

Regression Discontinuity Designs

PP5001 · Week 7 · Martinmas 2026

Professor David A. Jaeger

Sharp Regression Discontinuity

Let X_i be the running variable and x_0 the known cutoff:

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

Assignment is **deterministic**: once we know X_i , we know D_i .

Treatment changes discontinuously at x_0 . The running variable itself may change by an arbitrarily small amount while treatment changes from 0 to 1.

Sharp Discontinuity

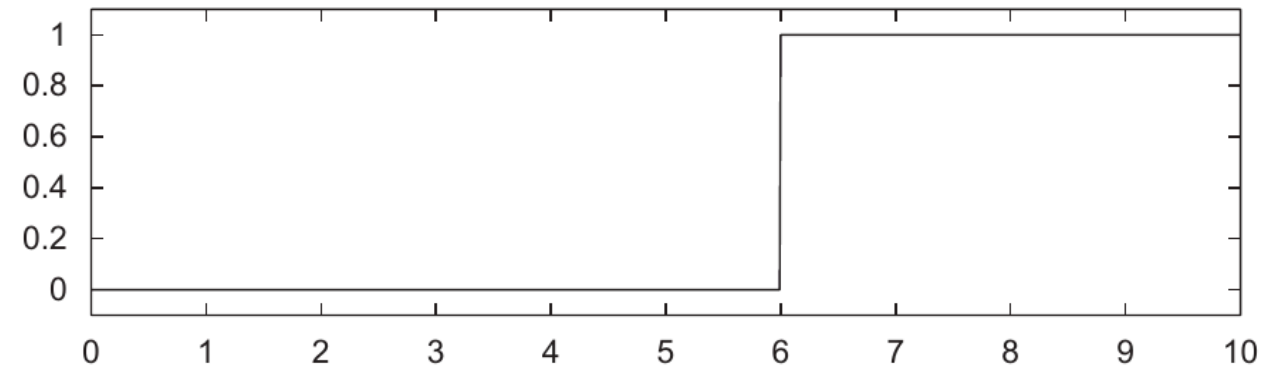


Fig. 1. Assignment probabilities (SRD).

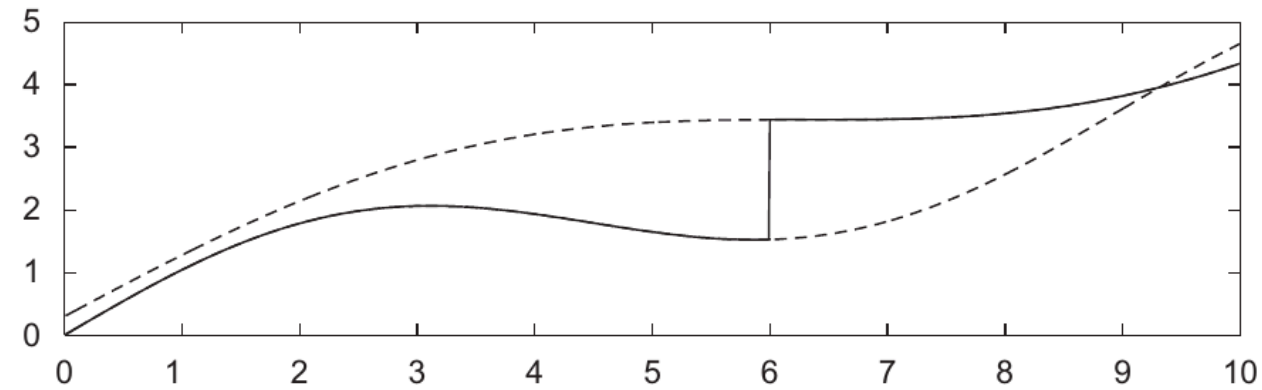


Fig. 2. Potential and observed outcome regression functions.

Imbens and Lemieux (2008), Figures 1–2.

The Causal Effect at the Cutoff

Using our potential-outcome notation, define

$$\rho = E \left[Y_i^1 - Y_i^0 \mid X_i = x_0 \right] .$$

The observed outcome is

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 .$$

To the right of the cutoff we observe treated outcomes. To the left we observe untreated outcomes.

With a continuously distributed score, identification uses the limiting conditional means as we approach x_0 from each side.

The Continuity Assumption

For each potential outcome, the conditional mean approaches the same value from either side of the cutoff:

$$\lim_{x \uparrow x_0} E [Y_i^0 \mid X_i = x] = \lim_{x \downarrow x_0} E [Y_i^0 \mid X_i = x] = E [Y_i^0 \mid X_i = x_0] .$$

$$\lim_{x \uparrow x_0} E [Y_i^1 \mid X_i = x] = \lim_{x \downarrow x_0} E [Y_i^1 \mid X_i = x] = E [Y_i^1 \mid X_i = x_0] .$$

Average outcomes under either treatment state would change smoothly as we cross the cutoff.

The two potential-outcome means can differ from one another. That difference at x_0 is the treatment effect.

Deriving the Sharp RD Effect

Immediately to the right, $D_i = 1$. By continuity,

$$\lim_{x \downarrow x_0} E[Y_i | X_i = x] = E[Y_i^1 | X_i = x_0] .$$

Immediately to the left, $D_i = 0$. By continuity,

$$\lim_{x \uparrow x_0} E[Y_i | X_i = x] = E[Y_i^0 | X_i = x_0] .$$

Subtracting gives

$$\rho = \lim_{x \downarrow x_0} E[Y_i | X_i = x] - \lim_{x \uparrow x_0} E[Y_i | X_i = x] .$$

Lee (2008): Elections and Incumbency

Lee studies US House elections and the effect of winning office on subsequent electoral success.

- **Running variable:** the Democratic winning margin in the current election.
- **Cutoff:** a margin of zero.
- **Treatment:** a Democratic victory in the current election.
- **Outcomes:** subsequent candidacy, vote share, and victory.

Why should narrow winners and narrow losers be informative about the effect of incumbency?

Lee: Subsequent Democratic Vote Share

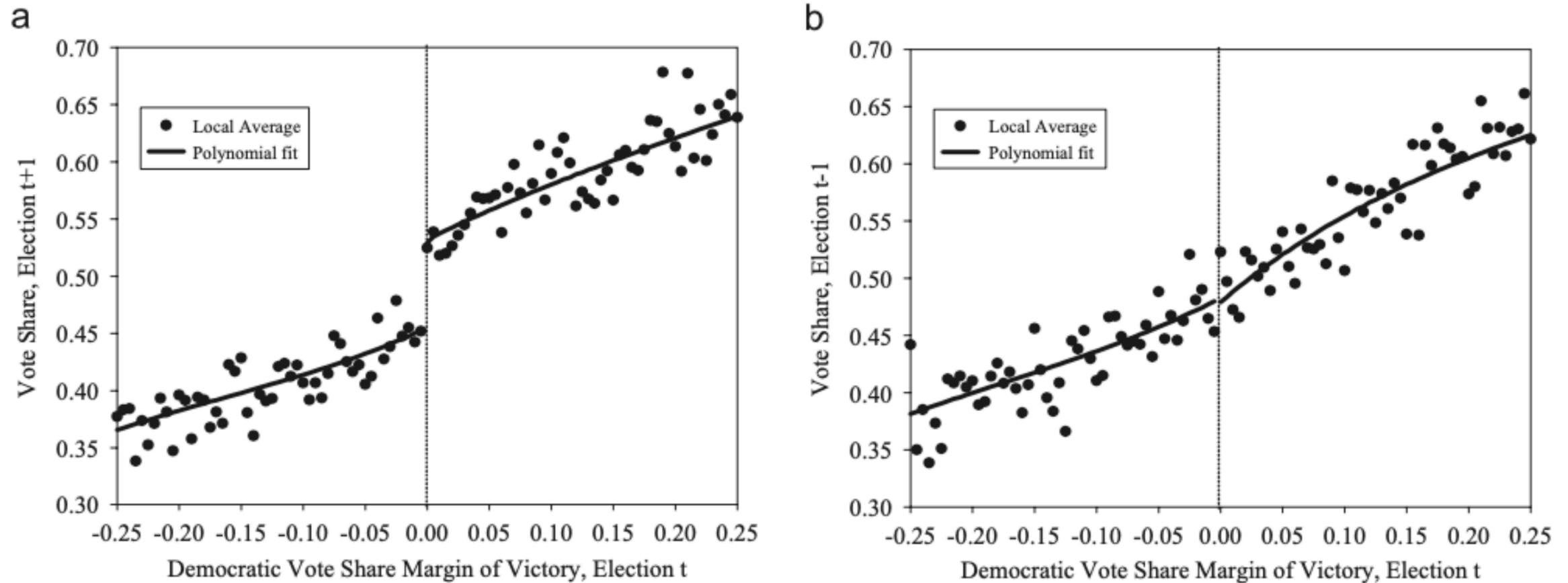


Fig. 4. (a) Democrat party's vote share in election $t+1$, by margin of victory in election t : local averages and parametric fit. (b) Democratic party vote share in election $t-1$, by margin of victory in election t : local averages and parametric fit.

Lee (2008), Figure 4.

RD as a Regression

Centre the running variable: $R_i = X_i - x_0$.

Suppose, locally,

$$E[Y_i^0 | X_i] = \beta_0 + \beta_1 R_i, \quad Y_i^1 = Y_i^0 + \rho.$$

Then

$$Y_i = \beta_0 + \beta_1 R_i + \rho D_i + \epsilon_i.$$

Here $\epsilon_i = Y_i^0 - E[Y_i^0 | X_i]$ is the disturbance: the difference between an individual's untreated outcome and its conditional mean. Thus $E[\epsilon_i | X_i] = 0$.

At $R_i = 0$, the fitted untreated outcome is β_0 and the fitted treated outcome is $\beta_0 + \rho$.

The difference between the intercepts is the treatment effect at the cutoff.

Allowing Different Slopes

Allow the relationship between the score and the outcome to differ on the two sides:

$$Y_i = \beta_0 + \rho D_i + \beta_1 R_i + \gamma_1 D_i R_i + \epsilon_i.$$

- Below the cutoff: intercept β_0 , slope β_1 .
- Above the cutoff: intercept $\beta_0 + \rho$, slope $\beta_1 + \gamma_1$.

Because $R_i = 0$ at the cutoff, the interaction contributes zero there. The jump remains ρ .

Polynomial Regression

If the conditional mean is curved, we can allow powers of the centred score:

$$Y_i = \beta_0 + \rho D_i + \sum_{j=1}^p \beta_j R_i^j + \sum_{j=1}^p \gamma_j D_i R_i^j + \epsilon_i.$$

The two sums allow different shapes on either side of the cutoff.

At $R_i = 0$, all the terms in those sums are zero, so ρ remains the jump.

A high-order polynomial over a wide range can fit poorly near the cutoff and confuse curvature with a discontinuity.

Local Regression and Bandwidth

Use observations within a **bandwidth** h of the cutoff: $-h < R_i < h$.

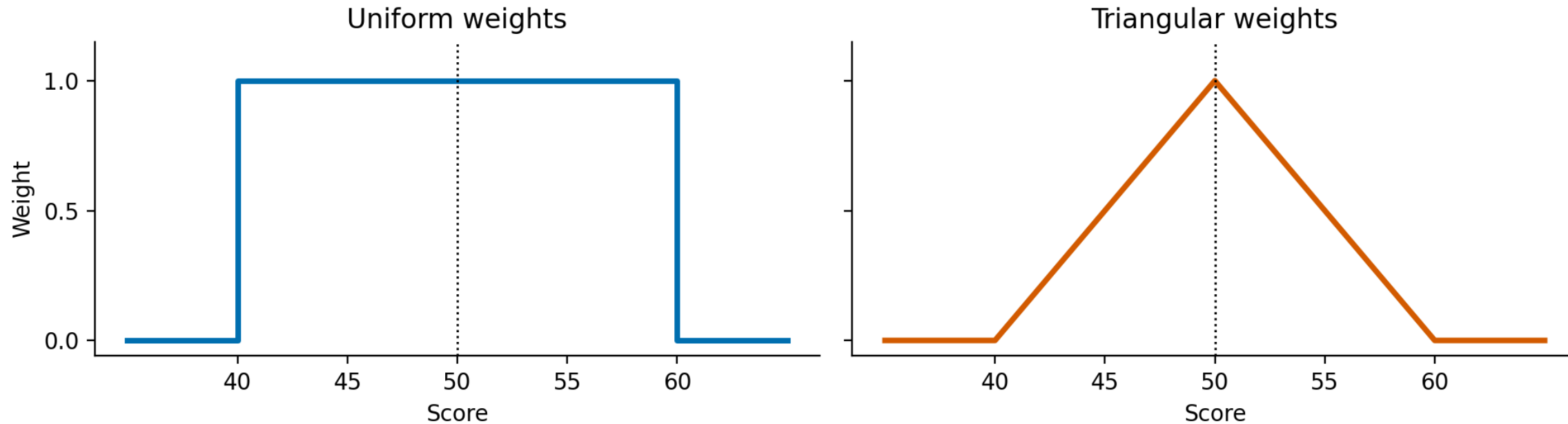
Fit a separate line on each side and compare the fitted values at $R_i = 0$.

With a cutoff of **50**, $h = 10$ uses scores between **40 and 60**. With $h = 5$, use **45 to 55**.

- A smaller bandwidth gives a more local comparison with fewer observations.
- A wider bandwidth provides more observations, but fitting a simple function over a wider range can introduce approximation bias.

Kernel Weights and Local Linear Regression

A **kernel** assigns weights according to distance from the cutoff.



With triangular weights, $w_i = 1 - |R_i|/h$ inside the bandwidth and zero outside it.

Estimate by weighted least squares:

$$\min_{\beta_0, \rho, \beta_1, \gamma_1} \sum_i w_i (Y_i - \beta_0 - \rho D_i - \beta_1 R_i - \gamma_1 D_i R_i)^2.$$

Bandwidth Choice and Inference

A data-driven bandwidth balances squared bias and variance: **mean squared error**.

At an MSE-optimal bandwidth, approximation bias can still be large enough to matter for confidence intervals.

Robust bias-corrected inference estimates that bias and accounts for the uncertainty from estimating the correction.

The `rdrobust` package implements bandwidth selection, local polynomial estimation, and robust bias-corrected inference.

An Illustration

Generate independent scores and disturbances:

$$X_i^* \sim U(-2, 2), \quad A_i, u_i \sim N(0, 1).$$

$$Y_i^0 = 10 + 0.5X_i^* + 2A_i + u_i, \quad Y_i^1 = Y_i^0 + 2.$$

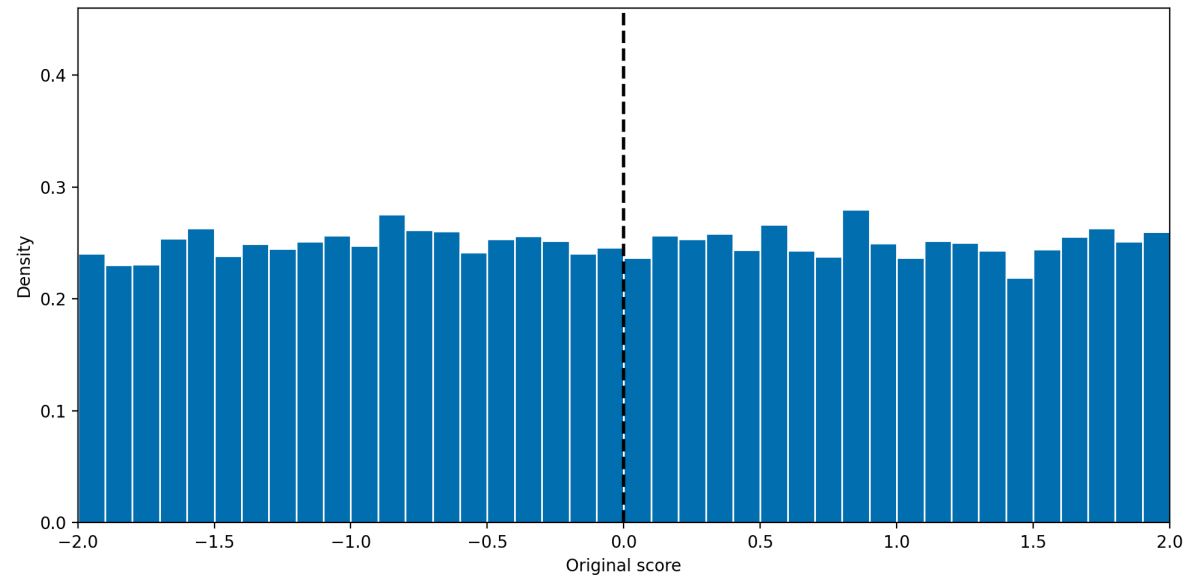
Initially $X_i = X_i^*$ and $D_i = 1 (X_i \geq 0)$. The true effect is **2**.

Now let people with $-0.4 \leq X_i^* < 0$ and $A_i > 0$ report $X_i = -X_i^*$.

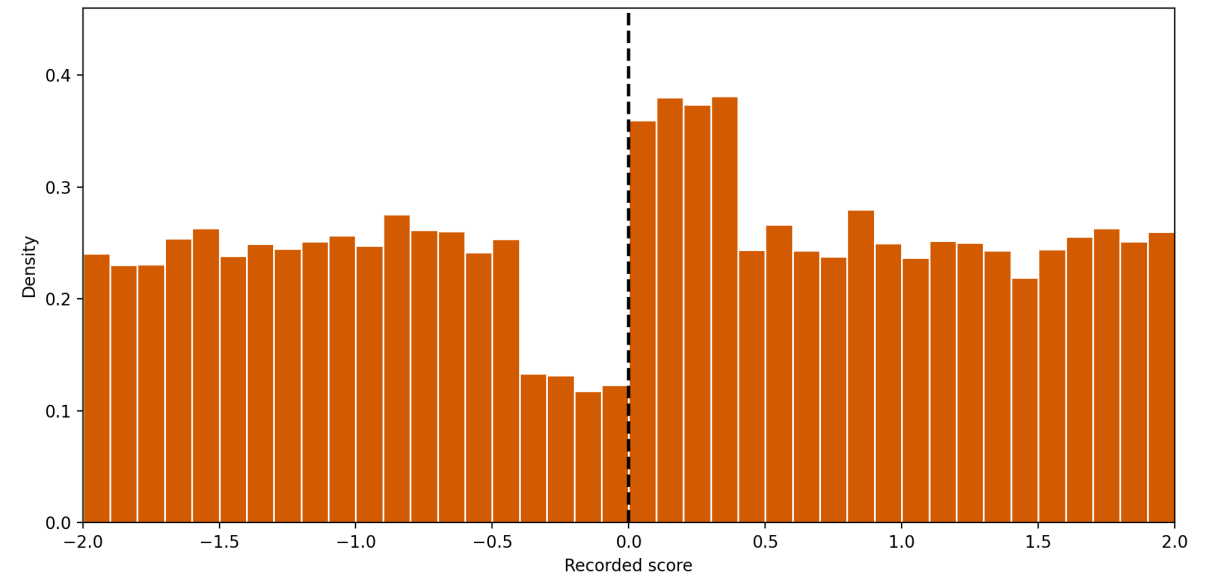
Their ability and untreated outcomes stay the same. Crossing zero gives them treatment. Everyone else keeps their original score.

Before and After Manipulation

Before

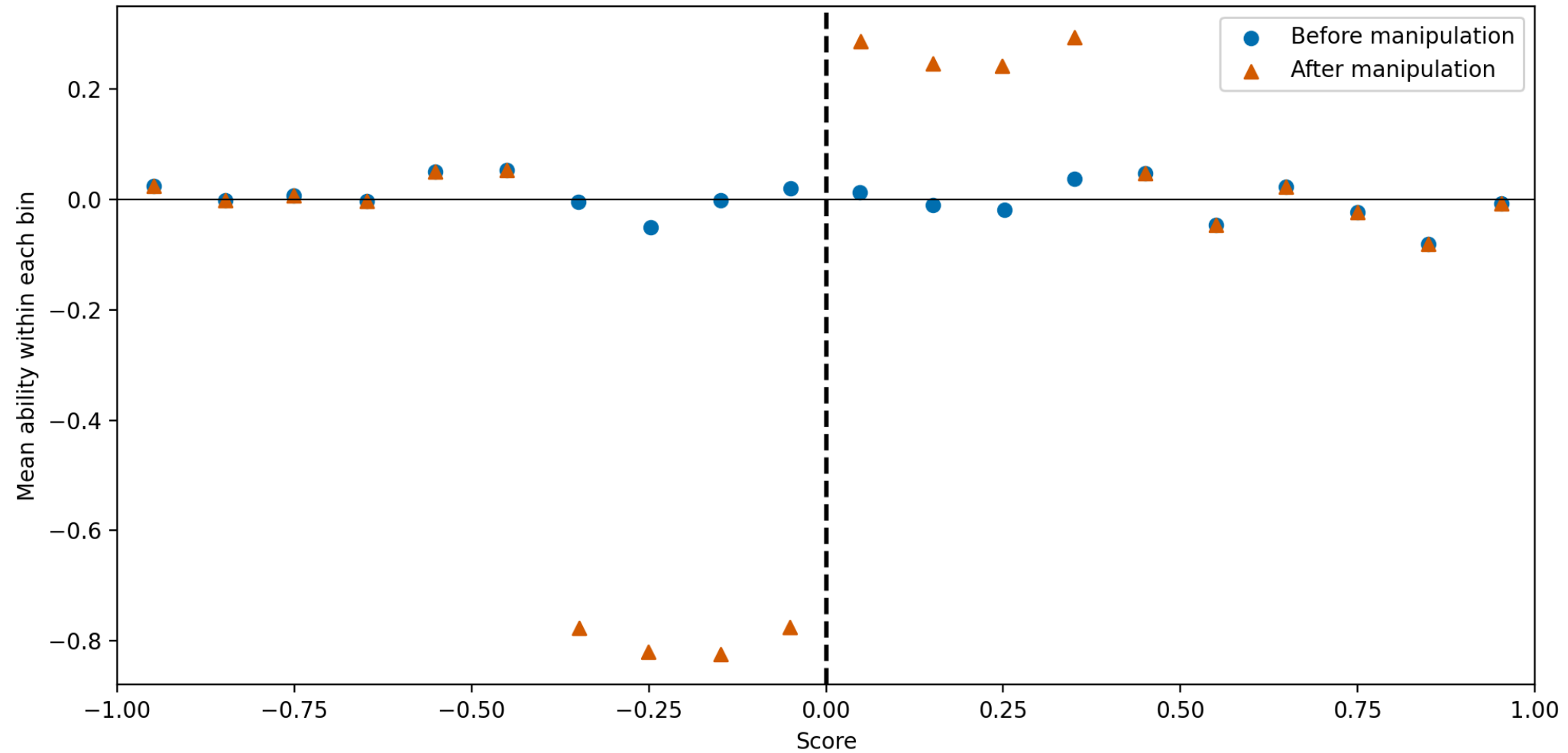


After



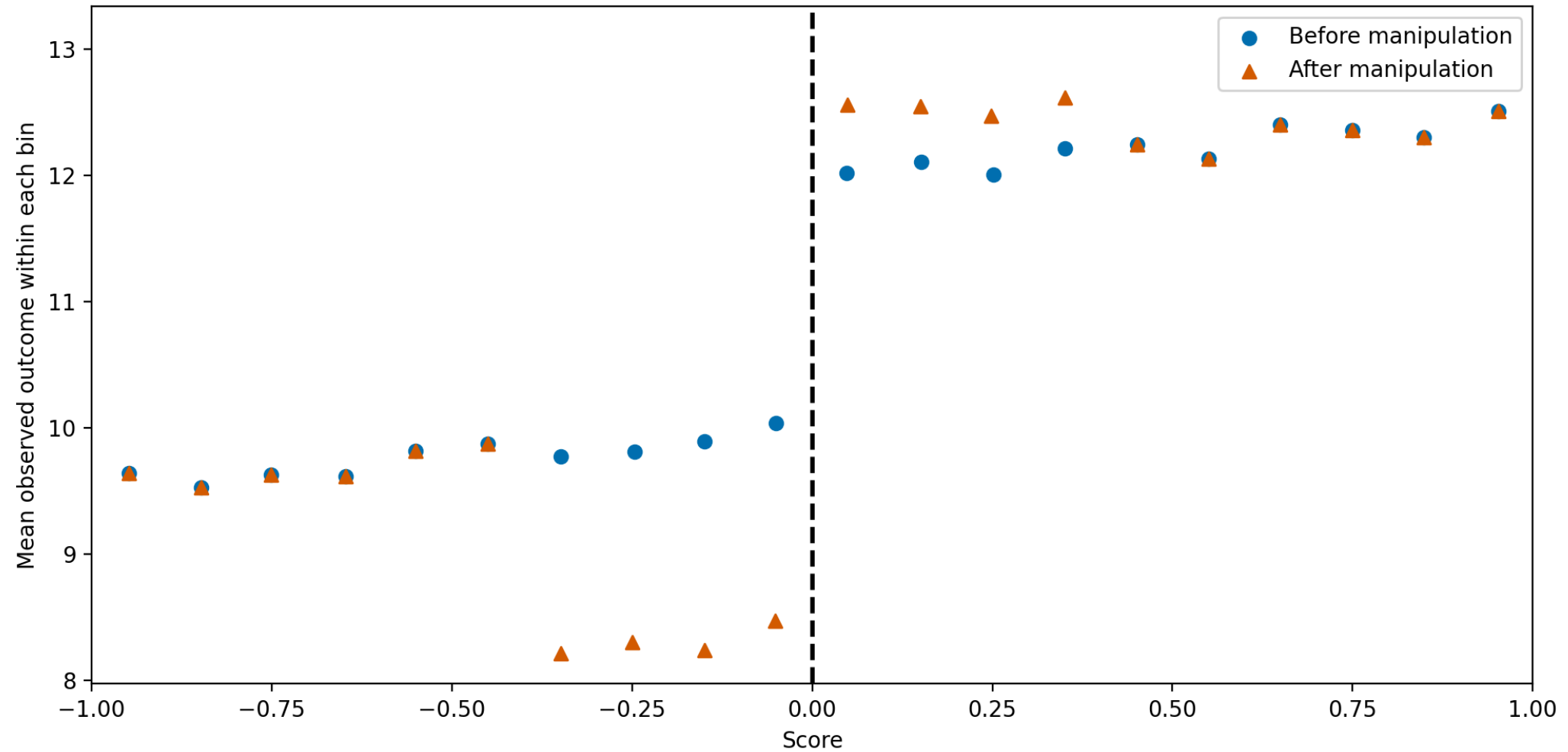
Same people, bins, and vertical scale. Only the recorded scores and resulting treatment change.

The Composition of the Groups Changes



Circles: before manipulation. Triangles: after manipulation. Each point is a bin mean.

The Outcome Jump Changes Too



The true treatment effect remains 2 for every person.

What Does the Density Test Ask?

Let $f_X(x)$ denote the density of the recorded running variable.

The null hypothesis is

$$H_0 : \lim_{x \uparrow x_0} f_X(x) = \lim_{x \downarrow x_0} f_X(x) .$$

A histogram illustrates how observations are distributed. The test estimates the density on each side and assesses the uncertainty in the estimated difference.

The question concerns the **number of observations near the cutoff**, rather than their average outcome.

How the Cattaneo–Jansson–Ma Test Works

1. Construct the **empirical cumulative distribution**: the proportion of observations with a score at or below each value.
2. Fit local polynomials to this distribution on each side of the cutoff, using nearby observations.
3. The **slope** of each fitted curve at the cutoff estimates the density on that side.
4. Test whether the difference between the two estimated densities is zero, allowing for sampling uncertainty and smoothing bias.

The test uses individual observations, with a selected bandwidth and kernel weights. Histogram bins are used only for our descriptive graph.

Density-Test Results and Interpretation

Recorded score	Test statistic	<i>p</i> -value
Before manipulation	−0.697	0.486
After manipulation	6.205	< 0.001

Cattaneo–Jansson–Ma test, 20,000 observations, default bandwidth selection.

- **Reject continuity:** investigate sorting, recording practices and institutional rules.
- **Do not reject:** no statistically detectable density jump. The test's precision matters.

Combine this evidence with predetermined-characteristic checks and the assignment process. Some forms of sorting can leave the density smooth.

Sorting and Identification

For discussion

- In the simulation, explain why the outcome jump changes even though the treatment effect stays at 2.
- What can a researcher learn from the score distribution when ability is unobserved?
- What further evidence would you seek?

Meyersson: Question and Data

For discussion

- Explain the policy question, the unit of observation, and how the 1994 municipal elections are linked to the 2000 census.
- Define the main educational outcome and the population it describes.
- Why might municipalities with Islamic and secular mayors differ even without a causal effect of the mayor?

Meyersson: The Study and Assignment Rule

Does electing an Islamic mayor affect educational attainment?

The study links **1994 Turkish municipal elections** with the **2000 census**. A main outcome is high-school completion among women aged 15–20 in 2000.

The running variable is

$$X_i = \text{largest Islamic party vote share}_i - \text{largest secular party vote share}_i.$$

An Islamic party wins when its margin is positive. The cutoff is **zero**.

Compare municipalities where it narrowly won with those where it narrowly lost.

Meyersson: Identification Strategy

For discussion

- Define the running variable, cutoff, treatment and comparison.
- Explain why the cutoff is zero rather than a 50 percent vote share.
- State the continuity assumption in terms of potential educational outcomes.
- Discuss a plausible threat to identification and explain which municipalities the estimated effect describes.

Meyersson: Figure 4

For discussion

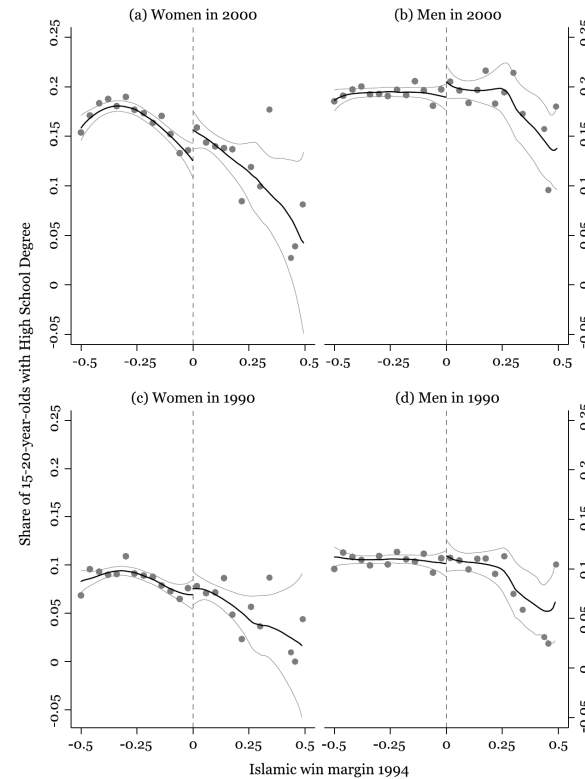


FIGURE 4.—Graphical illustration of RD design: High School Education in 2000 and 1990. The graphs show unconditional means in 4-percent bins for the share of women (graphs (a) and (c)) and men (graphs (b) and (d)), respectively, between 15 and 20 years of age with a high school degree in 2000 (graphs (a) and (b)) and 1990 (graphs (c) and (d)). The solid black line represents the predicted values of a local linear smoother on each side of the threshold at zero. The outer gray lines denote 95 percent confidence intervals.

Meyersson (2014), Figure 4. High school education in 2000 and 1990.

Meyersson: Table II, Panel A

For discussion

TABLE II									
ISLAMIC RULE AND HIGH SCHOOL EDUCATION ^a									
Outcome	Completed High School in 2000								Enrollment
Age Cohort	15–20								15–30
Control Function	None		Linear				Quadratic	Cubic	Linear
Bandwidth	Global		$\hat{h}$	$\hat{h}/2$		$2\hat{h}$	$\hat{h}$	$\hat{h}$	$\hat{h}$
Covariates	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Women									
Outcome mean	0.163	0.163	0.152	0.152	0.144	0.166	0.152	0.152	0.127
Islamic mayor in 1994	−0.026*** (0.006)	0.012** (0.006)	0.032*** (0.010)	0.028*** (0.007)	0.032*** (0.011)	0.022*** (0.006)	0.028*** (0.011)	0.043*** (0.016)	0.014*** (0.005)
Bandwidth	1.000	1.000	0.240	0.240	0.120	0.480	0.240	0.240	0.205
R ²	0.01	0.55	0.03	0.65	0.65	0.58	0.65	0.65	0.48
Observations	2629	2629	1020	1020	589	2049	1020	1020	904
(Continues)									

Meyersson (2014), Table II. Women.

Interpreting the Policy Effect

The comparison identifies the effect of electing an Islamic mayor in municipalities near the electoral threshold.

That treatment changes a bundle of policies, practices and political representation.

For discussion: How far can the results distinguish the mechanisms through which education changes?
To which other municipalities, elections or policy reforms would you apply the findings?

Fuzzy Discontinuity

Crossing the cutoff changes treatment probability without determining treatment completely.

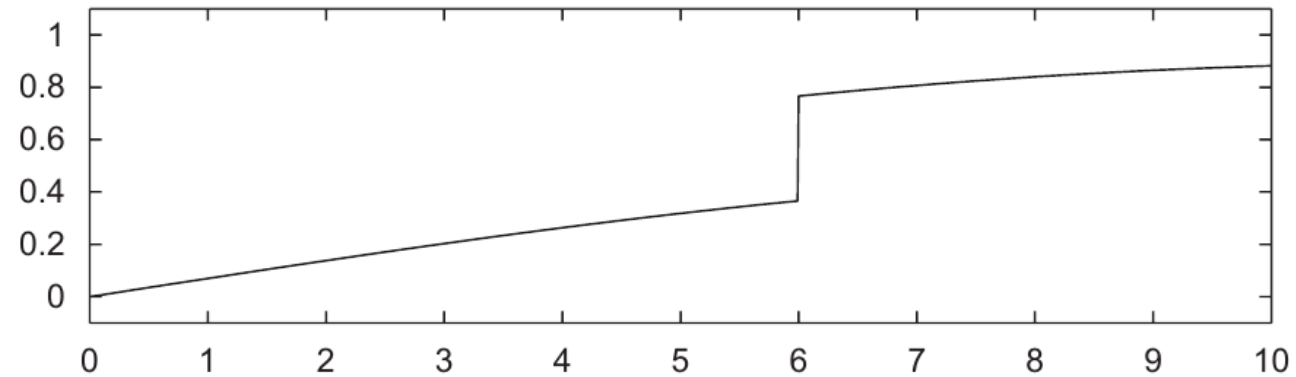


Fig. 3. Assignment probabilities (FRD).

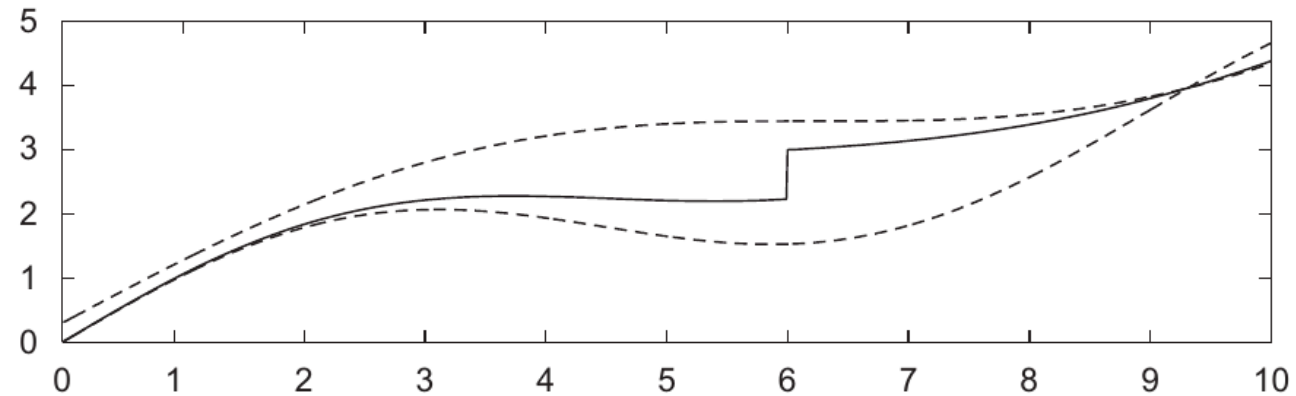


Fig. 4. Potential and observed outcome regression (FRD).

Imbens and Lemieux (2008), Figures 3–4.

The Fuzzy RD Estimand

Divide the outcome discontinuity by the treatment discontinuity:

$$\rho_{FRD} = \frac{\lim_{x \downarrow x_0} E[Y_i | X_i = x] - \lim_{x \uparrow x_0} E[Y_i | X_i = x]}{\lim_{x \downarrow x_0} E[D_i | X_i = x] - \lim_{x \uparrow x_0} E[D_i | X_i = x]}.$$

This is a **local Wald ratio**:

$$\frac{\text{reduced-form jump}}{\text{first-stage jump}}.$$

The denominator must be nonzero and estimated with sufficient precision.

What Makes the Fuzzy RD Ratio Causal?

For a binary treatment, the interpretation is a LATE at the cutoff for people whose treatment changes when eligibility changes.

We need:

- Continuity of potential outcomes and treatment behaviour through the cutoff, apart from the eligibility change.
- A nonzero first-stage discontinuity.
- Exclusion: eligibility affects the outcome through treatment.
- Monotonicity: eligibility changes treatment in the same direction for everyone at the cutoff.

Retain the well-defined-treatment and no-interference assumptions.

Fuzzy RD as Two-Stage Least Squares

Set $R_i = X_i - x_0$ and $Z_i = 1 (R_i \geq 0)$. Locally, write the first stage as

$$D_i = \pi_0 + \pi_1 Z_i + \pi_2 R_i + \pi_3 Z_i R_i + v_i.$$

The outcome equation is

$$Y_i = \beta_0 + \rho D_i + \beta_1 R_i + \gamma_1 Z_i R_i + \epsilon_i.$$

Use Z_i as the **excluded instrument** for D_i .

Include R_i and $Z_i R_i$ as controls in both equations. They allow separate slopes on the two sides.

Report the first-stage jump. A weak jump creates the denominator problem from Week 5.

Angrist and Lavy: Question and Data

For discussion

- Explain the class-size policy question, the data and the outcomes.
- Describe Maimonides' rule and illustrate what happens to predicted class size when enrolment rises from 40 to 41.
- Why might an OLS relationship between class size and achievement give a misleading estimate of the causal effect?

Angrist and Lavy (1999): Maimonides' Rule

The rule limits classes to 40 pupils. As grade enrolment crosses a multiple of 40, the implied number of classes increases.

For enrolment e , predicted class size is

$$m(e) = \frac{e}{\lfloor (e - 1) / 40 \rfloor + 1}.$$

For example, 40 pupils imply one class of 40. With 41 pupils, the rule implies two classes averaging 20.5.

Actual class size does not follow the rule perfectly, creating a fuzzy design.

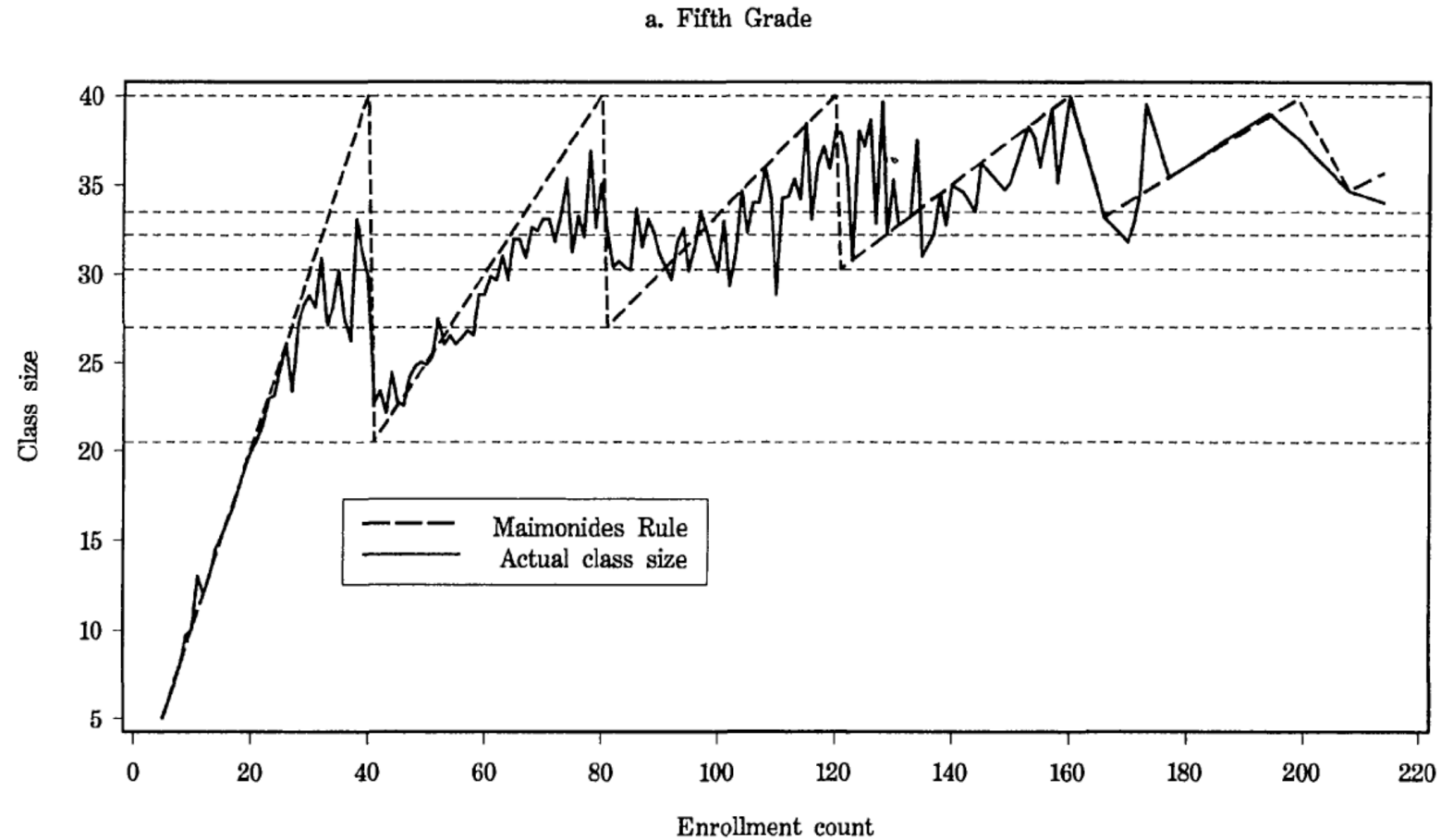
Angrist and Lavy: Identification Strategy

For discussion

- Use the class-size and achievement graphs to explain the first stage and reduced form.
- Explain why this is a fuzzy design.
- Discuss relevance, exclusion, local comparability and the direction of the class-size change induced by the rule.
- Could another school input change at the enrolment threshold?

Actual and Predicted Class Size

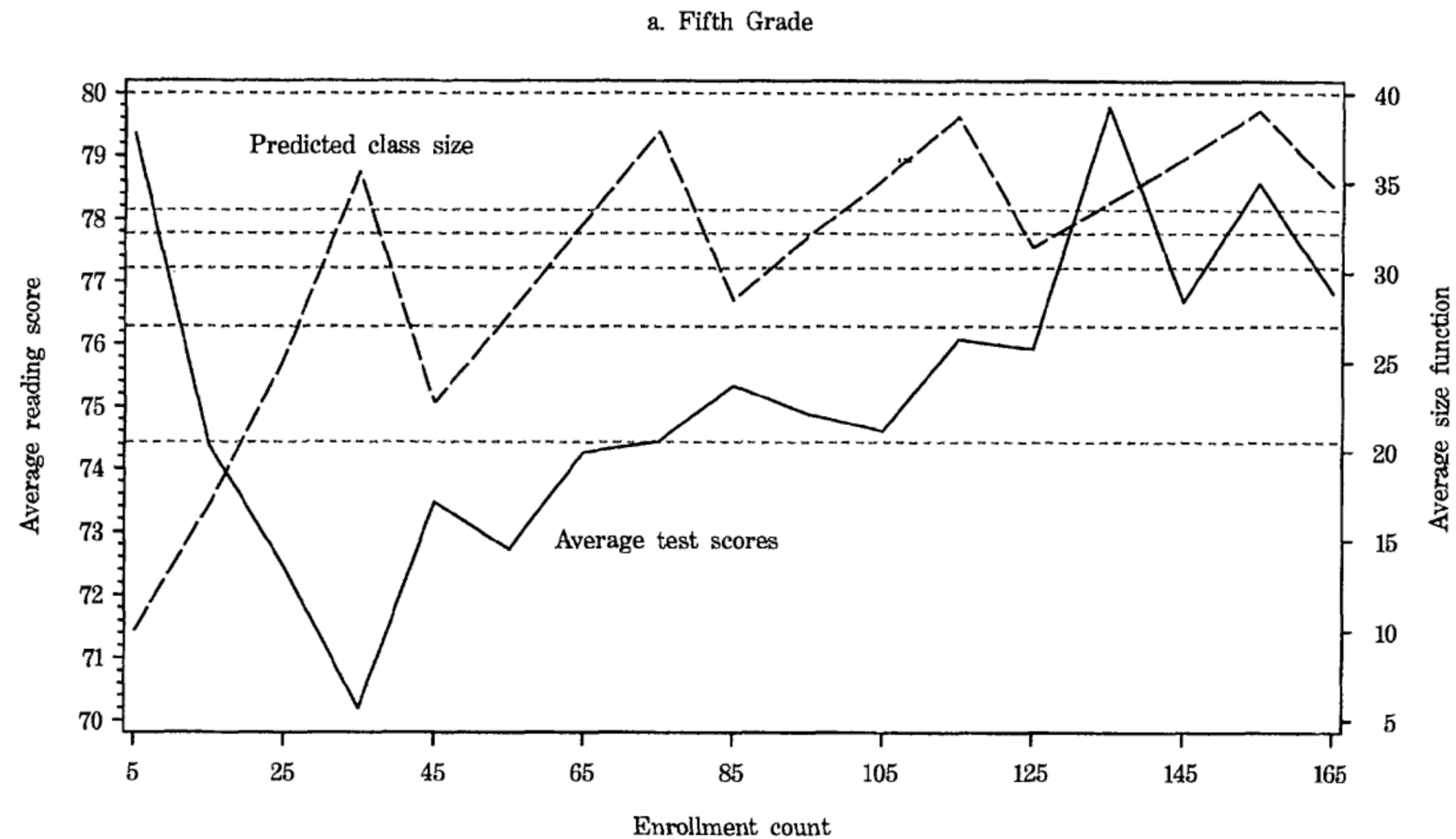
For discussion



Angrist and Lavy (1999), class-size figure from the original slides.

The Reduced Form

For discussion



Angrist and Lavy (1999), achievement and enrolment figure from the original slides.

Angrist and Lavy: Table IV

For discussion

TABLE IV 2SLS ESTIMATES FOR 1991 (FIFTH GRADERS)												
	Reading comprehension						Math					
					+/- 5 Discontinuity sample						+/- 5 Discontinuity sample	
	Full sample						Full sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Mean score</i>		74.4			74.5			67.3			67.0	
<i>(s.d.)</i>		(7.7)			(8.2)			(9.6)			(10.2)	
<i>Regressors</i>												
Class size	-.158 (.040)	-.275 (.066)	-.260 (.081)	-.186 (.104)	-.410 (.113)	-.582 (.181)	-.013 (.056)	-.230 (.092)	-.261 (.113)	-.202 (.131)	-.185 (.151)	-.443 (.236)
Percent disadvantaged	-.372 (.014)	-.369 (.014)	-.369 (.013)		-.477 (.037)	-.461 (.037)	-.355 (.019)	-.350 (.019)	-.350 (.019)		-.459 (.049)	-.435 (.049)
Enrollment		.022 (.009)	.012 (.026)			.053 (.028)		.041 (.012)	.062 (.037)			.079 (.036)
Enrollment squared/100			.005 (.011)						-.010 (.016)			
Piecewise linear trend				.136 (.032)						.193 (.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.

Angrist and Lavy (1999), Table IV. IV results.

Maimonides' Rule Redux

For discussion

- Explain what the newer data allow the authors to examine and why manipulation of reported enrolment matters.
- Describe the alternative enrolment measure based on birthdays.
- Use Table 2 to discuss the first stage, achievement estimates and uncertainty.
- How does the follow-up affect your interpretation of the original findings?

Maimonides' Rule Redux (2019)

The follow-up uses much larger Israeli samples from **2002–2011**.

- The rule still predicts class size, but the achievement estimates are close to zero.
- Reported enrolment shows manipulation near the thresholds.
- An alternative instrument uses enrolment imputed from birthdays.

Compare the estimates and confidence intervals with the original findings. For class size, interpret the effect **per pupil**, rather than as a binary-treatment LATE.

What does the follow-up teach us about the original design and its policy interpretation?

Redux: Table 2

For discussion

TABLE 2—CLASS SIZE EFFECTS ESTIMATED USING BIRTHDAY-BASED IMPUTED ENROLLMENT (2002–2011)

	Language				Math			
	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	OLS (5)	2SLS (6)	2SLS (7)	2SLS (8)
Class size	0.0070 (0.0172)	0.0012 (0.0608)	−0.0089 (0.0666)	−0.0061 (0.0657)	0.0386 (0.0231)	−0.0366 (0.0814)	−0.0623 (0.0889)	−0.0616 (0.0878)
SES index	−0.4254 (0.0602)	−0.4263 (0.0610)	−0.4246 (0.0609)	−0.4252 (0.0609)	−0.3570 (0.0799)	−0.3680 (0.0809)	−0.3636 (0.0810)	−0.3644 (0.0809)
Birthday-based enrollment	0.0033 (0.0035)	0.0038 (0.0060)	0.0163 (0.0184)		0.0037 (0.0046)	0.0099 (0.0081)	0.0418 (0.0239)	
Enrollment squared/100			−0.0068 (0.0086)				−0.0173 (0.0109)	
Piecewise linear trend				0.0113 (0.0151)				0.0318 (0.0203)
Observations		225,108				226,832		

Notes: This table reports OLS and 2SLS estimates of equation (3). The endogenous variable is June class size; Maimonides' rule is constructed using birthday-based imputed enrollment. Standard errors reported in parentheses are clustered at the school and year level. The dependent variable is a math or language test score. Additional covariates include student characteristics (a gender dummy, parents' years of schooling, number of siblings, a born-in-Israel indicator, and ethnic-origin indicators), year fixed effects, an indicator for religious schools, the school socioeconomic index, and interactions of the socioeconomic index with dummies for the 2002–2003 and 2008–2011 periods. The reported SES coefficient is for 2004–2007.

Angrist, Lavy, Leder-Luis and Shany (2019), Table 2. Birthday-based imputed enrolment.

References

- Jacob, R., P. Zhu, M.-A. Somers, and H. Bloom (2012). *A Practical Guide to Regression Discontinuity*. MDRC.
- Imbens, G. W., and T. Lemieux (2008). “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics* 142(2): 615–635.
- Lee, D. S. (2008). “Randomized Experiments from Non-random Selection in U.S. House Elections.” *Journal of Econometrics* 142(2): 675–697.
- Meyersson, E. (2014). “Islamic Rule and the Empowerment of the Poor and Pious.” *Econometrica* 82(1): 229–269.
- Angrist, J. D., and V. Lavy (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114(2): 533–575.
- Angrist, J. D., V. Lavy, J. Leder-Luis, and A. Shany (2019). “Maimonides’ Rule Redux.” *American Economic Review: Insights* 1(3): 309–324.
- McCrary, J. (2008). “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics* 142(2): 698–714.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). “Simple Local Polynomial Density Estimators.” *Journal of the American Statistical Association* 115(531): 1449–1455.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica* 82(6): 2295–2326.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). “Regression Discontinuity Designs Using Covariates.” *Review of Economics and Statistics* 101(3): 442–451.
- Imbens, G., and K. Kalyanaraman (2012). “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *Review of Economic Studies* 79(3): 933–959.
- Gelman, A., and G. Imbens (2019). “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs.” *Journal of Business & Economic Statistics* 37(3): 447–456.
- Lee, D. S., and T. Lemieux (2010). “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48(2): 281–355. **Advanced reading.**