Regression Discontinuity

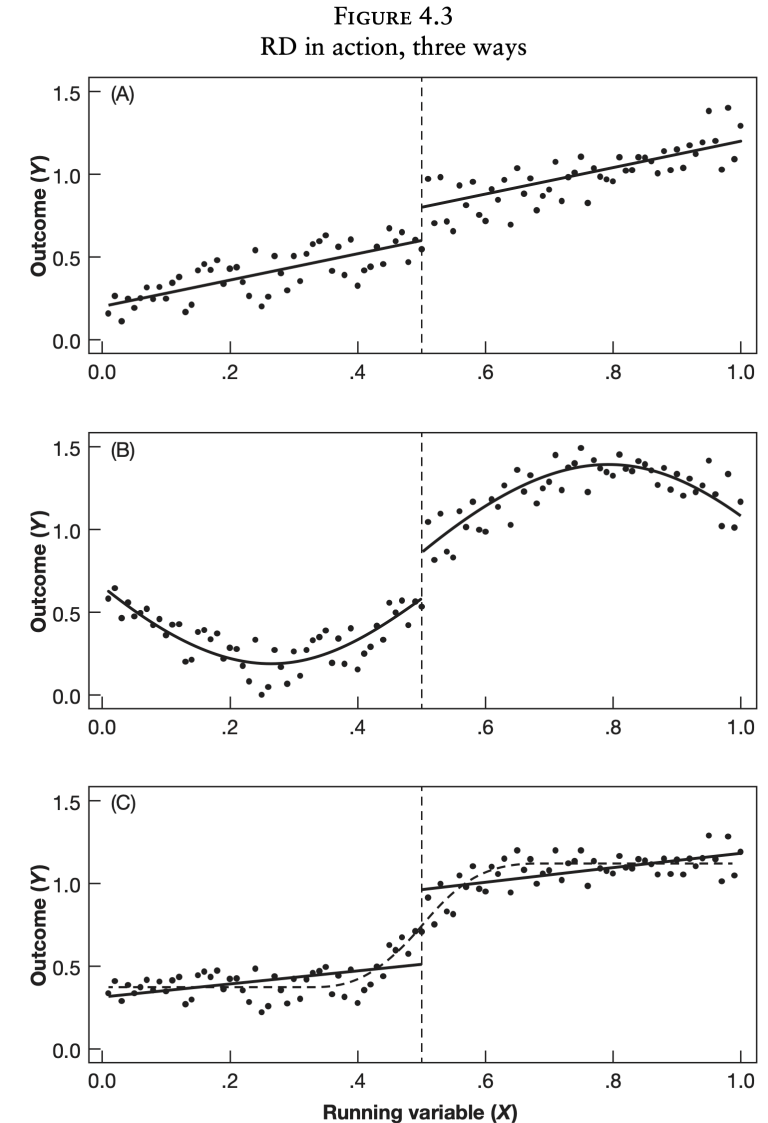## Part II

Fernando Hoces la Guardia

08/02/2022

Today's Lecture

- Finish Sharp RDD
  - Non-linearities
  - Interpreting results
- Start Fuzzy RDD

But Is There a Jump?

- The key question to identify causality, is whether relationship between running variable and outcome is well represented by a linear control on age.
- Two approaches to reduce the likelihood of mistakes when modeling this relationship: (i) modeling non-linear relationships, and (ii) focusing only on data around the cut-off. We will spend most of the time in (i).
- In addition to logs, non-linearities can be modeled with two additional tools: polynomials and interactions.

FIGURE 4.3
RD in action, three ways

Notes: Panel A shows RD with a linear model for $E[Y_i|X_i]$; panel B adds

Modeling Non-Linear Relationships: Polynomials

- Curves are usually modeled using polynomials (powers of the regressors).
- Higher polynomials (higher powers) introduce more flexibility but they are also likely to hide a disconitinuity when there is one.
- The choice of how much more flexibility is enough is a judgment call.
- Ideally the results should not vary much as you add higher order polynomials (powers of 3, 4 or more).
- In our example there might be a small curvature in the data, so we add a quadratic term for the running variable:

$$\overline{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

- We are not interested interpreting the effect of age, only on controlling for any non-linear behaviour.

Modeling Non-Linear Relationships: Interactions 1/3

- An interaction is defined as the multiplication of two regressors. Where typically one is a binary regressor.

- Adding an interaction in any regression (or any equation) is a way of capturing changes in (regression) coefficients change for certain groups.

  - Example with just a constant
  - Example with constant and slope
  - Example with both.

- In here we add an interaction and standardize the running variable, so rho can continue to be interpreted as the difference of average outcomes at the cutoff.

- The standardization part might add some confusion, so first let's focus only on adding the interaction to capture a potential shift in the slope that connects age (a) with mortality rates ($\overline{M}_a$):

$$\overline{M}_a = \alpha + \rho D_a + \gamma a + \delta a \times D_a + e_a$$

- The goal of the standardization is to have an easy interpretation of ρ as the difference of mortality around the cut-off. We could define the a new variable $\tilde{a} = a - 21$ which would represent the standardized age $(a - 21)$. This would give us the regression:

$$\overline{M}_a = \alpha + \rho D_a + \gamma \tilde{a} + \delta \tilde{a} \times D_a + e_a$$

- A more generic version would allow for the cut-off to be any number so instead of 21, put a_0. Giving us the standardized formulation of the book:

$$\overline{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta(a - a_0) \times D_a + e_a$$

- The most important part here is understanding the interactions, if you find the standardization distracting, focus on the first two equations but make sure to remember that "we standardize to be able to interpret ρ as the treatment effect"

- (If we want to extrapolate effects awway from the cut-off, we need to be aware that the treatment effect is $\rho + \delta(a - a_0)$)

Non-Linear Relationships: Interactions And Polynomials

Here are polynomials:

$$\overline{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

Here are polynomials:

$$\overline{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

Here are interactions:

$$\overline{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta(a - a_0) \times D_a + e_a$$

Here are polynomials:

$$\overline{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + e_a$$

Here are interactions:

$$\overline{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta(a - a_0) \times D_a + e_a$$
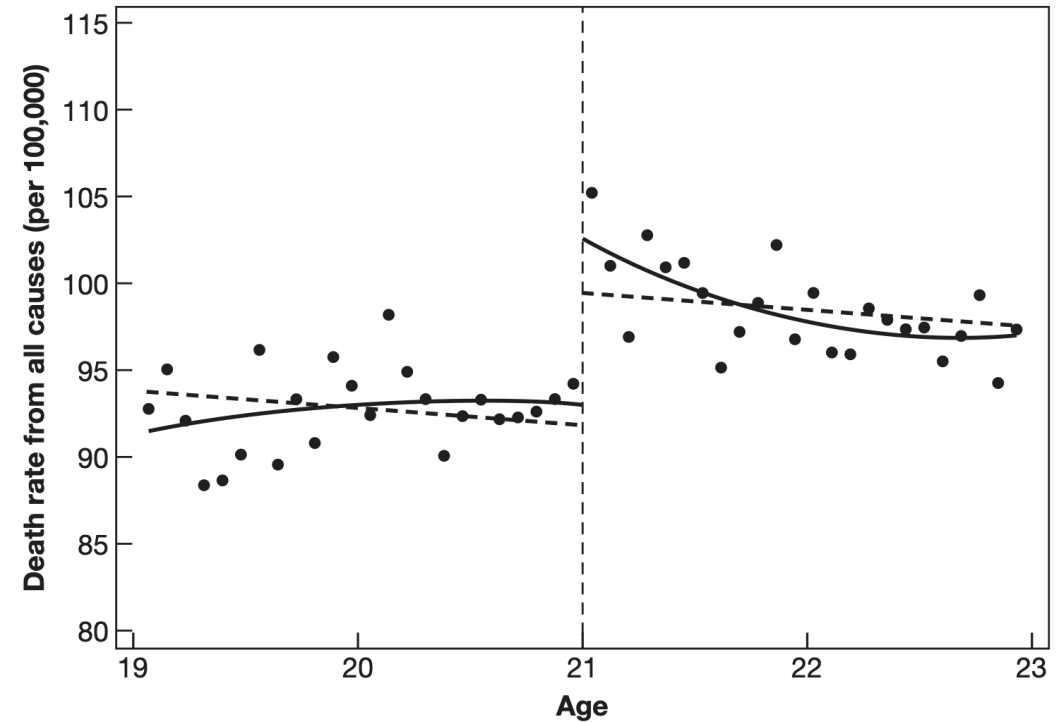
Here are combined:

$$\overline{M}_a = \alpha + \rho D_a + \gamma_1(a - a_0) + \gamma_2(a - a_0)^2 +$$
$$\delta_1\left[(a - a_0)D_a\right] + \delta_2\left[(a - a_0)^2 D_a\right] + e_a$$

We can now capture curvature and changing slopes in the relationship between a and $(\overline{M}_a)$, reducing the risk that we incorrectly find a discontinuity where there is none (figure 4.3-C).

- Effect of 21st birthday seems robust to this new specifications.

- Effect also persist substantially up to the 23rd birthday suggesting lasting effects.

- This last point demonstrates the value of a visual inspection of RDD estimates.

FIGURE 4.4
Quadratic control in an RD design

Notes: This figure plots death rates from all causes against age in months. Dashed lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months. The solid lines plot fitted values from a regression of mortality on an over-21 dummy and a quadratic in age, interacted with the over-21 dummy (the vertical dashed line indicates the minimum legal drinking age [MLDA] cutoff).

TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22		Ages 20–21	
	(1)	(2)	(3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

Now All in One Table

TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22		Ages 20–21	
	(1)	(2)	(3)	(4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related	44	80	74	1.03

Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

Non-Paramteric RDD

- The second way in which can handle non-linearities is by removing parametrical assumptions (about the slopes and how they change).
- This involves either taking simple averages, or computing linear regressions but only arround on a narrow bandiwth around the cut-off.
- This approach does not have the problems trying to get the relationship between a and $(\overline{M}_a)$ right, but it discard a large amount of data (information).
- The main challenge is how to choose the bandwidth to balance the trade of between bias (incorrectly attributing discontinuities) and variance (due to smaller sample size). The choice of this bandwidth is a judgement call, and results should not rely on one specific choice.
- It also has several "fancy" (more complex) methodological challenges that we ignore for now.

Fuzzy RDD

(Same content as MM 4.2, but different order)

- More specifically: Do students that attend the best exam school in Boston (Boston Latin School, or BLS) perform better because of having better performing peers? This potential effect is known as "peer effect" in the (academic) literature.

- **Outcome** (Y_i)**:** Math Score in 7th and 8th Grade (1 or 2 years after entering the exam school). Standardized.

- **Treatment** $(\overline{X}_{(i)})$**:** Average score of peers before entering the exam school (4th grade). Proxy measure of peer quality.

- **Running Variable** (R_i)**:** Score in entrance exam, measured as distance to BLS cut-off threshold.

- **Discontinuity**: Crossing the eligibility threshold in entrance exame for elite school in Boston (BLS).
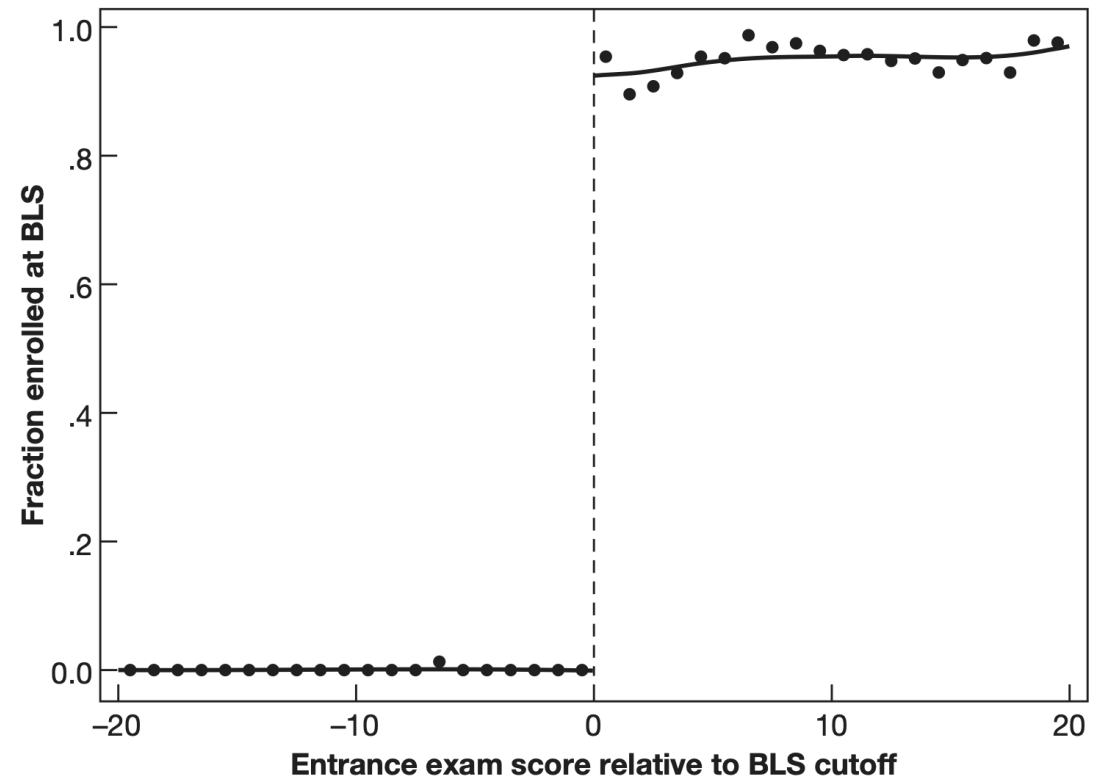
- Regression:

$$Y_i = \theta_0 + \theta_1 \overline{X}_{(i)} + \theta_2 X_i + u_i$$

- OLS Regression estimates for $\theta_1 = 0.25$

- Selection problem: these schools are by definition selecting the best students, so comparisons between peers in exam school versus the rest will be contaminated by selection bias.

- Let's use RDD to address this selection bias problem.

- Let's start with **a** discontinuity.

- Enrollment to BLS and distance from exam cut-off

- This is not the discontinuity will end up focusing on, but it helps to illustrate the concept of Fuzzy RDD and how it connects with the notion of compliance.
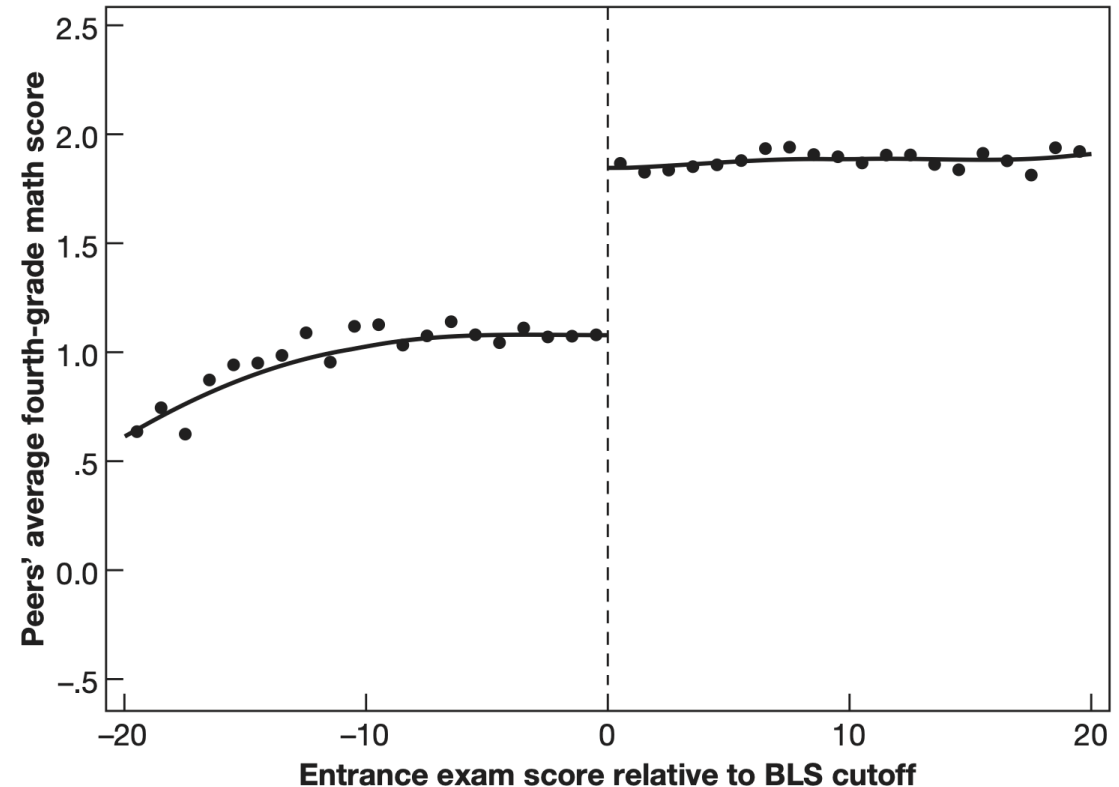
Notes: This figure plots enrollment rates at Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the BLS admissions cutoff. Solid lines show fitted values from a local linear regression estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- Now let's switch to figure 4.8, which shows the treatment we care about (peer effects) as a function of the running variable.
- (compliers here are harder to describe: "those with peers who notably improve their performance after crossing the BLS threshold")
- The **instrument** here is defined as the variable that captures crossing the threshold.
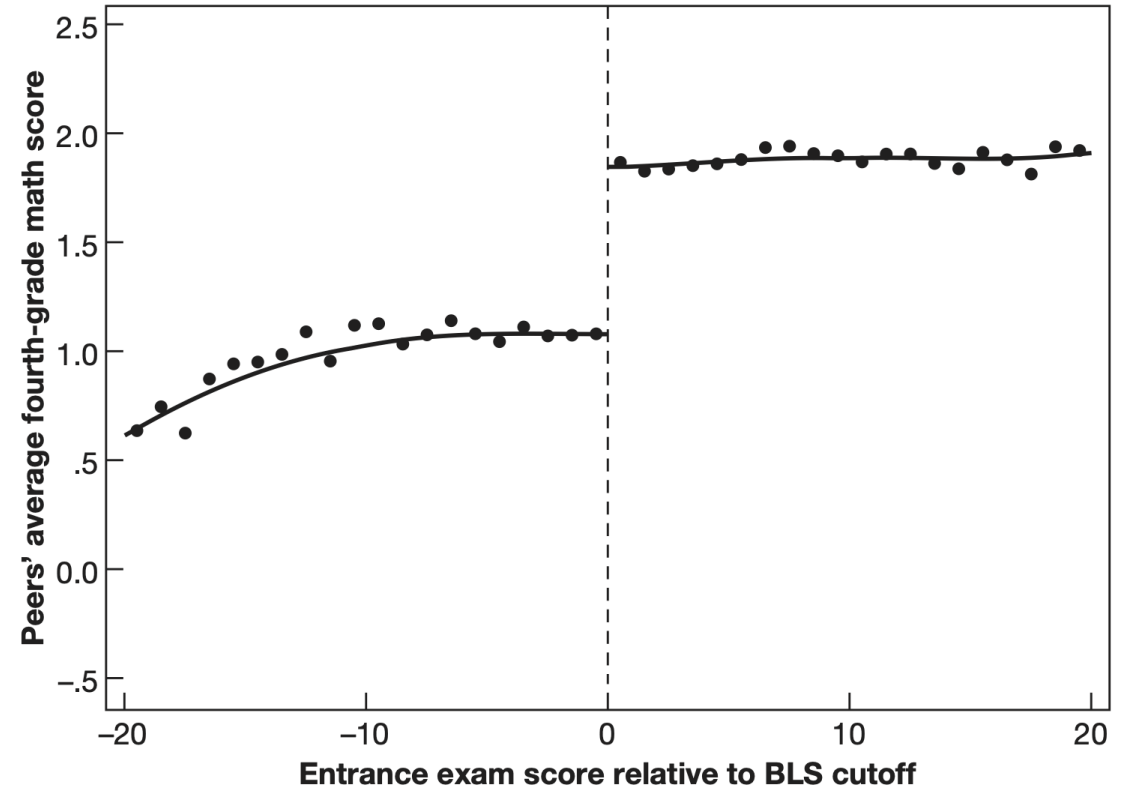
FIGURE 4.8
Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- Beware of confusions: in sharp RDD this variable represents the treatment, in fuzzy represents the instrument (akin to the offers in KIPP and OHP examples).
- To add to the confusion the instrument here is labeled as D_i (instead of Z_i)
- If this is the instrument, what is the first stage?

FIGURE 4.8
Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- First Stage:

$$\overline{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i}$$

- Reduced Form:

- First Stage:

$$\overline{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i}$$

- Reduced Form:

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}$$

- Second Stage (for 2SLS):

- First Stage:

$$\overline{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i}$$

- Reduced Form:

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}$$
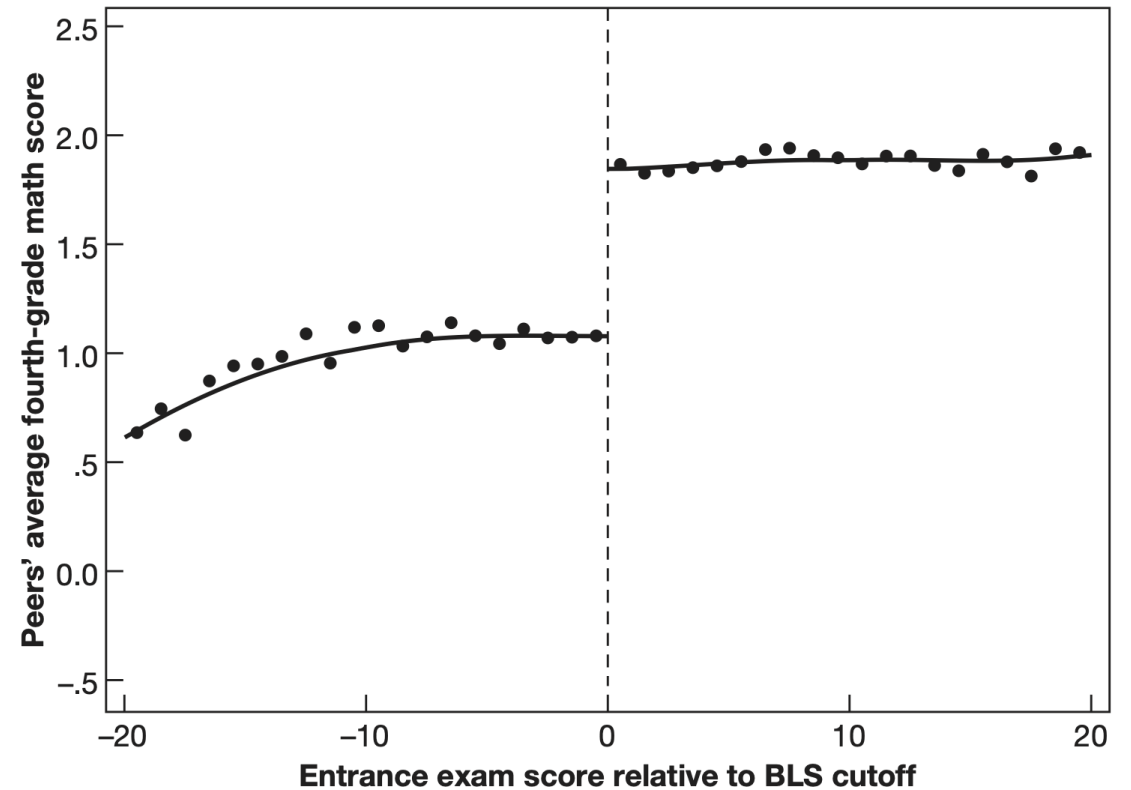
- Second Stage (for 2SLS):

$$Y_i = \alpha_2 + \lambda \widehat{\overline{X}}_{(i)} + \beta_2 R_i + e_{2i}$$
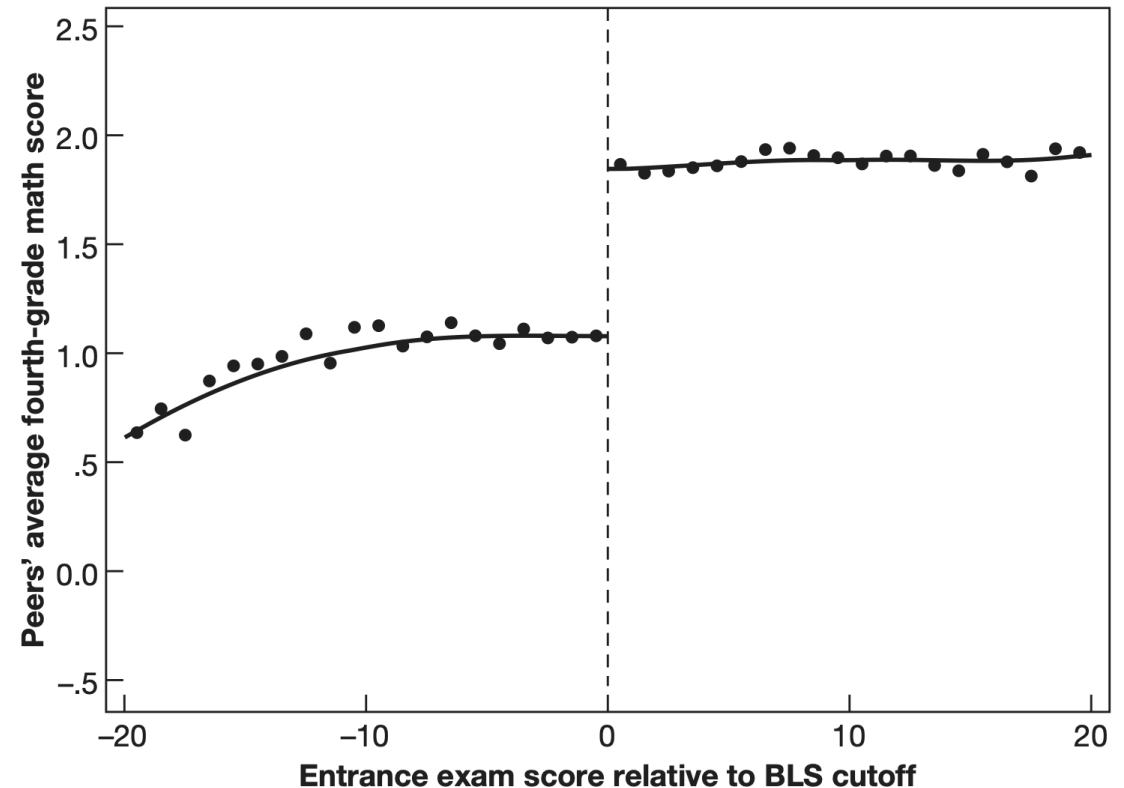
FIGURE 4.8
Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- **Relevancy:** See figure 4.8. Effect of instrument on treatment is an increase in 0.8σ (very big)

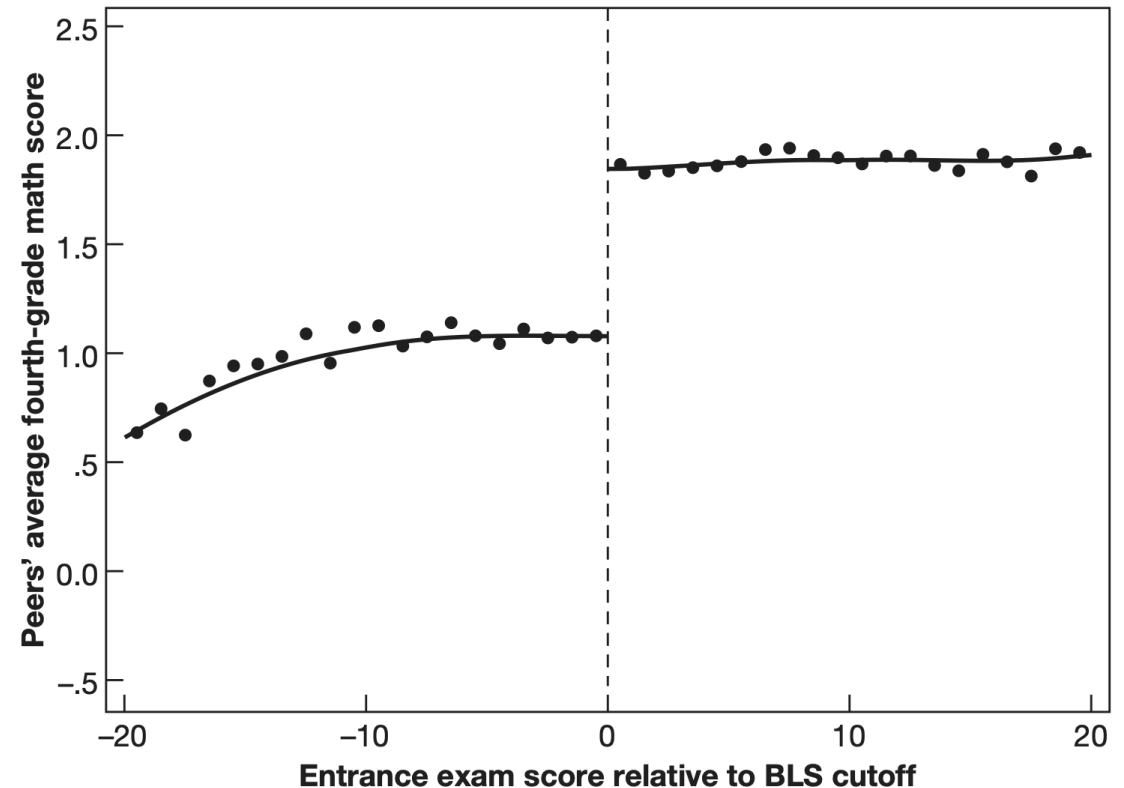FIGURE 4.8
Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- **Relevancy:** See figure 4.8. Effect of instrument on treatment is an increase in 0.8σ (very big)

- **Independence:** Yes for the same reason that Sharp RDD does not have OVB: Instrument is a deterministic function of a running variable.

FIGURE 4.8
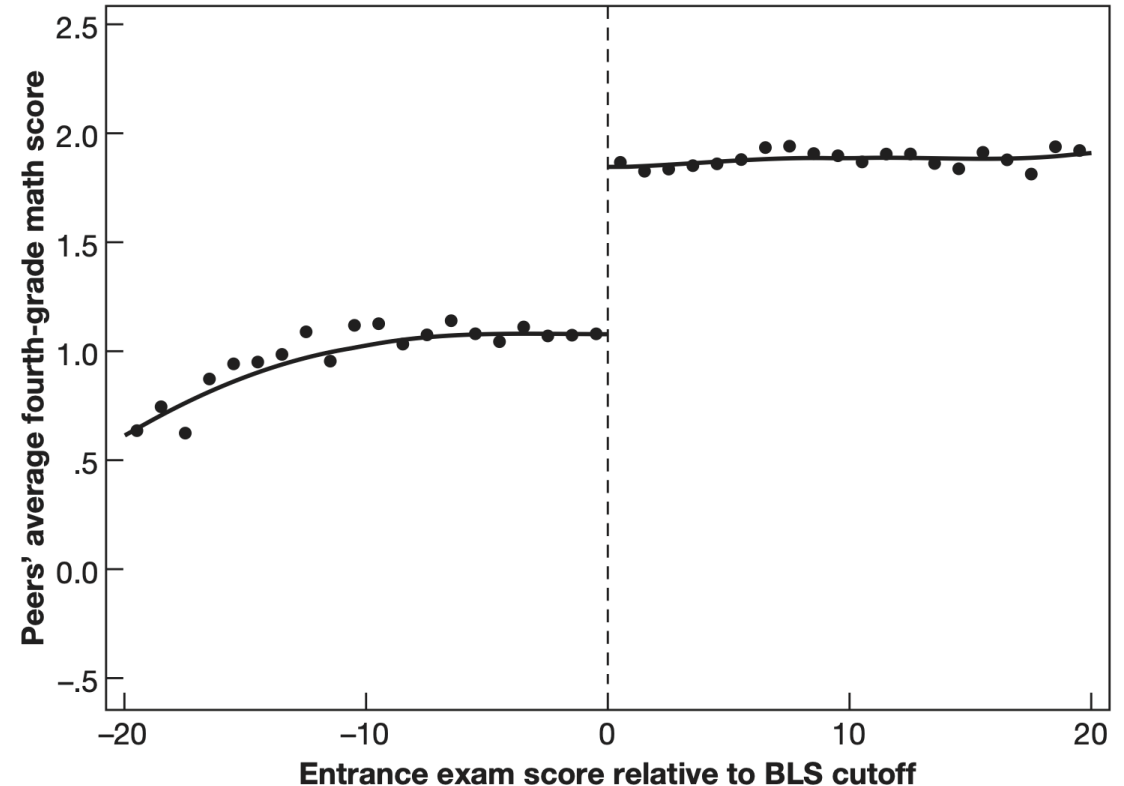Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- **Relevancy:** See figure 4.8. Effect of instrument on treatment is an increase in 0.8σ (very big)

- **Independence:** Yes for the same reason that Sharp RDD does not have OVB: Instrument is a deterministic function of a running variable.

- **Exclusion (Restriction):** the cut-off variable (instrument) is influencing the math scores (outcome) only through peer quality (treatment). There are probably other channels, so this assumption probably doesn't hold.
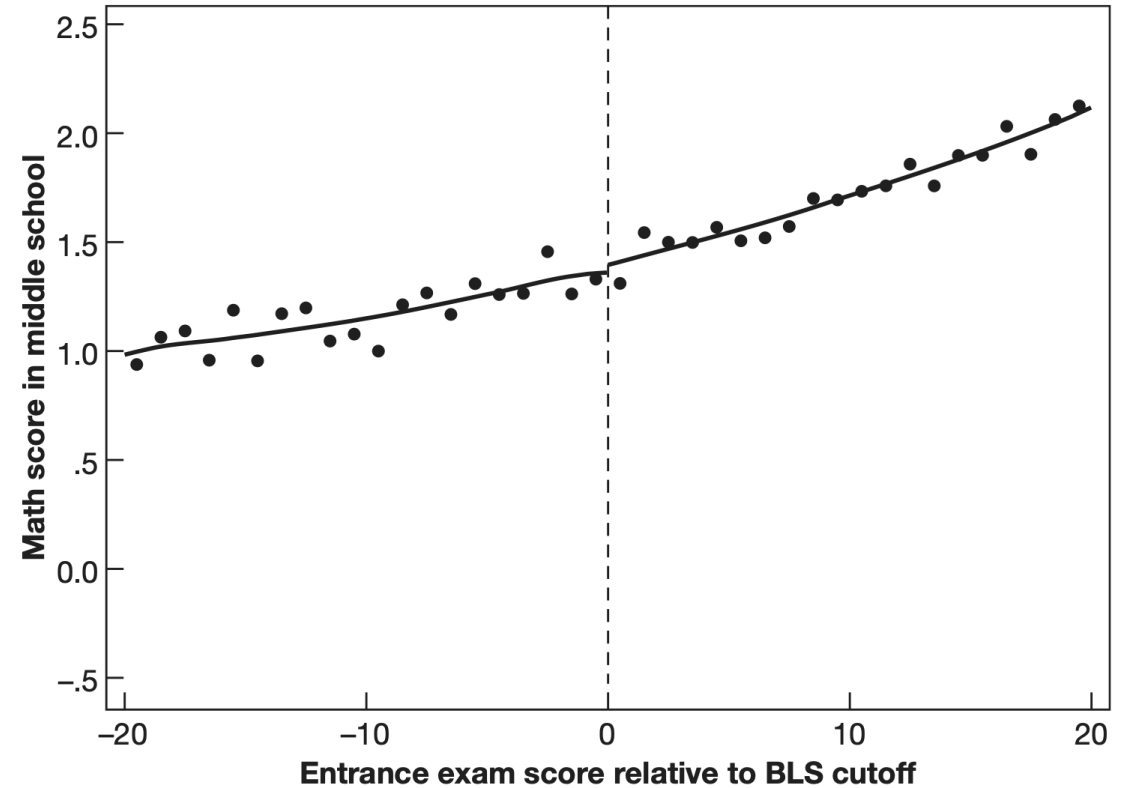
FIGURE 4.8
Peer quality around the BLS cutoff

Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

- First Stage: $\phi = 0.8$
  (no SE reported). Strong first stage.

FIGURE 4.9
Math scores around the BLS cutoff

Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Results

- First Stage: $\phi = 0.8$
  (no SE reported). Strong first stage.

- Reduced Form: $\rho = -0.02$
  (SE = 0.1). Statistical zero.
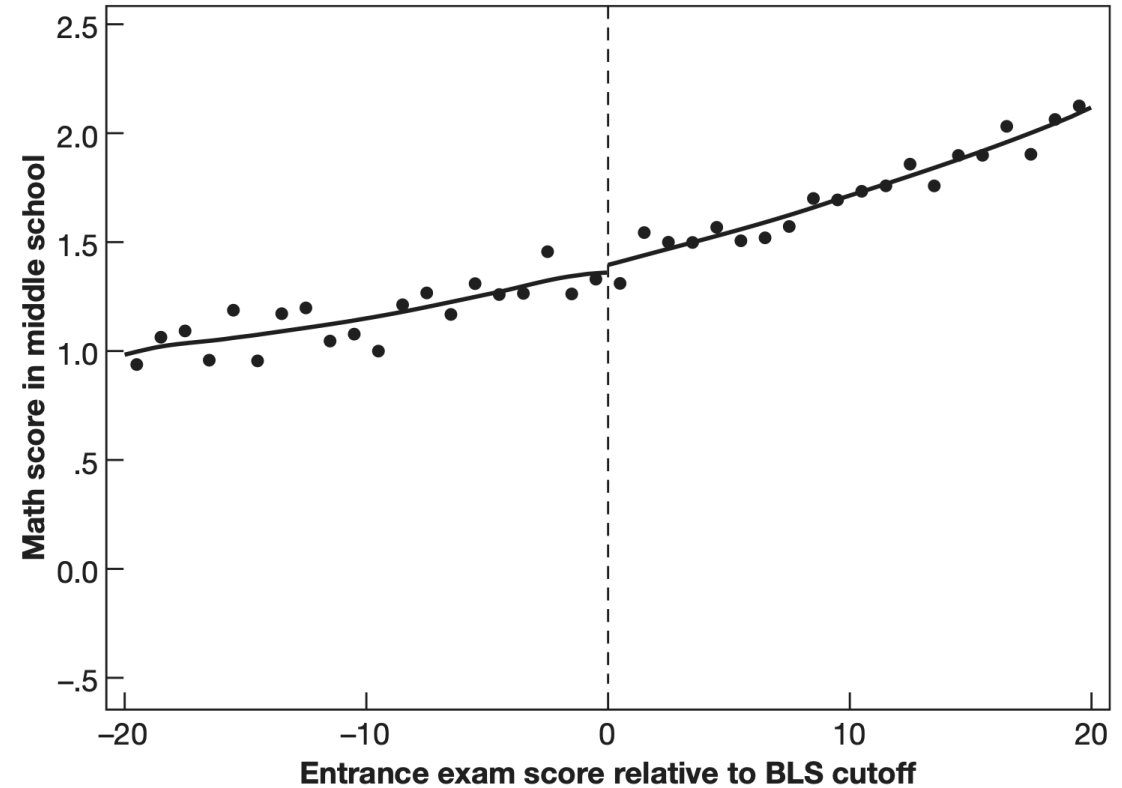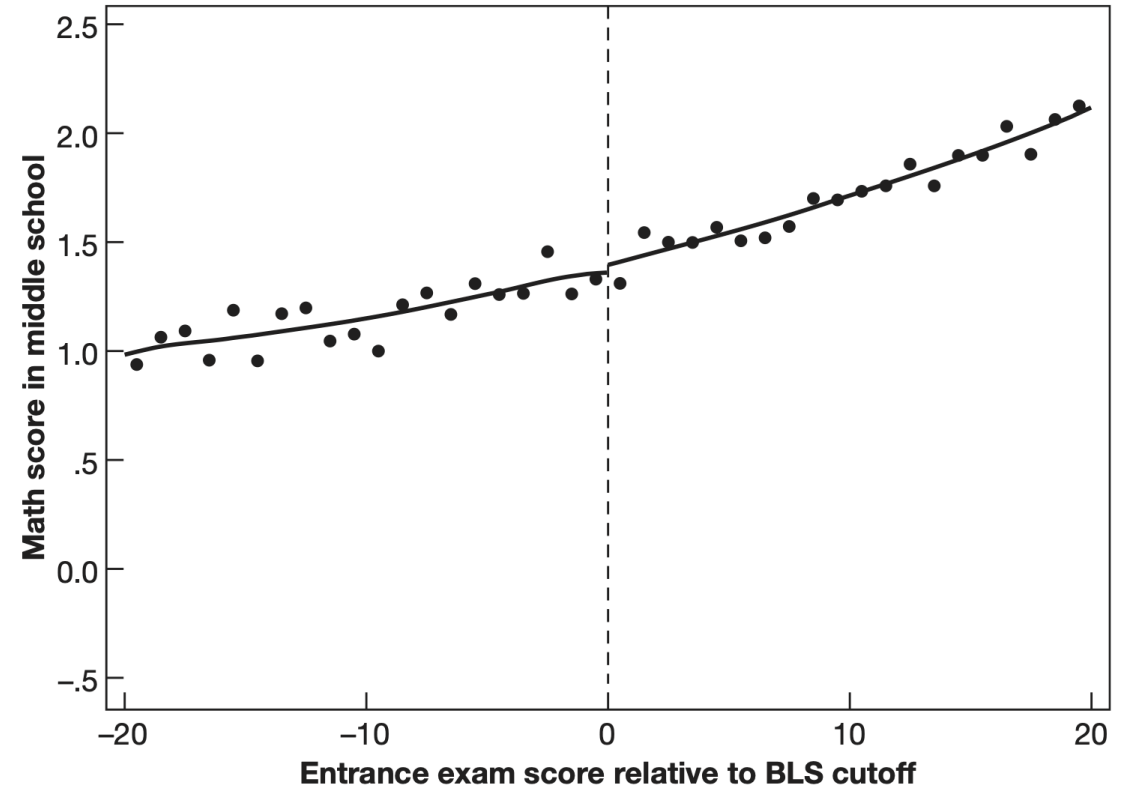
FIGURE 4.9
Math scores around the BLS cutoff

Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Results

- First Stage: $\phi = 0.8$
  (no SE reported). Strong first stage.

- Reduced Form: $\rho = -0.02$
  (SE = 0.1). Statistical zero.

- 2SLS LATE: $\lambda = -0.023$
  (SE = 0.132). Zero again.

- OLS: $\theta_1 = 0.25$
  (no SE reported). Strong positvie "effect"

Math scores around the BLS cutoff

Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Back to the Exclusion Restriction

- We saw that the exclusion assumption probably doesn't hold, so why bother with the estimation?

- The key is that the reduce form has zero effect.

- Whatever other channels (of the same instrument) will be captured in the reduce form.

- So no effect in the reduce form for this instrument, means no effect for any treatment/channel this instrument is instrumenting.

- Additionally, the an OVB analysis of the OLS estimates shows us that most (all?) potentially omitted variables produce $OVB > 0$ (practice question for the exam!).

RDD: Final Considerations 1/2

- Visual inspection of RDD estimates are important but remember to keep an eye on the range of the y-axis

- Notice here that we cannot interpret the result of regression as a matched group, because we do not have individuals in the same cell (say age 20) with both treatment and control. The validity of RDD depends on our willingness to extrapolate across the running variable, at least around a narrow neighborhood around the cut-off.

- This extrapolation limits the policy questions that can be answered with RDD evidence. RDD can answer questions about changes in the margin (from 21 to 22 or 19) but not complete rearrangements of a policy (prohibiting or eliminating restrictions completely).

RDD: Final Considerations 2/2

- There is one important assumption for RDD that MM does not discuss, and it is pretty important (but I will not test you on it): RDD works as long as the threshold cannot be manipulated. This means that individuals cannot place themselves on either side of the threshold at will. This probably can be connected to the exclusion restriction, but requires a deeper dive into Fuzzy RDD. For those interested in more RDD I suggest following up this class from Andrew Heiss.

Acknowledgments

- *MM*