaping/bunching

- ▶ McCrary density test (McCrary, 2008)

**Run placebo tests** based on pre-treatment characteristics

- ▶ there should be no jump at the discontinuity
- ▶ if there is, that's a problem

**One solution**: a donut hole estimator (leave out points close to the discontinuity)

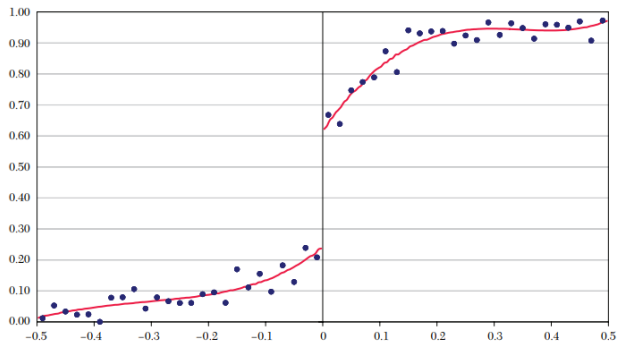# RD: The Cookbook

**3 General Rules: plot, plot, plot!**

**1) Explain the Identification Strategy**

- ▶ why is there a discontinuity?
- ▶ and what is the treatment that changes?
- ▶ what is the scope for manipulation?

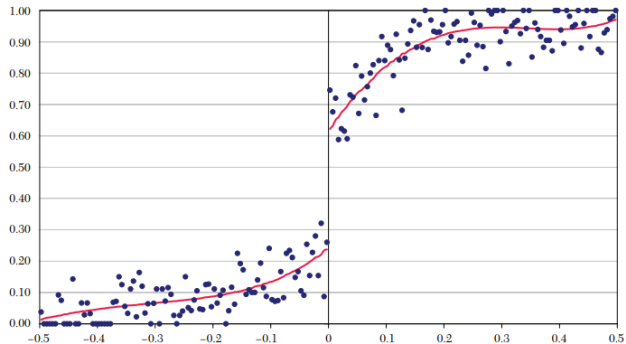**2) Produce and discuss the main graph**

- ▶ Outcome plotted against the running variable
- ▶ Best to use binned scatters
- ▶ Important to find the right functional form
- ▶ Use practitioners' guides, for example Lee & Lemieux (2010)

# RD: The Cookbook



{Source: Lee & Lemieux (2010)}

# RD: The Cookbook
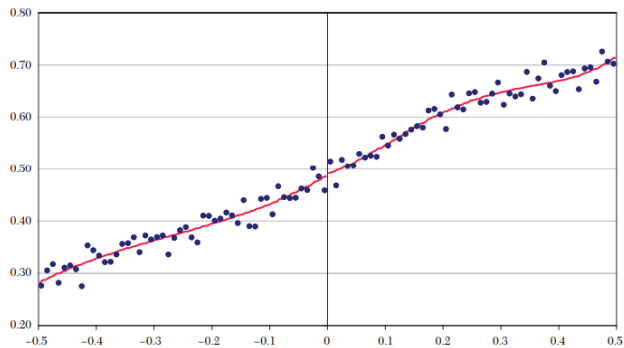


{Source: Lee & Lemieux (2010)}

# RD: The Cookbook

**3) Report estimates based on different methods**

- ▶ careful when using polynomials
- ▶ local linear regression
- ▶ kernel methods

**4) Density and placebo tests**

- ▶ Inspect if there is heaping at the discontinuity
- ▶ Run a McCrary density test
- ▶ Plot pre-treatment characteristics against the running variable
- ▶ Check out the latest literature. If you don't run the latest tests, your referees will ask you to (If you're lucky). . .

# RD: The Cookbook



{Source: Lee & Lemieux (2010)}

# Appendix

# RDD in R

The following slides show how an RDD can be implemented with R. Most of this code is from Cunningham's Mixtape and Huntington-Klein's Effect.

You need the following packages: `tidyverse`, `stargazer`, `causaldata`, and `rdrobust` and `rddensity`.

We first simulate a dataset and then show how we run regressions.

# RDD in R – Simulated Dataset

```r
# simulate the data
dat <- tibble(
  x = rnorm(1000, 50, 25)
) %>%
  mutate(
    x = if_else(x < 0, 0, x)
  ) %>%
  filter(x < 100)

# cutoff at x = 50
dat <- dat %>%
  mutate(
    D  = if_else(x > 50, 1, 0),
    y1 = 25 + 0 * D + 1.5 * x + rnorm(n(), 0, 20)
  )
```

# RDD in R – RDD plot with gpplot

```r
# Simulate data with discontinuity
dat <- dat %>%
  mutate(
    y2 = 25 + 40 * D + 1.5 * x + rnorm(n(), 0, 20)
  )

# Plot
ggplot(aes(x, y2, colour = factor(D)), data = dat) +
  geom_point(alpha = 0.5) +
  geom_vline(xintercept = 50, colour = "grey", linetype = 2) +
  stat_smooth(method = "lm", se = F) +
  theme_minimal() +
  theme(text = element_text(size = 20), legend.position = "none") +
  labs(x = "Test score (X)", y = "Potential Outcome (Y)")
```

The code from the previous slide gives this
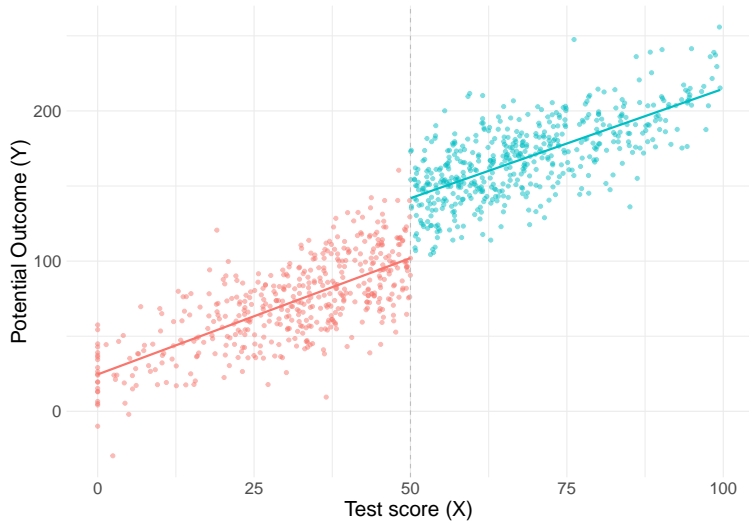
# RDD Regressions in R

The regressions are based on the package `causaldata`, which has data on a government transfer that was administered based on an income cutoff. The original study was done by Manacorda et al (2011).

# RDD Regressions in R

Data preparation next slide

- ▶ binned data for bin scatter
- ▶ add kernel weights

# RDD Regressions in R

```r
gt <- causaldata::gov_transfers

# Use cut() to create bins, using breaks to make sure it breaks at 0
# (-15:15)*.02/15 gives 15 breaks from -.02 to .02
binned <- gt %>%
    mutate(Inc_Bins = cut(Income_Centered,
            breaks = (-15:15)*(.02/15))) %>%
    group_by(Inc_Bins) %>%
    summarize(Support = mean(Support),
    Income = mean(Income_Centered))

# Add a triangular kernel weight
kweight <- function(x) {
    # To start at a weight of 0 at x = 0, and impose a bandwidth of .01,
    # we need a "slope" of -1/.01 = 100,
    # and to go in either direction use the absolute value
    w <- 1 - 100*abs(x)
    # if further away than .01, the weight is 0, not negative
    w <- ifelse(w < 0, 0, w)
    return(w)
}

gt <- gt |> mutate(kweight=kweight(Income_Centered))
```
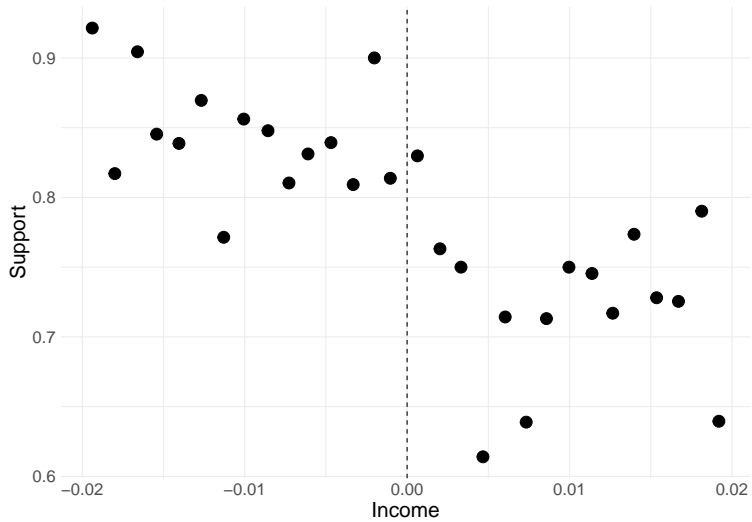
# RDD Regressions in R

Binned scatter

```r
# Taking the mean of Income lets us plot
# data roughly at the bin midpoints

ggplot(binned, aes(x = Income, y = Support)) +
    geom_point(size=5) +
    theme_minimal() +
    theme(text = element_text(size = 20)) +
    # Add a cutoff line
    geom_vline(aes(xintercept = 0), linetype = 'dashed')
```

# RDD Regressions in R

```r
# Reg 1: linear term
  m1 <- lm(Support ~ Income_Centered*Participation, data = gt)

# Reg 2: model with square
  m2 <- lm(Support ~ Income_Centered*Participation +
      I(Income_Centered^2)*Participation, data = gt)

# Reg 3: model with Kernel weights
 m3 <- lm(Support ~ Income_Centered*Participation, data = gt,
        weights = kweight)

# Regression results:


stargazer(m1, m2, m3, header=FALSE, type='latex', font.size="tiny",
          single.row=TRUE)
```

Table 1:

| | Dependent variable: | | |
|---|---|---|---|
| | Support | | |
| | (1) | (2) | (3) |
| Income_Centered | −0.179 (1.916) | −11.567 (7.777) | −23.697*** (6.430) |
| Participation | 0.100*** (0.030) | 0.093** (0.046) | 0.033 (0.037) |
| I(Income_Centered^2) | | 562.247 (372.182) | |
| Income_Centered:Participation | −1.442 (2.516) | 19.300* (10.445) | 26.594*** (8.916) |
| Participation:I(Income_Centered^2) | | −101.103 (500.196) | |
| Constant | 0.730*** (0.023) | 0.769*** (0.034) | 0.819*** (0.027) |
| Observations | 1,948 | 1,948 | 1,948 |
| $R^2$ | 0.034 | 0.036 | 0.041 |
| Adjusted $R^2$ | 0.033 | 0.034 | 0.038 |
| Residual Std. Error | 0.313 (df = 1944) | 0.313 (df = 1942) | 0.217 (df = 933) |
| F Statistic | 22.954*** (df = 3; 1944) | 14.625*** (df = 5; 1942) | 13.457*** (df = 3; 933) |

Note: *p<0.1; **p<0.05; ***p<0.01

# RDD Regressions in R

On the previous slide we can see that

▶ The estimated treatment effect is large when we fit linear or quadratic regressions (9.3pp, statistically significant)
▶ But it is small when we fit a triangular Kernel regression

Note: the Kernel regression uses a lot fewer observations (bias-variance trade-off...)

# Using `rdrobust`

`rdrobust` is a very useful package because it has many useful in-built functions

- ▶ generates **RD plots**
- ▶ allows for **optimal bandwidth selection**
- ▶ allows for **fuzzy RDD**

# Using `rdrobust`

```r
# Estimate regression discontinuity and plot it
m <- rdrobust(gt$Support, gt$Income_Centered, c = 0, p=1)
summary(m)
```

```r
# Note, by default, rdrobust and rdplot use different numbers
# of polynomial terms. You can set the p option to standardize them.
rdplot(gt$Support, gt$Income_Centered, masspoints="off", p=1)
```

Nice feature: can change the order of polynomial through the p option

Not so nice feature: hard to change the layout of the plot

RD Plot
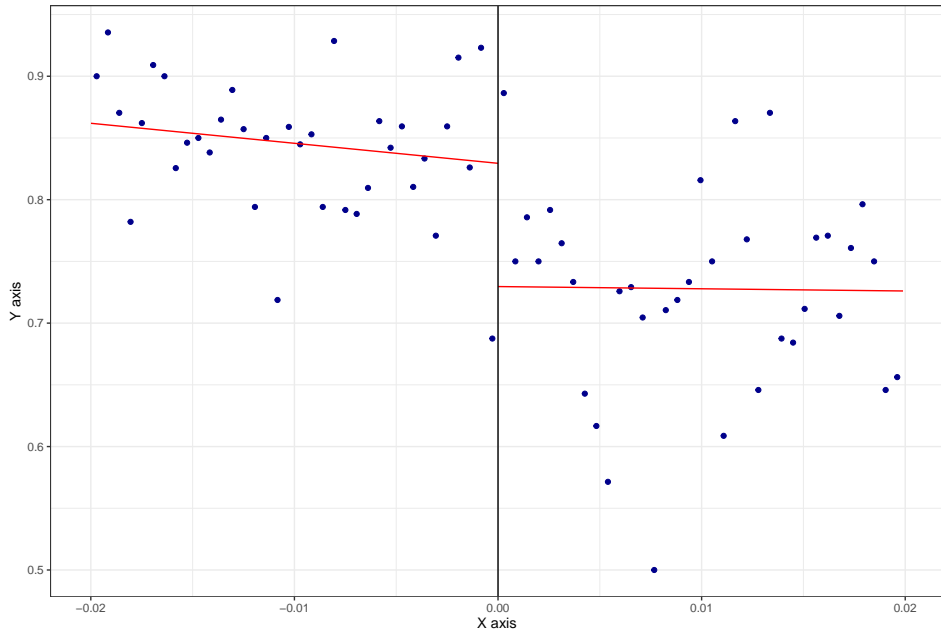
# Density test with `rddensity`

```r
gt <- causaldata::gov_transfers_density  %>%
    filter(abs(Income_Centered) < .02)

# Estimate the discontinuity
gt %>%
    pull(Income_Centered) %>%
    rddensity(c = 0) %>%
    summary()
```

# References

Angrist, Joshua, & Lavy, Victor. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, **114**(2), 533–575.

Angrist, Joshua, & Pischke, Jörn-Steffen. 2009. *Mostly Harmless Econometrics - An Empiricist's Companion*. Princeton University Press.

Calonico, Sebastian, Cattaneo, Matias D., & Titiunik, Rocio. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, **82**(6), 2295–2326.

Carpenter, Christopher, & Dobkin, Carlos. 2009. The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age. *American Economic Journal: Applied Economics*, **1**(1), 164–82.

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

Green, Donald P., Leong, Terence Y., Kern, Holger L., Gerber, Alan S., & Larimer, Christopher W. 2009. Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks. *Political Analysis*, **17**(4), 400–417.

Hoekstra, Mark. 2009. THE EFFECT OF ATTENDING THE FLAGSHIP STATE UNIVERSITY ON EARNINGS: A DISCONTINUITY-BASED APPROACH. *The Review of Economics and Statistics*, **91**(4), 717–724.

Imbens, Guido, & Kalyanaraman, Karthik. 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, **79**(01), 933–959.

Lee, David S., & Lemieux, Thomas. 2010. Regression Discontinuity Design in Economics. **48**(06), 281–355.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142**(2), 698–714. The regression discontinuity design: Theory and applications.

benjamin.elsner@ucd.ie

www.benjaminelsner.com

Sign up for office hours

YouTube Channel

@ben_elsner

LinkedIn

# Contact

**Prof. Benjamin Elsner**
University College Dublin
School of Economics
Newman Building, Office G206
benjamin.elsner@ucd.ie

Office hours by appointment. Please email me.