Causal Inference
5 - Regression Discontinuity and Kink Designs

Benjamin Elsner
benjamin.elsner@ucd.ie

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 <1500g 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

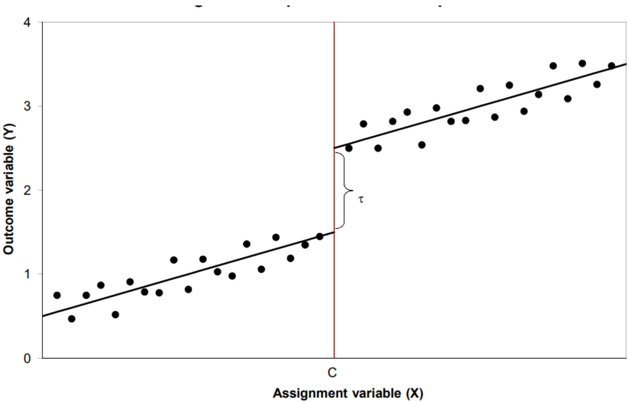
The forcing variable X

- 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

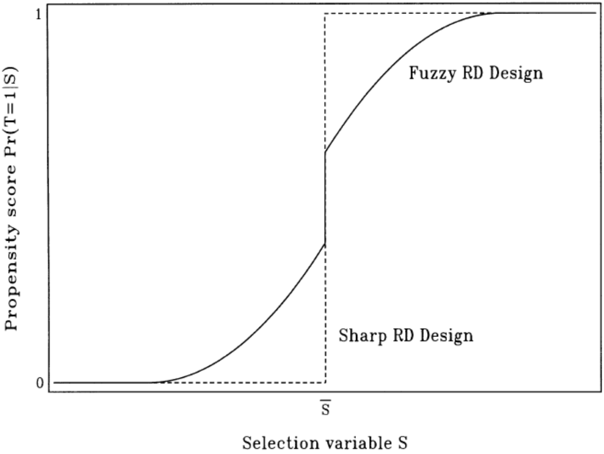


The aim is to estimate the **treatment effect** (here τ) **at the discontinuity**

Sharp and Fuzzy Regression Discontinuity

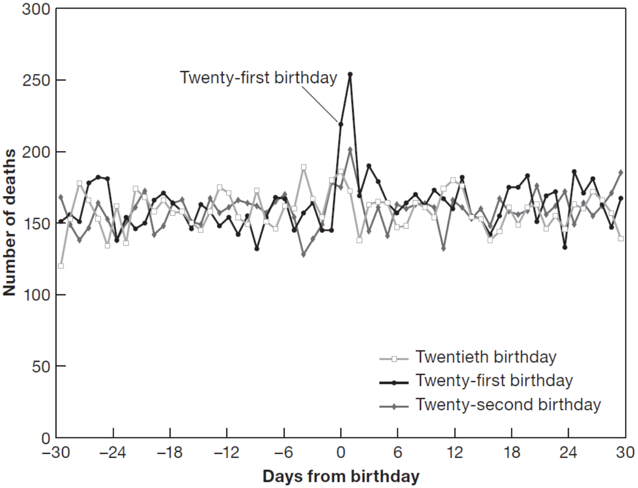
Sharp RDD: treatment probability jumps at X_0 from 0 to 1

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

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

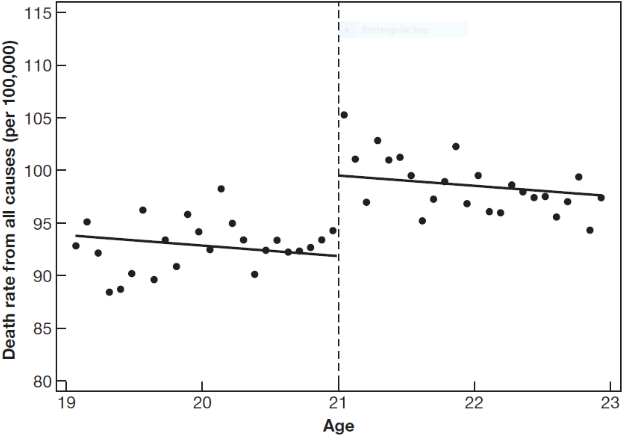
Treatment status is a **deterministic function of the running variable**

- 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 ρ 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

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

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

Assume a **constant effects model**

$$\begin{aligned} E\left[\mathrm{Y}_{0i} | x_i\right] &= \alpha + \beta x_i \\ \mathrm{Y}_{1i} &= \mathrm{Y}_{0i} + \rho \end{aligned}$$

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

The **corresponding regression** is

$$\mathrm{Y}_i = \alpha + \beta x_i + \rho \mathrm{D}_i + \eta_i$$

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

$$\mathrm{Y}_i = \alpha + f(x_i) + \rho \mathrm{D}_i + \eta_i$$

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

$$\mathrm{Y}_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \rho \mathrm{D}_i + \eta_i$$

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

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

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

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

Use the fact that D_i is a deterministic function of x_i

$$E\left[\mathrm{Y}_i|x_i\right] = E\left[\mathrm{Y}_{0i}|x_i\right] + E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}|x_i\right]\mathrm{D}_i$$

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

Substituting polynomials for conditional expectations yields the regression

$$
\begin{aligned}
\mathrm{Y}_i =& \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p \\
& + \rho \mathrm{D}_i + \beta_1^* \mathrm{D}_i \tilde{x}_i + \beta_2^* \mathrm{D}_i \tilde{x}_i^2 + \ldots + \beta_p^* \mathrm{D}_i \tilde{x}_i^p + \eta_i
\end{aligned}
$$

If we want to restrict the sample to a **bandwidth** δ

$$
\begin{aligned}
E\left[\mathrm{Y}_i | x_0 - \delta < x_i < x_0\right] &\simeq E\left[\mathrm{Y}_{0i} | x_i = x_0\right] \\
E\left[\mathrm{Y}_i | x_0 < x_i < x_0 + \delta\right] &\simeq E\left[\mathrm{Y}_{1i} | x_i = x_0\right]
\end{aligned}
$$

...the **estimate becomes**

$$
\lim_{\delta \to 0} E\left[\mathrm{Y}_i | x_0 < x_i < x_0 + \delta\right] - E\left[\mathrm{Y}_i | x_0 - \delta < x_i < x_0\right] = E\left[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | x_i = x_0\right]
$$

RD: Importance of Functional Form

RDs don't guarantee the estimation of a **causal effect**

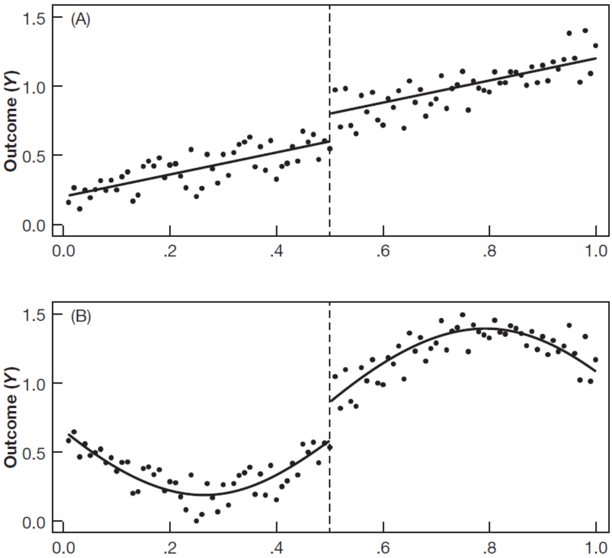
Problem: what looks like a discontinuous jump may actually be an **increase in a non-linear function**

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

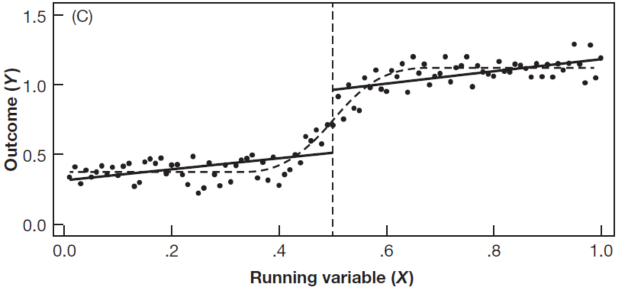
RD: Importance of Functional Form

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 ($a = a_0$)

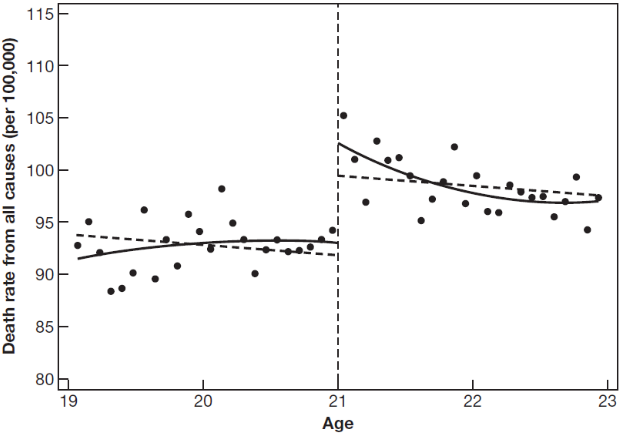
RD: Importance of Functional Form

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

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

RD: Importance of Functional Form

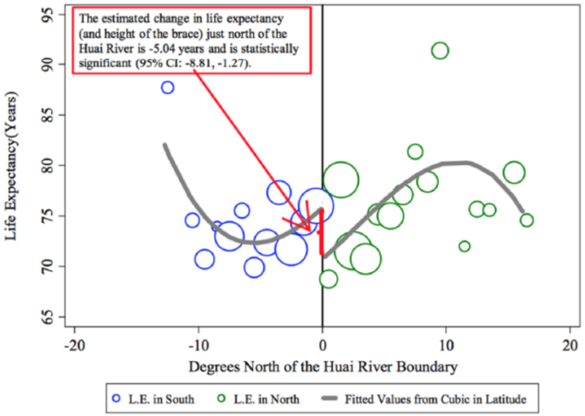
Linear vs quadratic functional form in Carpenter & Dobkin (2009)



Treatment effect is larger with quadratic controls

Overfitting? Crimes against Data?

Gelman & Zelizer (2015): polynomials can lead to **overfitting**



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

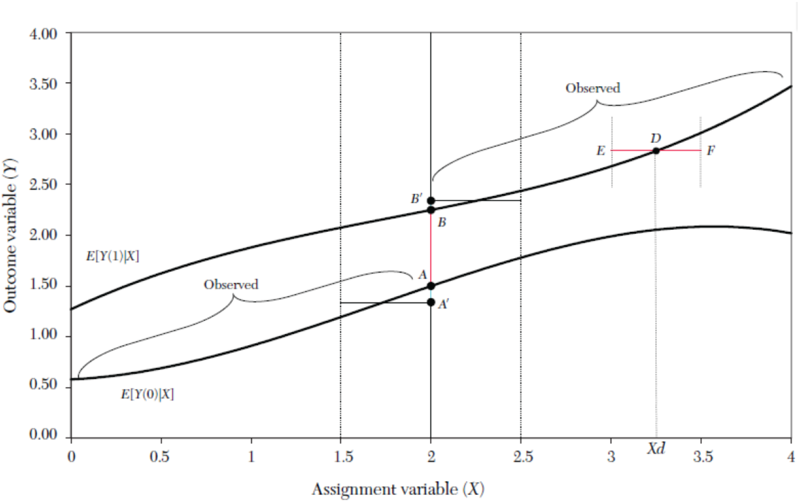
- smaller bandwidth ⇒ **smaller bias**
- smaller bandwidth ⇒ **less precision**

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

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**

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 δ

$$\lim_{\delta \to 0} \frac{E\left[\mathrm{Y}_i | x_0 < x_i < x_0 + \delta\right] - E\left[\mathrm{Y}_i | x_0 - \delta < x_i < x_0\right]}{E\left[\mathrm{D}_i | x_0 < x_i < x_0 + \delta\right] - E\left[\mathrm{D}_i | x_0 - \delta < x_i < x_0\right]} = \rho$$

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)

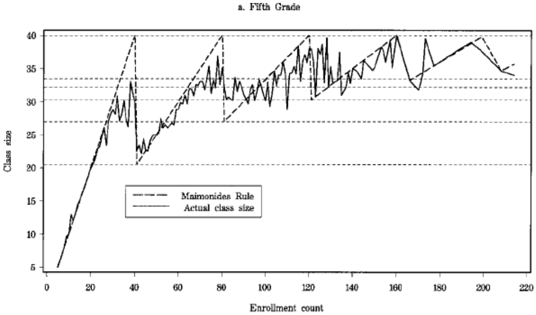
Fuzzy RD: Angrist & Lavy (1999)

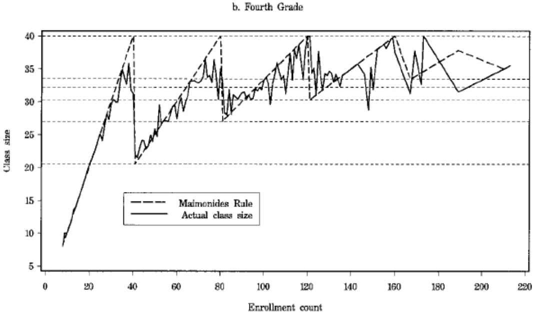
Angrist & Lavy (1999) study the **impact of class sizes on student achievement**

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)





Fuzzy RD: Angrist & Lavy (1999)

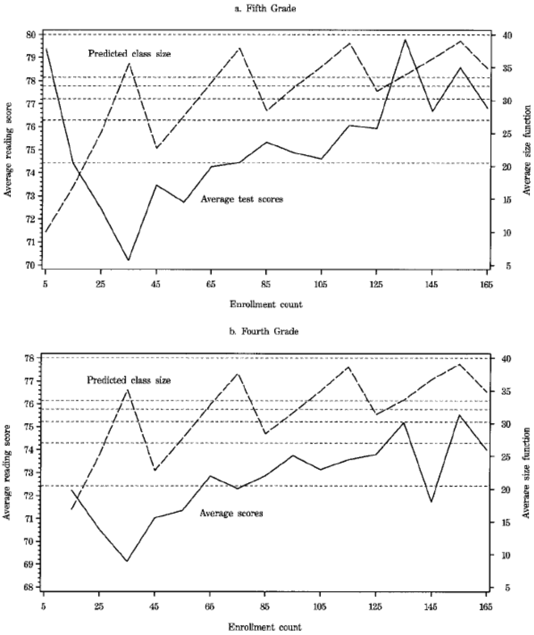
Angrist & Lavy (1999) use the **predicted class size as instrument for the actual class size**

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)



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

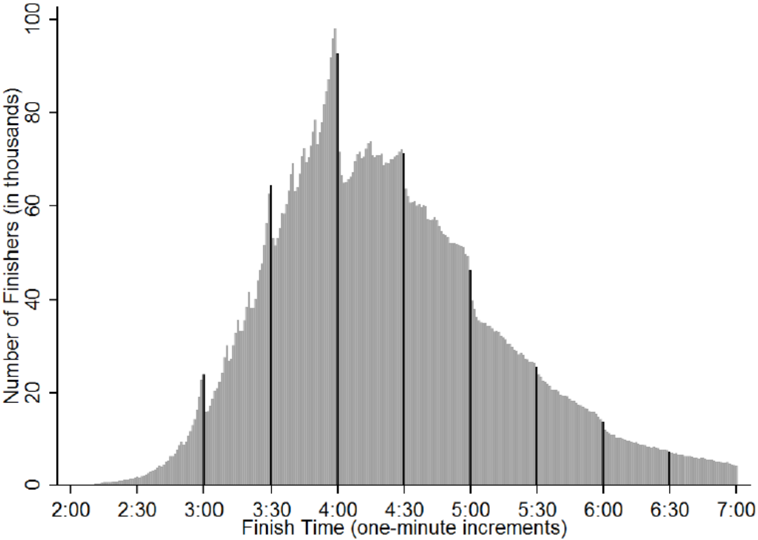
As with any method of causal inference, RD rests on an **untestable identification assumption**

- 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

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!

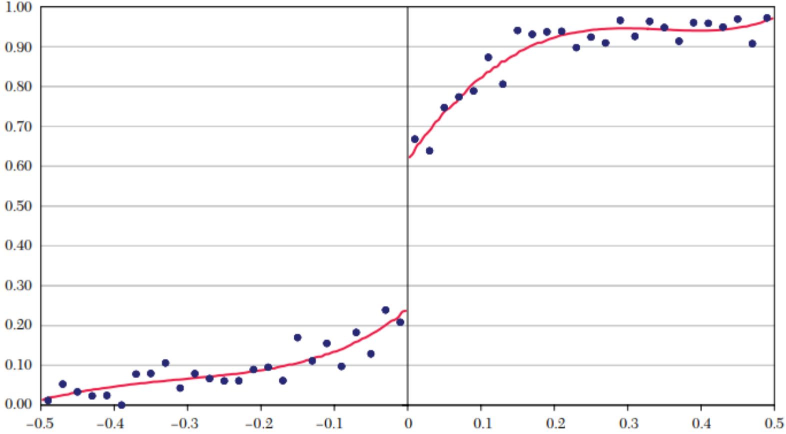
1) Explain the Identification Strategy

- 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

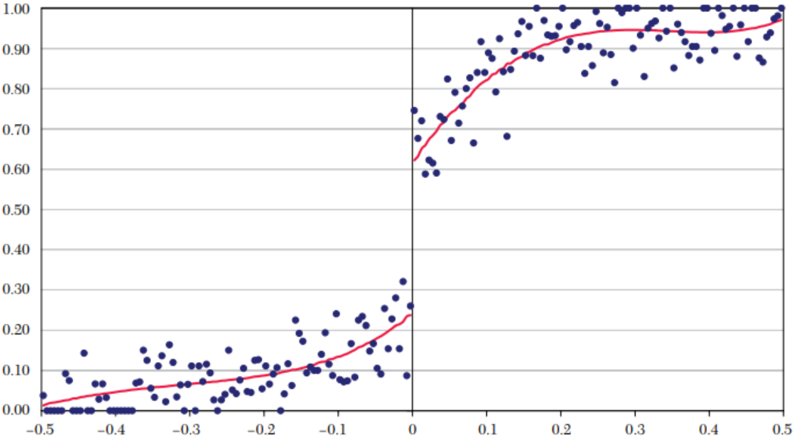
- Outcome plotted against the running variable
- Best to use binned scatters
- Important to find the right functional form
- Use practitioners' guides, for example Lee & Lemieux (2010)

RD: The Cookbook



Source: Lee & Lemieux (2010)

RD: The Cookbook



Source: Lee & Lemieux (2010)

RD: The Cookbook

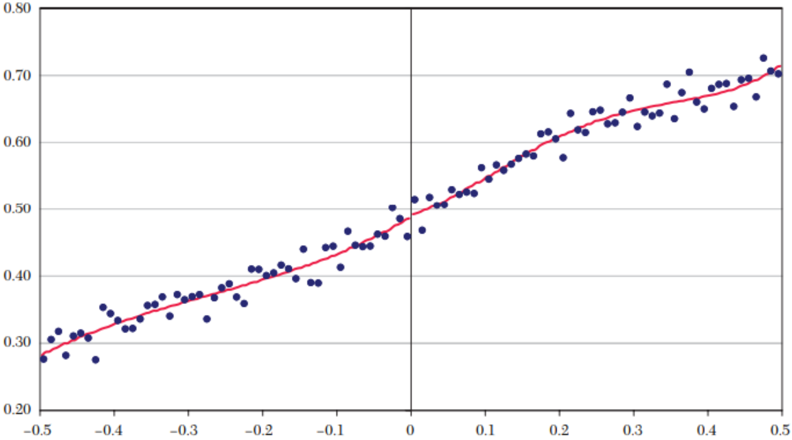
3) Report estimates based on different methods

- 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)

New Developments in RDD

Extrapolation away from the discontinuity

- Cattaneo *et al.* (2020): extrapolation based on multiple cut-offs

Disentangling multiple treatments at the discontinuity (Gilraine, 2020)

RDD and machine learning: coming soon...

The Regression Kink Design

Main Source: Card *et al.* (2015)

The Regression Kink Design

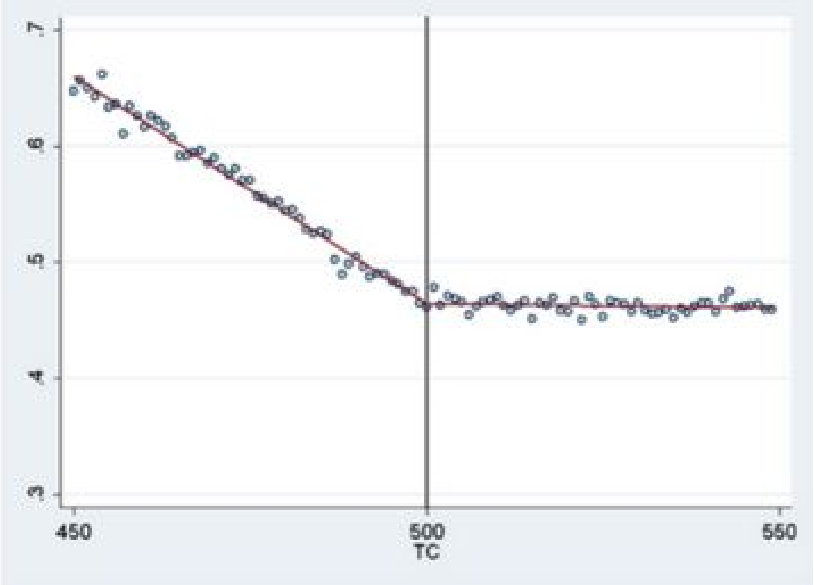
Card *et al.* (2015) introduced the **Regression Kink Design (RKD)**

Difference to RDD

- **RDD** exploits a **jump in the probability of treatment**
- **RKD** exploits a **kink in the likelihood of being treated**
- ...i.e. a sudden **change in the first derivative** of the assignment function
- ...which is often a **kink in a policy rule**

If the **likelihood of treatment exhibits a kink**, can we also see a **kink in the outcome?**

RKD Example I: Subsidies for Medication



Source: Simonsen *et al.* (2016)

RKD Example II: Unemployment Benefits

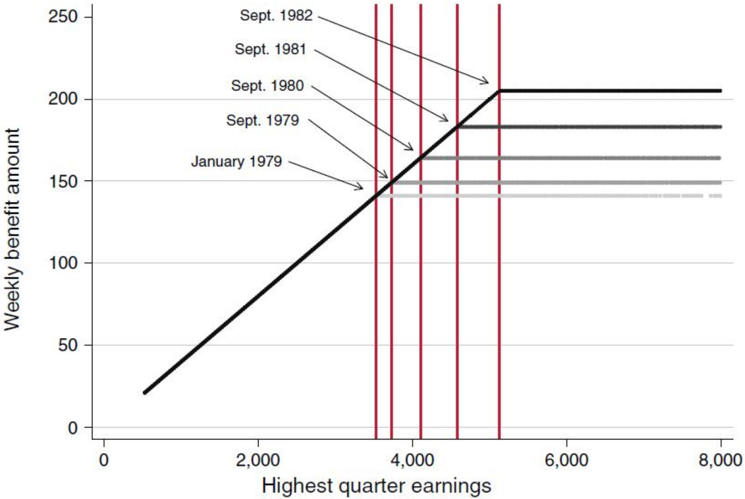
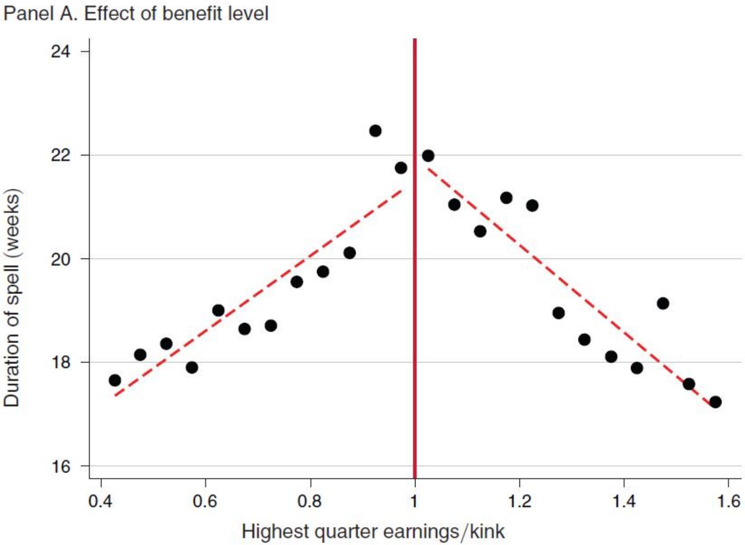


FIGURE 2. LOUISIANA: SCHEDULE OF UI WEEKLY BENEFIT AMOUNT, JAN. 1979–DEC. 1983

Source: Landais (2015)

RKD Example II: Unemployment Benefits

Is there a **kink in the outcome?**



Source: Landais (2015)

The RKD Estimand (Card *et al.*, 2015)

Assume a linear model

$$Y = \tau B + g(V) + \varepsilon$$

$B = b(V)$ is a deterministic function of V with a kink at $V = 0$

$$\tau = \frac{\lim\limits_{v_0 \to 0^+} \left.\frac{dE[Y|V=v]}{dv}\right|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \left.\frac{dE[Y|V=v]}{dv}\right|_{v=v_0}}{\lim\limits_{v_0 \to 0^+} b'(v_0) - \lim\limits_{v_0 \to 0^-} b'(v_0)}$$

Numerator: difference in outcome above and below kink
Denominator: first stage

RKD Identification (Card *et al.*, 2015)

Basic idea: units **above and below the kink are similar**

- assignment above/below **as good as random**

Identification assumptions

1. The **density of the assignment variable is smooth** at the kink point
2. The **treatment assignment rule** is **continuous** at the kink point

Units need not perfectly comply ⇒ **fuzzy RKD**

The Smooth Density Condition

People/firms should **not be able to manipulate their position**

- No bunching at the kink point
- Card *et al.* (2015) allow for small amounts of bunching

Covariates should be **continuous at the kink**

- this is a diagnostics check

The Continuity Condition

Simply put, a **setting with a discontinuous jump is no RKD**

The **RKD requires a kink** but no jump

If there is a jump, **one can use RDD**

Estimation of a Sharp RKD

Card *et al.* (2015) show that it is **sufficient to consider the effect of V on Y**

Linear regression with a **polynomial up to order $\overline{p}$** and $D = \mathbb{1}(V \geq v_0)$

$$E[Y|V = v] = \mu_0 + \left[\sum_{p=1}^{\overline{p}} \gamma_p (v - v_0)^p + \nu_p (v - v_0)^p \cdot D\right]$$

Coefficients of interest: ν_p (kinks in outcome)

Estimation of a Fuzzy RKD

In a fuzzy RKD, the first stage is no longer deterministic

Need to estimate the **reduced form** (previous slide) and the **first stage**

$$E[B|V=v] = \mu_0 + \left[\sum_{p=1}^{\bar{p}} \delta_p (v - v_0)^p + \kappa_p (v - v_0)^p \cdot D\right]$$

Obtain coefficient by **dividing the RF by the FS** $\tau = \frac{\nu_p}{\kappa_p}$

Estimate **standard errors by bootstrap**

Estimation: Refinements

RKD has **similar refinements to RDD**

- local linear regression vs. **polynomial** regression
- **Parametric** vs. **non-parametric methods**
- estimates **may depend on bandwidth**
- Can apply optimal bandwidth selection (Calonico *et al.*, 2014)

RKD is a **new methodology**; more refinements to follow

Example: Unemployment Benefits in Austria (Card *et al.*, 2015)

UI benefits: 55% of net daily earnings up to a cap

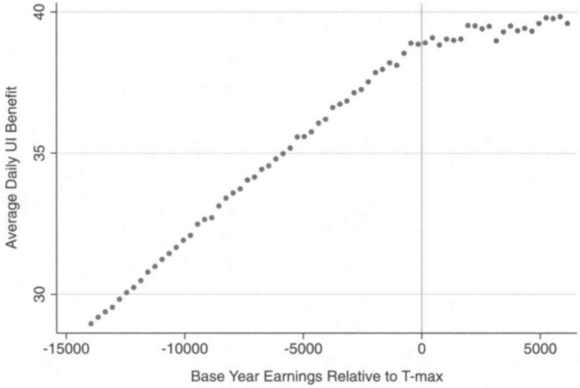
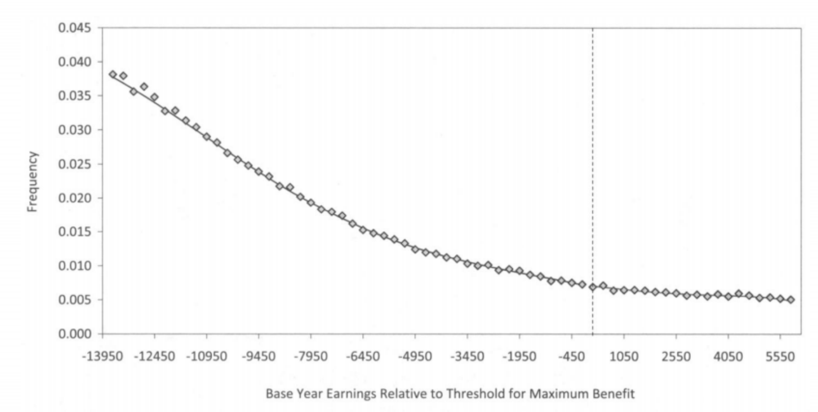


FIGURE 2.—Daily UI benefits.

Earnings are Smooth around the Kink



Effect on Unemployment Duration

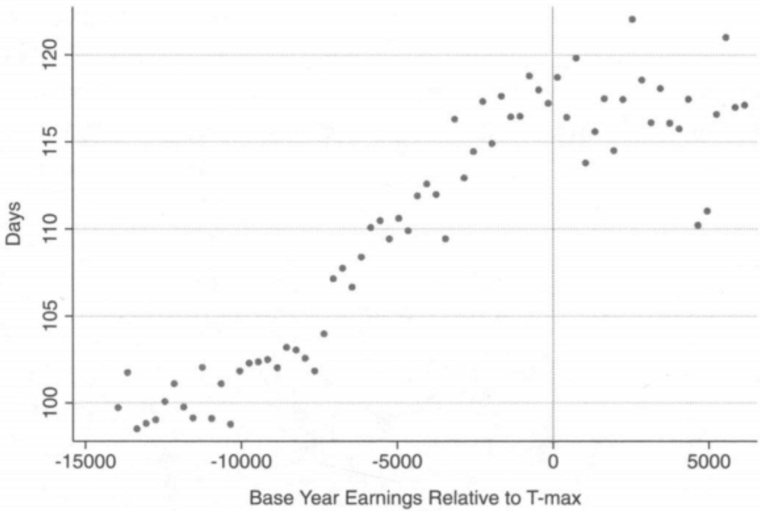
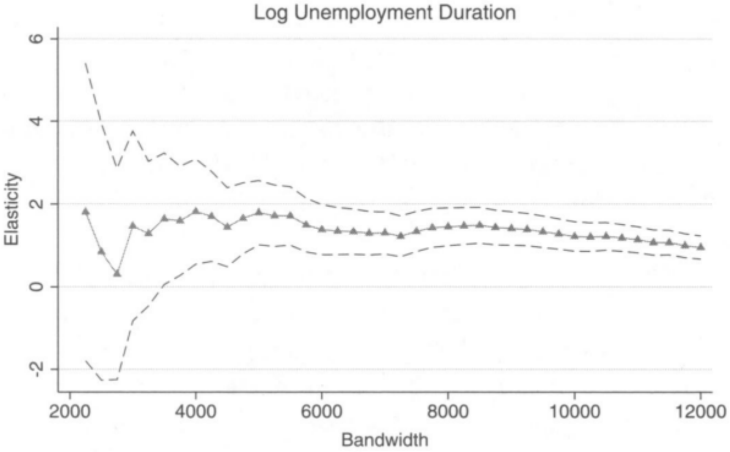


FIGURE 3.—Unemployment duration.

Effect is Robust to Bandwidth Selection



A Note on Inference

Ganong & Jäger (2018) show that **conventional SEs underestimate the uncertainty** in RKD designs

Problem: **polynomial regressions** may yield **spurious effects**

They **propose a permutation test**

- Run local linear regressions in areas without a kink
- Use the empirical distribution for hypothesis tests

RKD summary

RKD is a welcome **addition to the causal inference toolbox**

- **Straightforward to implement**
- Allows for the estimation of important **behavioral elasticities**

The **intuition is similar to RDD**

Requirement: **very detailed data!**

References I

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.

Card, David, Lee, David S., Pei, Zhuan, & Weber, Andrea. 2015. Inference on Causal Effects in a Generalized Regression Kink Design. *Econometrica*, **83**(6), 2453–2483.

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.

Cattaneo, Matias D., Keele, Luke, Titiunik, Rocío, & Vazquez-Bare, Gonzalo. 2020. Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs. *Journal of the American Statistical Association*, **0**(0), 1–12.

Ganong, Peter, & Jäger, Simon. 2018. A Permutation Test for the Regression Kink Design. *Journal of the American Statistical Association*, **113**(522), 494–504.

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.

Gelman, Andrew, & Zelizer, Adam. 2015. Evidence on the deleterious impact of sustained use of polynomial regression on causal inference. *Research & Politics*, **2**(1), 2053168015569830.

Gilraine, Michael. 2020. A Method for Disentangling Multiple Treatments from a Regression Discontinuity Design. *Journal of Labor Economics*, **38**(4), 1267–1311.

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.

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

Landais, Camille. 2015. Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design. *American Economic Journal: Economic Policy*, **7**(4), 243–78.

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

Simonsen, Marianne, Skipper, Lars, & Skipper, Niels. 2016. Price Sensitivity of Demand for Prescription Drugs: Exploiting a Regression Kink Design. *Journal of Applied Econometrics*, **31**(2), 320–337.