Acknowledgement of Country

I acknowledge the Traditional Owners and their custodianship of the lands on which we meet today.

On behalf of us all, I pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.

Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

Lecture 7 : Regression Discontinuity Design

Julie Moschion
University of Queensland

Reading for Lecture 7

- Mostly Harmless Econometrics (Angrist and Pischke): Chapter 6
- Additional Reading: Microeconometrics: Methods and Applications (Cameron and Trivedi): Chapter 25.6

Introduction

- **Regression discontinuity design** (RDD) exploit situations in which rules or laws create sharp thresholds in the implementation of policies and programs.
- For example, below the threshold people are not eligible but above the threshold they are eligible: we can exploit these **cutoffs for identification.**
- Essentially these are arbitrary rules that provide natural experiments: are people right above and right below the threshold that different?
- RDD exploits the fact that *cutoffs are often arbitrary*. This generates **quasi-experimental** settings in which you can compare people (or cities, firms, countries,...) who are **just affected** by the rule with people who are **just not affected** by the rule.
- We call these quasi-experiments because the design lacks the element of random assignment to treatment or control.

Introduction

- The world abounds with such rules:
  - Students receive a scholarship if their test scores are above above 3.0.
  - Children are allowed to start school if they turn 6 by 30th of June of that year.
  - Individuals are eligible for a micro-finance loan if they own less than 0.5 acres of land.
  - Legislators are elected if they receive over 50% of the vote.
- These arbitrary rules provide wonderful natural experiments!
- Essentially because the person who just misses out (with a test score of 2.9; born on 1 July...) is usually not that different from the person who only just made it in!

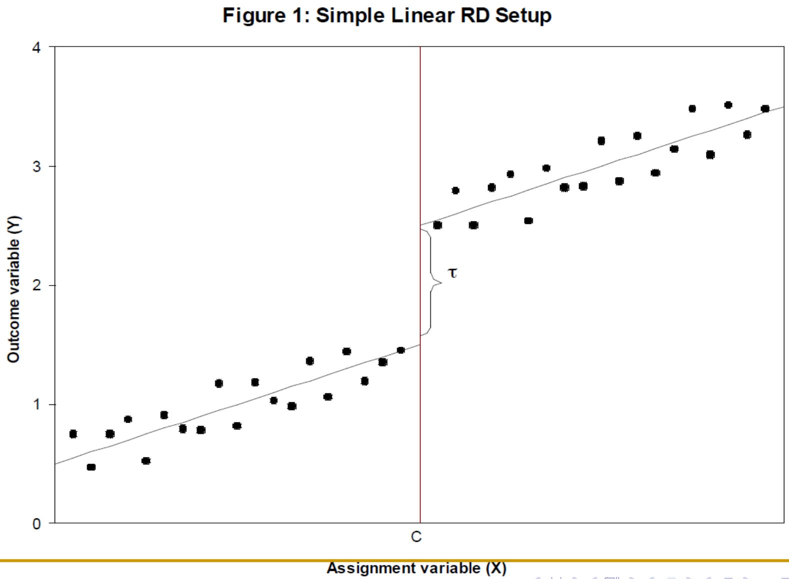
Introduction

- This idea goes as far back as Thistlethwaite and Campbell (1960) who analyzed the **impact of winning a merit scholarship on subsequent educational outcomes**, using the fact that the allocation of these awards was based on the **comparison of the observed test score with a threshold score.**
- Merit scholarships D were awarded to students if their test score x was above a certain threshold c.
  - x is called the running/forcing/assignment variable.
  - D is the (explanatory) variable of interest.
- The main idea behind the research design is that individuals with scores just **below the cutoff** (who did not receive the scholarship) were good comparisons to those just above the cutoff (who did receive the scholarship).
- This allows us to compare individuals (or cities, firms, countries,...) who were only just affected by the rule with those who are only just not affected by the rule and consider them to be a **valid counterfactual or comparison group**.

Introduction

- It took a long time before people realised the strength of RD designs.
- In the 1990s, there was a general shift towards empirical work focusing on credible natural experiments, including some well-published papers such as Angrist and Lavy (1999), van der Klaauw (2002), or Black (1999).
- Unlike the identification assumptions for IV, the assumptions for the RD design are (to a large extent) testable.
- In the right context, RD design can provide a highly credible and transparent way of estimating causal effects of programs/interventions.

Example of a linear RD Setup



Thistlethwaite and Campbell (1960)

- Back to TC (1960) to estimate the impact of winning a merit scholarship on subsequent educational outcomes.
  - Merit scholarships D were awarded to students if their test score x (i.e. assignment variable) was above a certain threshold c.
- D is the (explanatory) variable of interest, i.e. the treatment status $D \in \{1, 0\}$.

$$D = 1 \text{ if } x \geq c \text{ and } D = 0 \text{ if } x < c$$

- Thus there is a discontinuous jump in treatment at the test score threshold c.

Thistlethwaite and Campbell (1960)

- Let's denote the outcome by y.
- There appears to be no reason other than the merit award for y to be a discontinuous function of x.
- In other words, why would subsequent educational outcomes vary discontinuously with test scores exactly at the award threshold? Unless this specific cutoff is the cutoff for another program, there is no reason!
- This is the underlying motivation of the RD design.
- Assuming that **the relationship between y and x is otherwise linear**, a simple way of estimating the treatment effect τ is by estimating the linear regression:

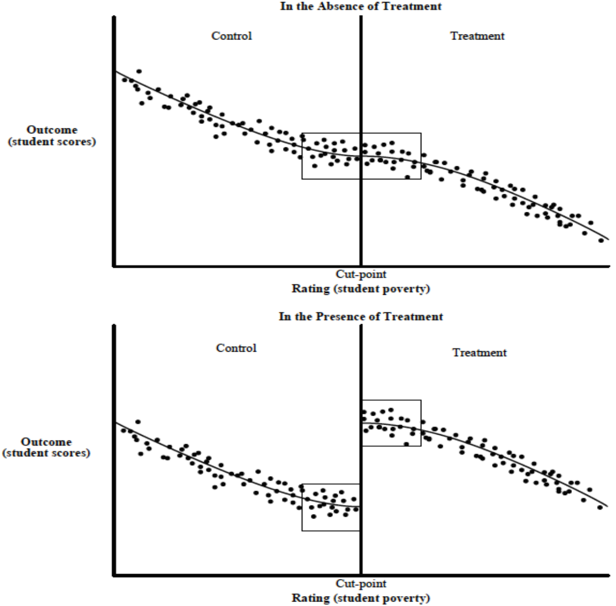
$$y = \alpha + D\tau + x\beta + u$$

- D is the treatment dummy and x is the assignment variable.

Thistlethwaite and Campbell (1960)

- Comparing individuals close to the discontinuity is thus a **natural estimate** of the treatment effect.
- In practice, we cannot only look at individuals super close to the cutoff since the closer we look, the less observations will be available.
- We **use data away from the cutoff** to get reasonable estimates for the treated and untreated values of y at $x = c$.
- If the underlying function is linear, then τ the coefficient on D from the OLS regression yields the **causal effect** of interest.

Two Ways to Characterize Regression Discontinuity Analysis



NOTE: Dots represent individual schools. The vertical line in the center of each graph designates a cut-point, above which candidates are assigned to the treatment and below which they are not assigned to the treatment. The boxes represent the proportion of the distribution proximal enough to the cut-point to be used in regression discontinuity analysis when the relationship is viewed as local randomization.

Sharp RD and Fuzzy RD

- Sharp RD: the treatment (D) is a deterministic function of x.
  Example: if your GPA is above 3.0, you receive a scholarship for sure.
- Fuzzy RD: exploits discontinuities in the probability of treatment conditional on x.
  Example: If your GPA is above 3.0 and you get an IELTS score above 6.5, then you are more likely to receive a scholarship.

Sharp RD

- Sharp RD designs exploit situations where the treatment status D is a deterministic and discontinuous function of an assignment variable x.
  $D = 1$ if $x \geq c$ (treatment group)
  $D = 0$ if $x < c$ (control group)
  Where c is a known threshold or cutoff.
- Once we know x we know D.
- There is no value of x at which you observe both treatment and control observations.

Sharp RD

- Formally, sharp RD leads to the following regression:

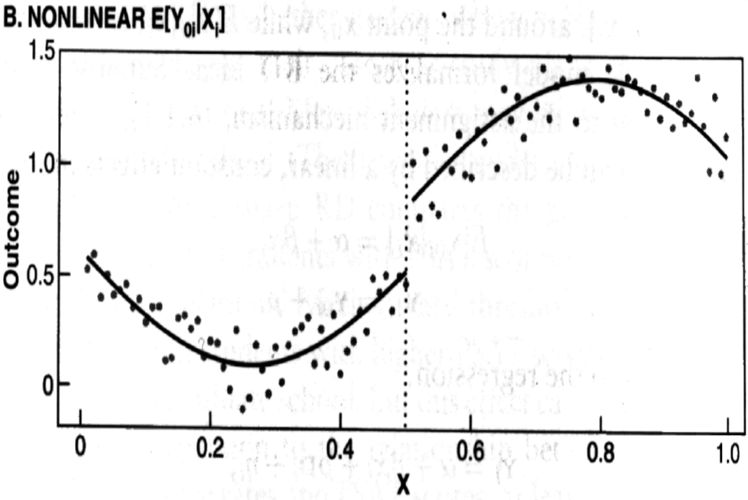
$$y = \alpha + D\tau + x\beta + u$$

- The **key identification assumption** is that 'all other factors' affecting y are continuous (no jumps) with respect to x (assignment variable) at cutoff c.
- There are no observed or unobserved determinant of y that experience a jump when x goes the over cutoff c.
- If that was the case, the effect of D would be confounded with the effect of this other variable experiencing a jump at the cutoff c.
- This assumption cannot be directly tested. But there are some tests which give suggestive evidence whether the assumption is satisfied (see below).

Sharp RD (nonlinear case)

- Until now, we assumed that the relationship between the assignment variable x and the outcome y was linear (i.e could be represented by a straight line: if x increases by 1 unit, y increases by β units, whatever the value of x).
- But the functional form can be nonlinear, because:
  - The relationship between the assignment variable x and the outcome y is nonlinear.
  - There is an interaction between the assignment variable x and the treatment D.
- In that case we can estimate:
$$y = \alpha + D\tau + \mathbf{f(x)} + u$$
- The Misspecification of the functional form may generate a bias in the treatment effect.
- The consequences of using an incorrect functional form are potentially more severe for RD than for other methods.

Sharp RD (nonlinear case)



- If this was specified linearly, the jump below and above the cutoff would appear as larger than it actually is and overestimate the treatment effect.

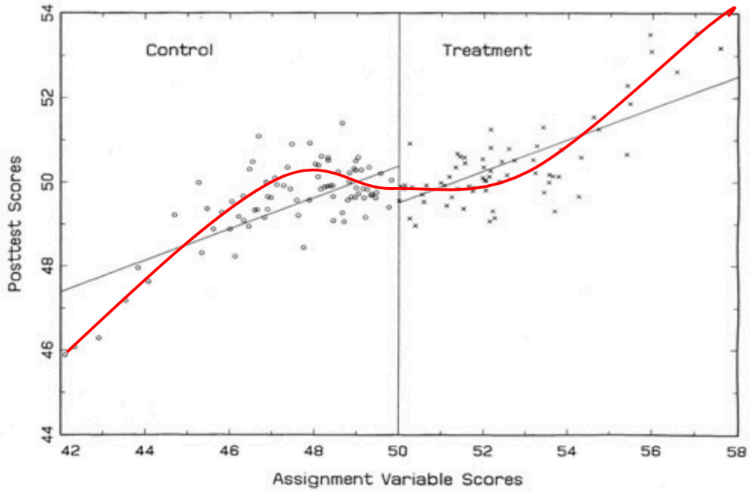
Sharp RD (nonlinear case)

- Solution: Use a pth-order polynomial:

$$y = \alpha + D\tau + x\beta_1 + x^2\beta_2 + ... + x^p\beta_p + u$$

- The x's are centered around the cutoff value prior to running the regression (x=0 at c).
- This ensures that the coefficient on D is the treatment effect.
- Common practice is to use a 4th order polynomial (p=4).

Sharp RD (nonlinear case)

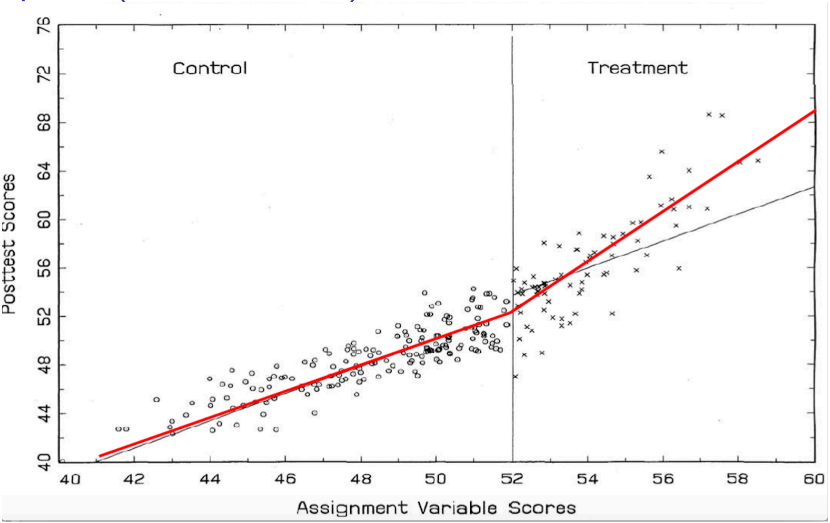


- The gap between the two straight lines at the cutoff suggests a negative treatment effect. But this is wrong because we modeled these data with a linear model when the underlying relationship was nonlinear. Using a 3rd-order polynomial actually shows no effects.

Sharp RD (nonlinear case)

- Let's now look at the case where there is an **interaction between the assignment variable x and the treatment D.**
- This happen for instance when the treatment works better for some people than for others.
  - For example, it is possible that children with higher pre-test scores benefit more from the treatment than do children with lower scores.
  - The interaction x*D allows to have different slopes for the effect of x on y before and after the cut-off. It captures how the effect of the treatment D differs for people below vs above the cutoff.
- If this interaction (between the assignment variable and treatment) is not modeled correctly, a false discontinuity will appear.

Sharp RD (nonlinear case)



- When the interaction term is not included, an artefactual discontinuity at the cutoff is produced (gap between straight lines). When the interaction term is included (x*D), there is no discontinuity around the cutoff, i.e. no treatment effect.

Sharp RD (nonlinear case)

- We can generalise the function $f(x)$ by allowing the coefficient on x to differ on both sides of the threshold by including it alone and with an interaction with D.

$$y = \alpha + D\tau + x\beta + x \cdot D\gamma + u$$

- As Lee and Lemieux (2010) note, allowing different functions on both sides of the discontinuity should be the main specification in an RD paper.

Fuzzy RD

- The fuzzy RD design exploits a **discontinuity in the probability of getting treatment** conditional on covariate x.
- D is no longer deterministically related to x but there is a jump in the probability of treatment at the cutoff c:

$$P(D_i = 1|x_i) = g_1(x_i) \text{ if } x_i \geq c$$
$$P(D_i = 1|x_i) = g_0(x_i) \text{ if } x_i < c$$

- where $g_1(x_i) \neq g_0(x_i)$
- $g_1(x_i)$ and $g_0(x_i)$ can be anything as long as they differ at c (a fixed probability, linear in x, a polynomial in x...).

Fuzzy RD

- The relationship between the probability of treatment and x can be written as:

$$P(D_i = 1|x_i) = g_0(x_i) + [g_1(x_i) - g_0(x_i)]\,T_i$$

- Where $T_i = 1(x_i \geq c)$
- The actual discontinuity at c becomes an instrumental variable for the treatment status D: it increases the chances of getting treated (first stage) and is exogenous (exclusion restriction).
- Fuzzy RD is an IV method that uses T as an IV for D.

Validity of the RD Design

- Set-up:
  - The treatment is fully or partially determined (sharp or fuzzy) by the assignment variable.
  - There is a **discontinuity in the allocation of treatment** at some cutoff value of the assignment variable.
- Identifying assumptions:
  - Participants and evaluators **cannot manipulate the assignment variable** to influence whether someone receives the treatment or not (No manipulation of the assignment variable).
  - **Other variables** that affect the outcome variable do **not change discontinuously at the cutoff point** (No jumps in other covariates).

Assumption 1: no manipulation of the assignment variable

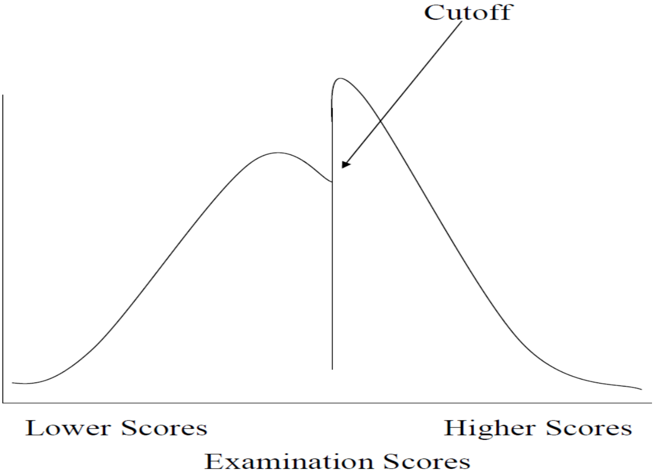
- We assume that there is **no manipulation** of the assignment variable.
- Manipulation occurs when participants or evaluators manipulate assignment scores so that participants on the "wrong" side of the cutoff receive or avoid treatment. For example:
  - If students know the test score threshold and have the option of re-taking the test, a select subsample of those with test scores just below the threshold may do so and invalidate the RDD.
  - If evaluators are aware of the importance of the threshold and of the consequences of failing, they may be reluctant to score an individual just below the threshold.

Assumption 1: no manipulation of the assignment variable

- The issue is that the researcher doesn't observe the real score.
- Manipulation can be suspected because it results in bunching of the assignment variable on one side of the threshold (see below), i.e. a discontinuity of the conditional density of the assignment variable at the threshold.
- A graphical analysis can help detect if it occurs.
- Note that in principle, the RD design does not require continuity of the density of x, but a discontinuity is suggestive of violations of the no-manipulation assumption.

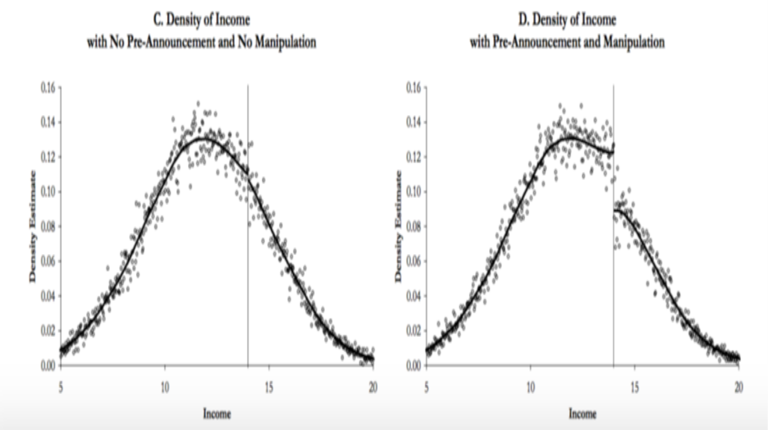
Assumption 1: no manipulation of the assignment variable

- Example of manipulation above the cut-off



Assumption 1: no manipulation of the assignment variable

- Example of manipulation below the cut-off

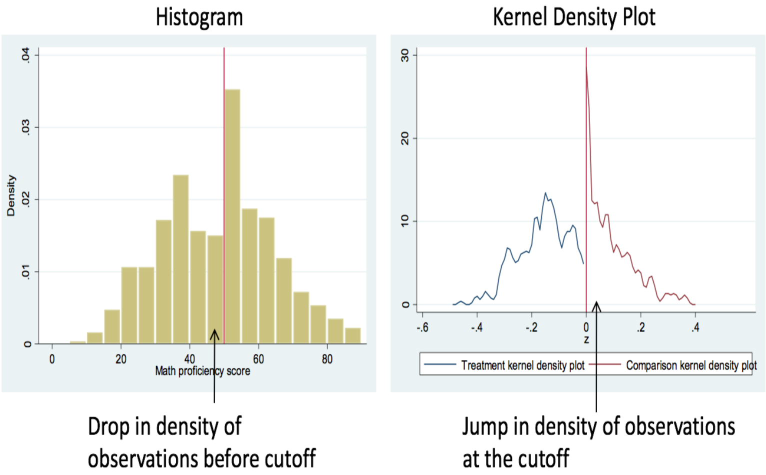


- Income support program in which those earning below 14,000 qualify for support.

Assumption 1 : no manipulation of the assignment variable

- Diagnostic test for manipulation: "McCrary Test" (2008).
- Basically a statistical test for assessing whether there is a discontinuity in the density of observations at the cut-off.
- This consists in testing the null hypothesis of the continuity of the density of the assignment variable at the discontinuity point, against the alternative of a jump in the density function at that point.
  - Create a histogram of the density of the assignment variable using a particular bin size, ensuring that no bin overlaps with the threshold.
  - Estimate a local linear regressions, one to the right and one to the left of the threshold. The midpoints of the histogram bins are treated as regressor and the number of observations falling into the bins as the outcome.
  - Test whether the difference in the slopes of the two regressions is statistically different from zero.

Assumption 1 : no manipulation of the assignment variable



Assumption 1: no manipulation of the assignment variable

- Being able to run a test is “reassuring” for the researcher, but is not sufficient for proving the validity of the RD design:
  - Test fails to reject the null but the RD design invalid: if sorting occurs in both directions with people trying to avoid and enter the treatment, cancelling out the discontinuity $\Rightarrow$ Test is more informative when manipulation occurs in one direction only.
  - Test rejects the null but RD design is valid: if some units who barely fail are given extra points so that they barely pass but these are chosen randomly from those who barely failed (the RD estimator can remain unbiased).
- Careful consideration must be combined with the test to decide how to interpret results.
- What to do next? Well if you have evidence of manipulation, we are back to the OVB problem.

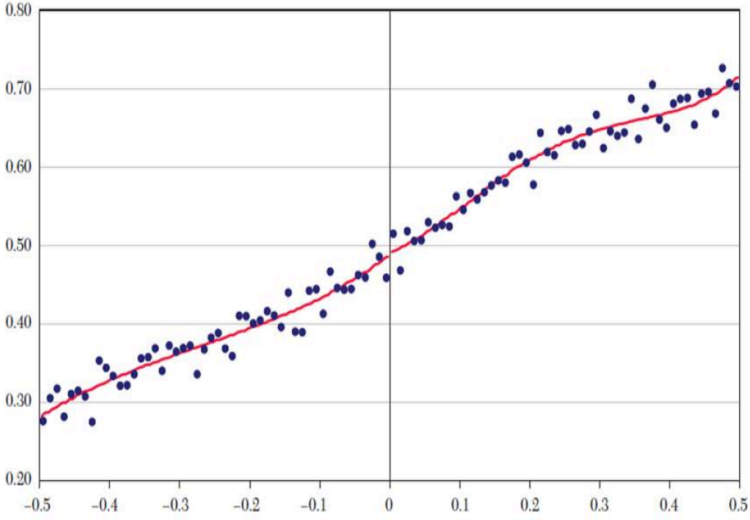
Assumption 2: no jumps in covariates at threshold

- In principle, controlling for covariates is not needed for identification with RD as it exploits the randomness of treatment allocation around the cut-off $\Rightarrow$ If there is no manipulation, adding covariates should not affect the coefficient very much (if it does, there is a problem!).
- But covariates can improve precision (reduce standard errors).
- All observed predetermined characteristics (and also unobserved) should have identical distributions on either side of the threshold: **there should be no discontinuities in the covariates**.

Assumption 2: no jumps in covariates at threshold

- **Graphical inspection** is key here - Two types of graph should always be created, where the assignment variable is graphed against:
  - The outcome: should show a discontinuity (if you can't see the main result with a simple graph, it is probably not there);
  - Other covariates: should show no discontinuity (this is equivalent to balancing test in randomized experiment).
- Test: the treatment effect should be insignificant on the covariates. If you find jumps on some covariates, it does not invalidate the RD design necessarily but it is critical to control for these covariates. Just like for RCTs or IVs, the validity then rests on the assumption that there are no unobservables that would jump at the threshold.

Assumption 2: no jumps in covariates at threshold



Specification Checks for RD

- Threats to identification
  - No manipulation of assignment variable (McCrary test)
  - No jumps in covariates (No treatment effect on covariates)
  - No discontinuities at other values of the assignment variable (placebo test)
- Other checks
  - Robustness to inclusion of higher order polynomials: results do not change drastically when including higher order polynomials of the assignment variable.
  - Sensitivity to bandwidth choice: test that the results are not affected when the window around the cutoff changes (standard errors may go up a bit but hopefully the point estimate does not change).

Angrist and Lavy (1999): Class size effect on test scores

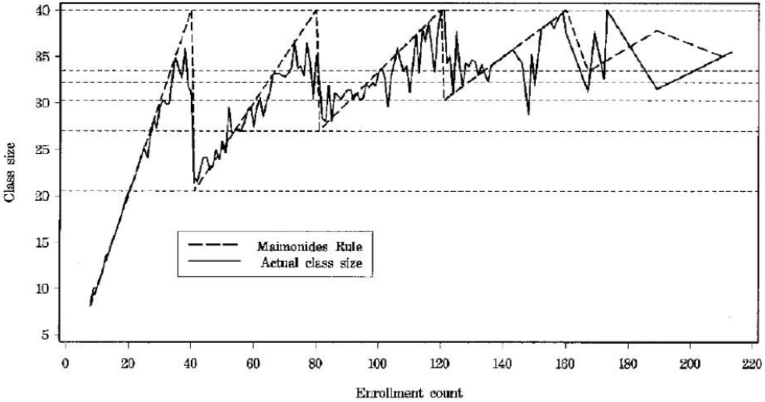
- Angrist and Lavy exploit an old Talmudic rule (Jewish law) applied in Israel that classes should be split if they have more than 40 students to estimate the effect of class size in a fuzzy RDD.
  - A school with 40 students has only one class → The class size 40.
  - A school with 41 students has two classes → The class sizes are 21 and 20.

$$m_{sc} = \frac{e_s}{int[\frac{e_s - 1}{40}] + 1}$$

Where int[a] is the integer part of real number a.

Angrist and Lavy (1999): Class size effect on test scores

- Maimonides (philosopher and key figure in Jewish religious tradition) rule and actual class size



- The rule is not followed completely and strictly. They therefore have a fuzzy discontinuity design.

Angrist and Lavy (1999): Class size effect on test scores

- They extend the fuzzy RD in two ways compared to the simple example presented earlier:
  - They use multiple discontinuities (at each multiple of 40).
  - The variable of interest (class size) takes on many values: instead of having a single treatment (small vs class size) with a probability that varies around the cutoff, there are continuous treatments (different class sizes) around each cutoff.
    $\rightarrow$ The first stage exploits discontinuities in average class size instead of probabilities of a single treatment.

Angrist and Lavy (1999): Class size effect on test scores

- They want to estimate the relationship between average achievement and class size.

$$y_i = \alpha + x_i\beta + classSize_i\tau + u_i$$

- Estimating this relationship with OLS may lead to biased results because class size is likely to be correlated with the error term. For example:
  - Parents from higher socioeconomic backgrounds may put their children in schools with smaller classes.
  - Principals may put weaker/troubling students in smaller classes.

Angrist and Lavy (1999): Class size effect on test scores

- Angrist and Lavy therefore use the Maimonides rule in a fuzzy RD design:

$$y_{isc} = \alpha + e_s\beta_1 + e_s^2\beta_2 + d_s\tau_1 + n_{sc}\tau_2 + e_{isc}$$

where y_{isc} is the test score of student i in school s and class c.
e_s is the number of enrolled students in school s (i.e. the assignment variable).
d_s is the percentage of disadvantage students in class.
n_{sc} is the class size, the variable of interest.

Angrist and Lavy (1999): Class size effect on test scores

- The variables relate to the fuzzy RD example as follows:
  - e_s (number of students in school s) plays the role of x_i.
  - n_{sc} (class size) plays the role of D_i (the treatment).
  - m_{sc} (the function describing Maimonides rule) plays the role of T_i.
- The first stage regression is:

$$n_{sc} = \gamma_0 + \delta_1 e_s + \delta_2 e_s^2 + \gamma_1 d_s + \phi m_{sc} + \epsilon_{isc}$$

The rule m_{sc} is used as an IV for the treatment n_{sc}.

Angrist and Lavy (1999): Class size effect on test scores

- OLS results: there is a positive relationship between class size and test scores (better students assigned to larger classes).
- With controls for percentage disadvantaged and total enrollment, the relationship is insignificant.

	Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean score		74.3			67.3	
(s.d.)		(8.1)			(9.9)	
Regressors						
Class size	.221	−.031	−.025	.322	.076	.019
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)
Percent disadvantaged		−.350	−.351		−.340	−.332
		(.012)	(.013)		(.018)	(.018)
Enrollment			−.002			.017
			(.006)			(.009)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30
R^2	.036	.369	.369	.048	.249	.252
N		2,019			2,018	

Angrist and Lavy (1999): Class size effect on test scores

- RD second stage results (reading): the effect of class size is significantly negative!
- The piecewise linear trend aims at controlling piece by piece for the linear trend in the variable that generates the discontinuity (enrolments) to keep just the variation around it.

	Full sample			
	(1)	(2)	(3)	(4)
Mean score		74.4		
(s.d.)		(7.7)		
Regressors				
Class size	−.158	−.275	−.260	−.186
	(.040)	(.066)	(.081)	(.104)
Percent disadvantaged	−.372	−.369	−.369	
	(.014)	(.014)	(.013)	
Enrollment		.022	.012	
		(.009)	(.026)	
Enrollment squared/100			.005	
			(.011)	
Piecewise linear trend				.136
				(.032)
Root MSE	6.15	6.23	6.22	7.71
N		2019		1961

Internal and external validity

- The strength of RDD is its internal validity, arguable the strongest of any quasi-experimental design.
- External validity may be limited though as it is a LATE:
  - Sharp RD provides estimates for the sub-population which is right at the threshold.
  - Fuzzy RD restricts the estimates even further to compliers at the threshold (like for IVs), i.e those whose treatment status change because of the IV.

SUPPLEMENTARY MATERIAL

Practical advice for RD: Figures

- Present the main RD figures using a fixed number of (non overlapping) binned local averages (grouping variables by intervals of the assignment variable).
- DO NOT plot the raw data without a minimal amount of local averaging or present a continuum of nonparametric estimates (with a single break at the threshold): this will superficially give the impression of a discontinuity even when there is none.
- Lee and Lemieux (2010) recommend "undersmoothing" (larger bins) to really get a sense of the shape of the underlying function.

Practical advice for RD: Testing the continuity of the density of the assignment variable

- McCrary's test (2008): "Ho: continuity of the density of the assignment variable at the discontinuity point, c".
- Rejecting Ho suggests that there is a discontinuity of x around the threshold.
- In the first step, you partition the assignment variable into bins and calculate frequencies (number of observations) in the bins.
- In the second step, you treat those frequency counts as dependent variable in a local linear regression.
- McCrary adopts the nonparametric framework for asymptotics. See his website (http://www.econ.berkeley.edu/ jmccrary/DCdensity/) for details on the test.

Practical advice for RD: Testing jump in covariates

- Test whether other covariates exhibit a jump at the threshold: re-estimate the RD model with the covariate as the dependent variable.
- if the result is significant, the covariate exhibits a jump.
- This is a type of placebo test.

Practical advice for RD: Testing for jumps outside of the threshold

- Imbens and Lemieux (2008) suggest to look at one side of the discontinuity and take the median of the forcing variable in that section and test whether you can find a discontinuity in that part.
- Do the same on the other side.
- This is another type of placebo test.