REGRESSION DISCONTINUITY DESIGN

2023/11/30

What is R.D.?

- A way of estimating treatment effects in a nonexperimental setting where treatment is determined by whether an observed “assignment” variable exceeds a known cutoff point.

- “assignment” variable is also referred to as the “forcing” variable or the “running” variable

The main idea

We’re interested in the jumps/changes that make very similar people get very dissimilar results.

Example: impact of scholarship on future academic outcomes

- If all students above a given grade—for example 80%—are given the scholarship
- A student scoring 79% is likely to be very similar to a student scoring 81%—given the pre-defined threshold of 80%, however, one student will receive the scholarship while the other will not.
- Comparing the outcome of the awardee (treatment group) to the counterfactual outcome of the non-recipient (control group) will hence deliver the local treatment effect.

Sharp and Fuzzy Designs

- Sharp RD: When all units in the study comply with the treatment condition they have been assigned

  Treatment is a deterministic function of the running variable.

- Fuzzy RD: exploits discontinuities in the probability of treatment conditional on a covariate X (the discontinuity is then used as an IV)

Sharp and Fuzzy Designs

Imbens and Lemieux (2008)

G.W. Imbens, T. Lemieux / Journal of Econometrics [redacted] ([redacted]) [redacted]–[redacted]

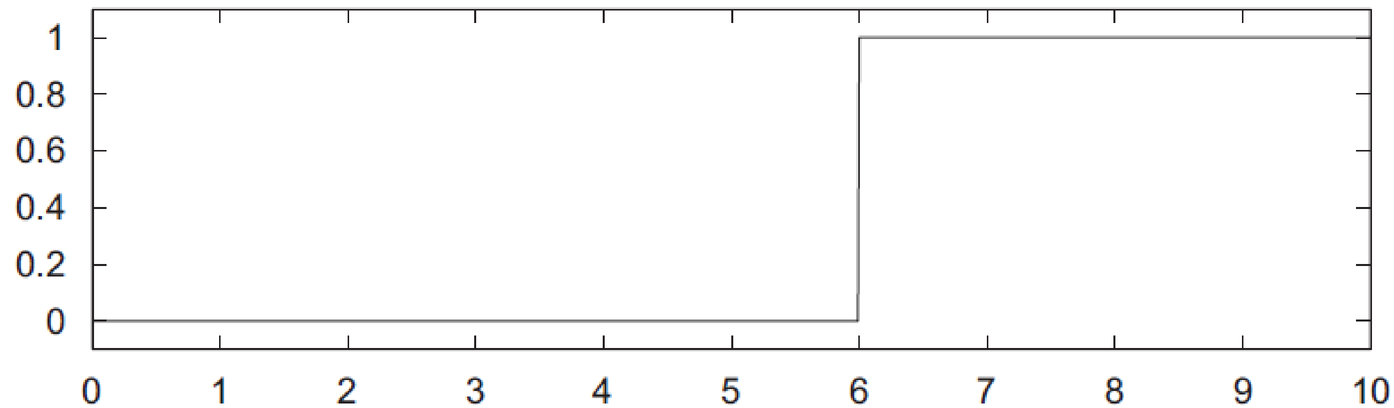


Fig. 1. Assignment probabilities (SRD).

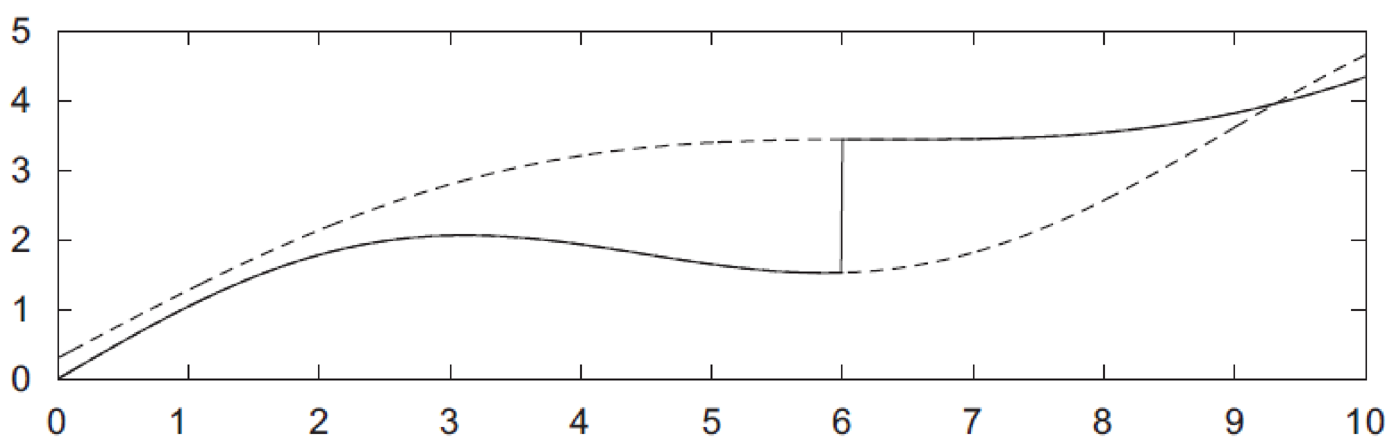
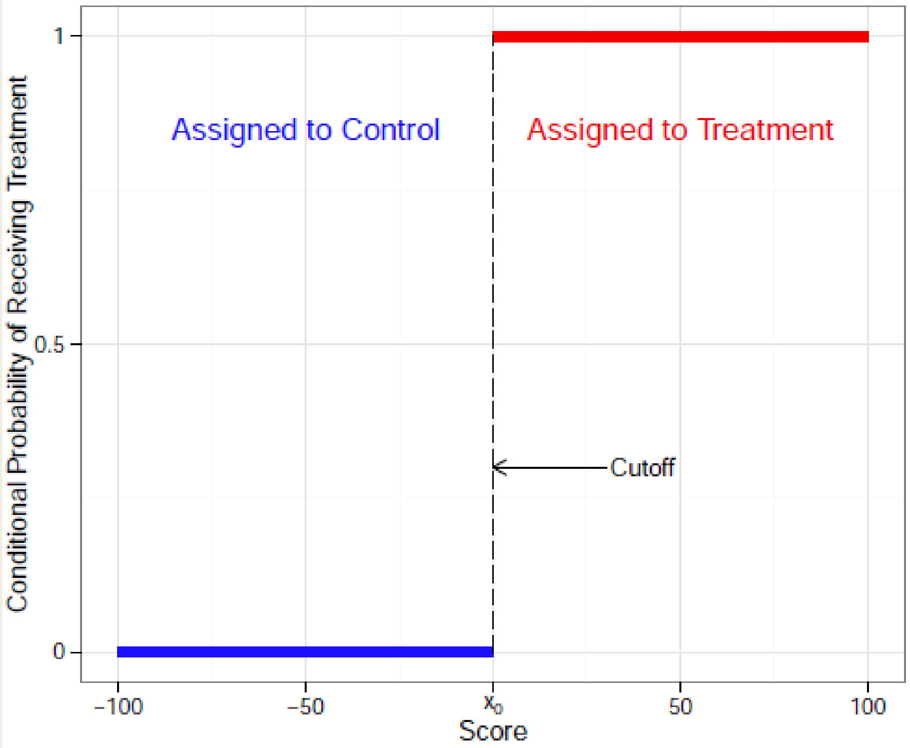
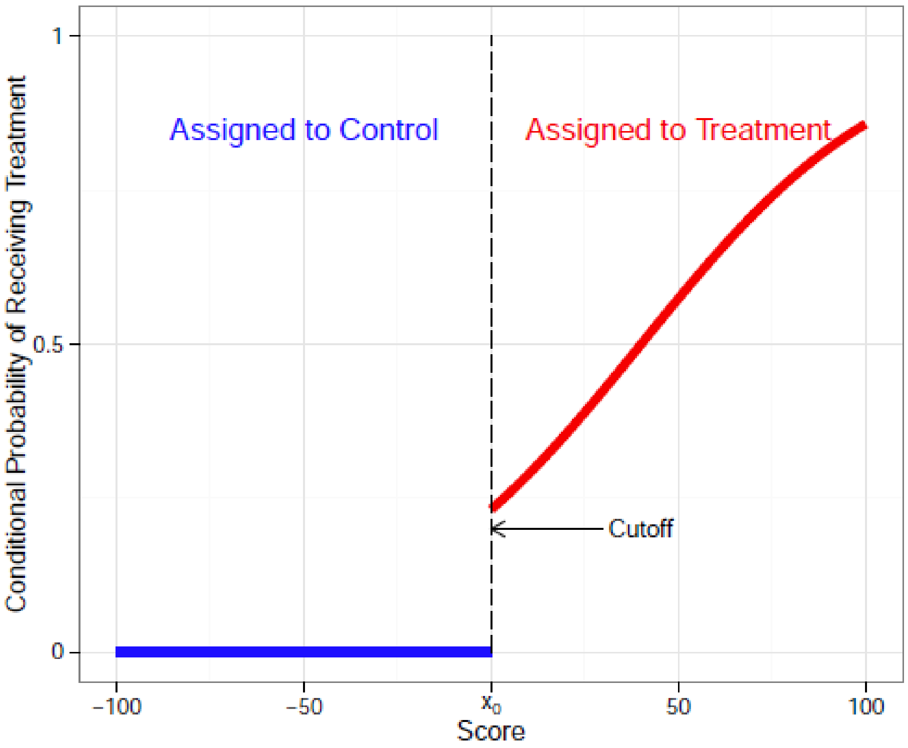


Figure 1: Conditional Probability of Receiving Treatment in Sharp vs. Fuzzy RD Designs



(a) Sharp RD



(b) Fuzzy RD (One-Sided)

Sharp RD

- In Sharp RD designs you exploit that treatment status is a deterministic and discontinuous function of a running variable xi .

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

- xo is a known threshold or cutoff
- Once we know xi we know Di .
- Functional form specification is important

Sharp RD – Simple Linear RD Setup

- **There are two commonly used RD designs. The first one is Local Linear Regression (LLR):**

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

- → $$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i,$$

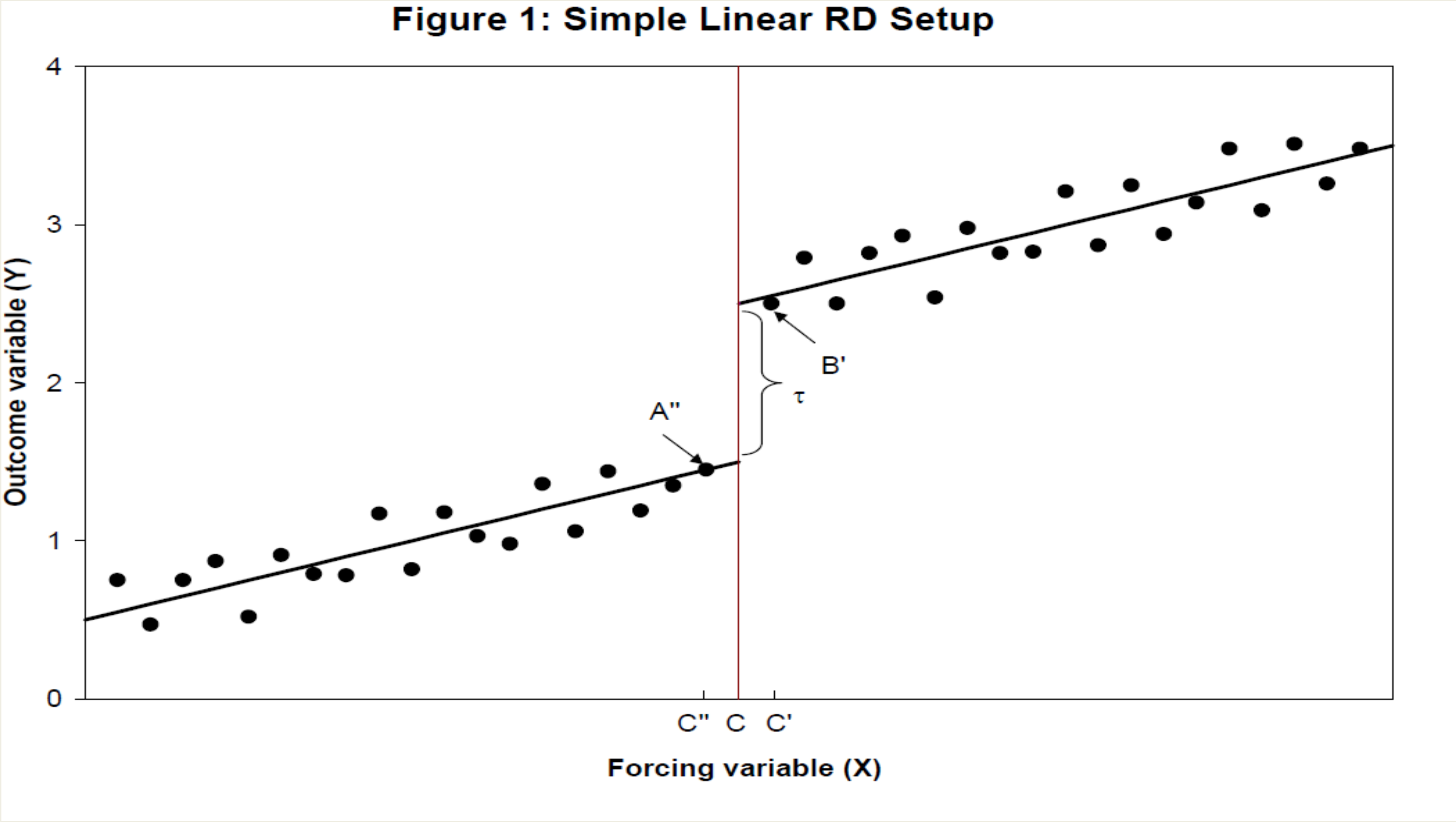
- The estimation is carried out using observations with a distance of *h* from the cutoff point.

Sharp RD

Key identifying assumption:

- $E[Y0i \mid Xi]$ and $E[Y1i \mid Xi]$ are continuous in Xi at $X0$.
- **This means that all other unobserved determinants of Y are continuously related to the running variable X.**
- This allows us to use average outcomes of units just below the cutoff as a valid counterfactual for units right above the cutoff.
- **Hard to test → make more graphs on other observables**

A. Simple Linear RD Setup



- Forcing Variable :
- Point of Discontinuity
- Outcome Variable

Sharp RD – general format

- A more general format:

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

- The function, *f(xi),* is in general flexible in form. But it usually take the form of higher polynomial orders of the forcing variable, and interaction terms with D up to a certain degree.

$$E[Y_{0i}|X_i] = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \ldots + \beta_{0p}\tilde{X}_i^p$$
$$E[Y_{1i}|X_i] = \alpha + \rho + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + \ldots + \beta_{1p}\tilde{X}_i^p$$

$$\tilde{X}_i = X_i - X_0$$

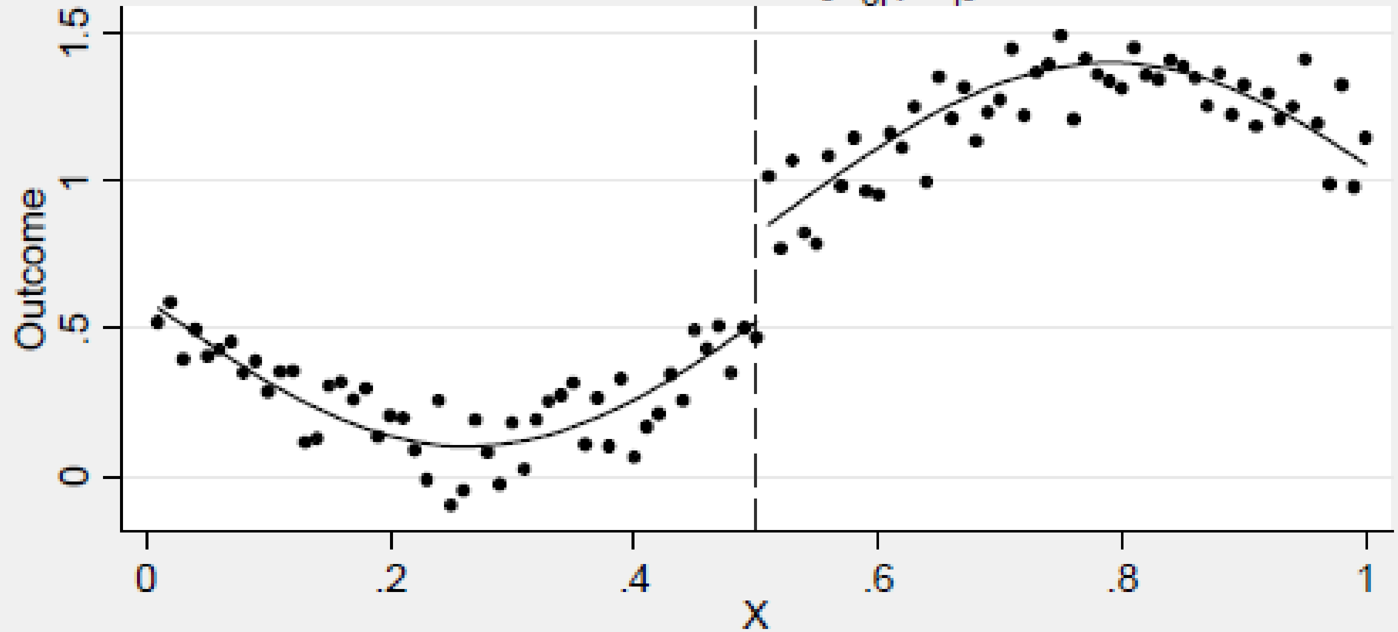
Sharp RD – general format

- The function, *f(xi),* is in general flexible in form. But it usually take the form of higher polynomial orders of the forcing variable, and interaction terms with D up to a certain degree.

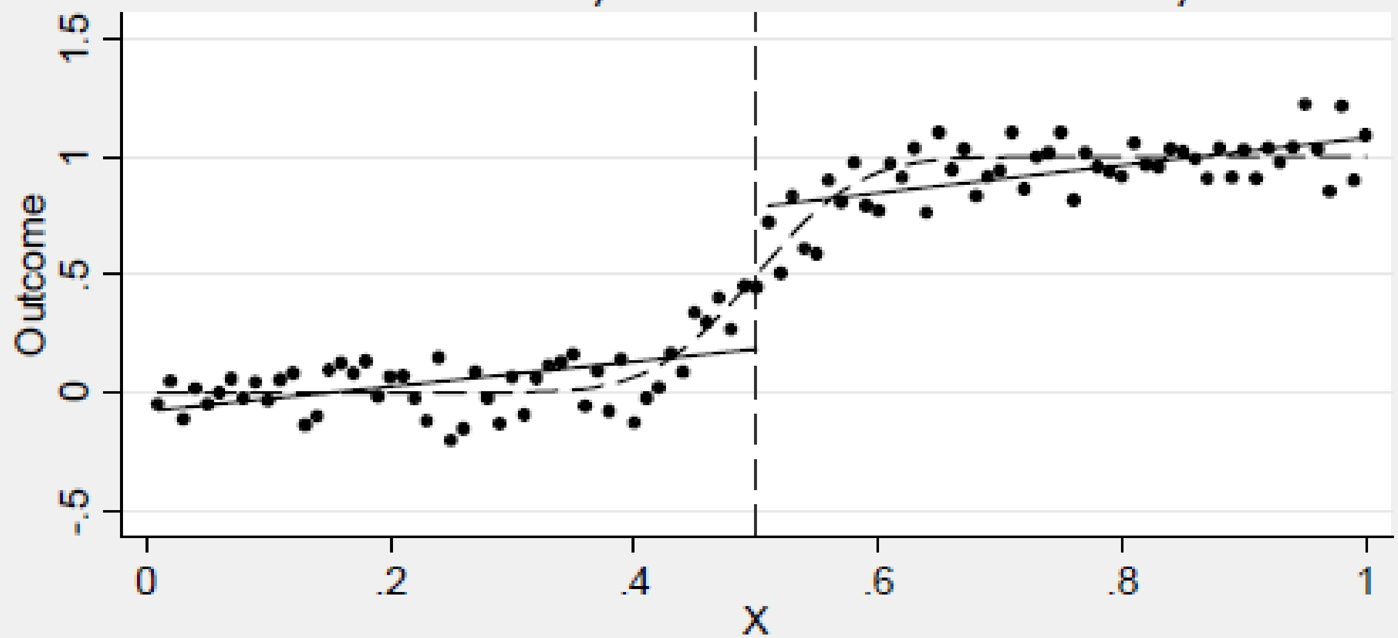
$$\begin{aligned} \mathrm{Y}_i \quad = \quad & \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p \\ & + \rho \mathrm{D}_i + \beta_1^* \mathrm{D}_i \tilde{x}_i + \beta_2^* \mathrm{D}_i \tilde{x}_i{}^2 + \ldots + \beta_p^* \mathrm{D}_i \tilde{x}_i{}^p + \eta_i,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

$\beta_1^* = \beta_{11} - \beta_{01}, \beta_2^* = \beta_{21} - \beta_{21}$ and $\beta_p^* = \beta_{1p} - \beta_{0p}$

B. Nonlinear $E[Y_{0i} | X_i]$

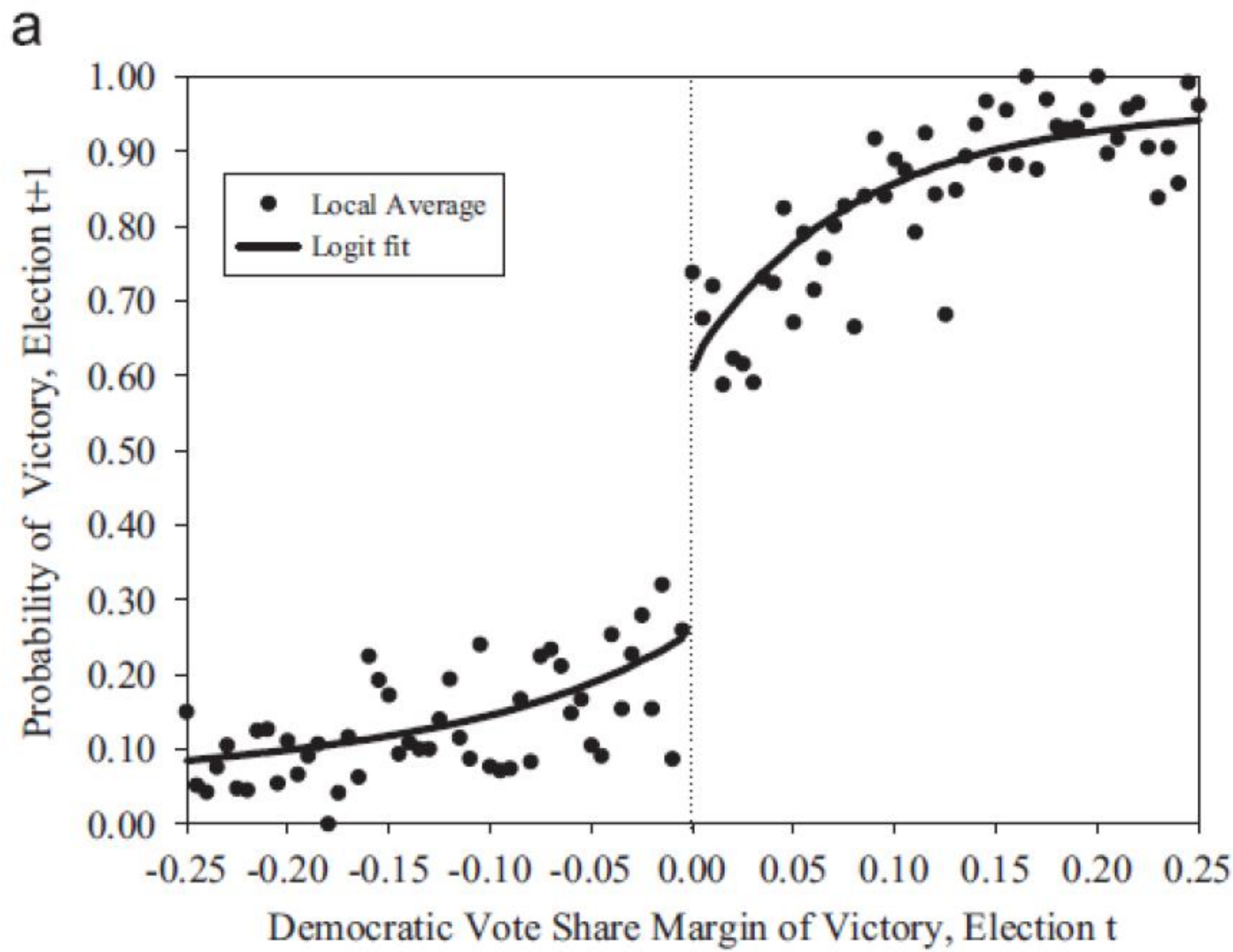


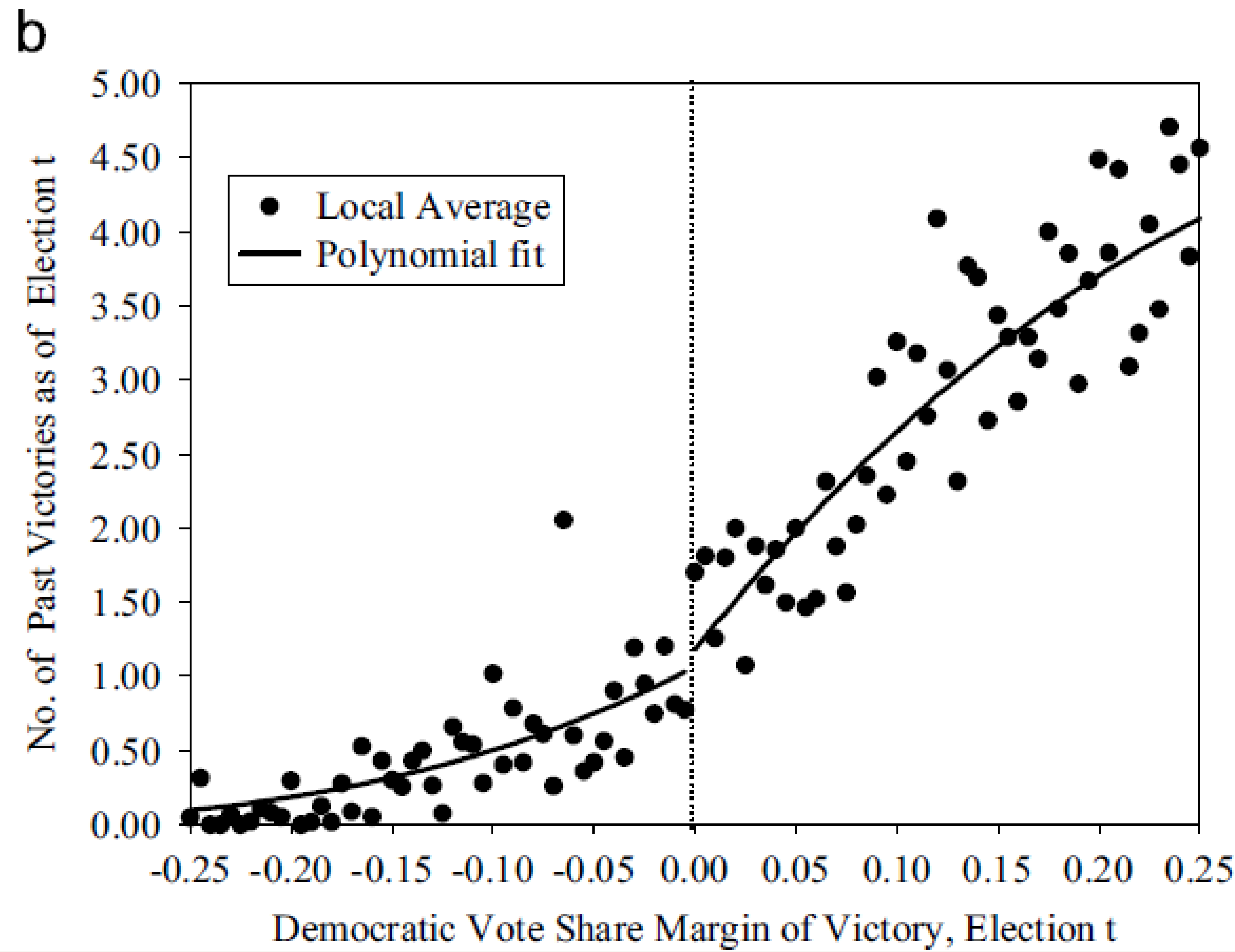
C. Nonlinearity mistaken for discontinuity



Example Sharp RD: Lee (2008) Incumbency Effects

- Lee (2008) uses a sharp RD design to estimate the probability that the incumbent wins an election.
- A large political science literature suggests that incumbents may use privileges and resources of office to gain an advantage over potential challengers.
- An OLS regression of incumbency status on election success is likely to be biased because ?
- Lee analyzes the incumbency effect using Democratic incumbents for US congressional elections.
- He analyzes the probability of winning the election in year t+1 by comparing candidates who just won compared to candidates who just lost the election in year t.


a
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Local Average
Logit fit
Probability of Victory, Election t+1
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Democratic Vote Share Margin of Victory, Election t


b
Local Average
Polynomial fit
No. of Past Victories as of Election t
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Democratic Vote Share Margin of Victory, Election t

Sharp discontinuity

- The discontinuity precisely determines treatment
- Equivalent to random assignment in a neighborhood
- E.g. Social security payment depend directly and immediately on a person's age

Fuzzy discontinuity

- exploits discontinuities in the probability of treatment conditional on a covariate X
- The discontinuity is then used as an IV. (Discontinuity is highly correlated with treatment .
  - E.g. Rules determine eligibility but there is a margin of administrative error.
  - Use the assignment as an IV for program participation.

Fuzzy RD

- Due to the existence of non-compliance, we have a fuzzy RD design, as opposed to a sharp design where passing the threshold point completely predicts the treatment. In the fuzzy case:

$$P[D_i = 1|X_i] = \{\begin{matrix} g_1(X_i) \text{ if } x_i \geq x_o \\ g_0(X_i) \text{ if } x_i < x_o \end{matrix}, \text{ where } g_1(X_i) \neq g_0(X_i)$$

- In this case we can use eligibility as the IV

  Let Ti=1 if Xi>X0

  Ti=0 otherwise

- Ti can be our instrument variable

First Stage & Reduced form

- First Stage

$$D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + \ \dots \ + \gamma_p X_i^p + \pi T_i + \xi_{1i}$$

- Reduced from

$$Y_i = \mu + \kappa_1 X_i + \kappa_2 X_i^2 + \ \dots \ + \kappa_p X_i^p + \rho \pi T_i + \xi_{2i}$$

With interaction terms

- Again, we can also allow interaction terms of T and Xi

  In this case, instrumental variables include:

  $$\left\{\mathrm{T}_i,\ \tilde{x}_i\mathrm{T}_i,\ \tilde{x}_i^2\mathrm{T}_i,\ .\ .\ .\ \tilde{x}_i^p\mathrm{T}_i\right\}$$

- Endogenous variables include

  $$\left\{\mathrm{D}_i,\ \tilde{x}_i\mathrm{D}_i,\ \mathrm{D}_i\tilde{x}_i{}^2,\ .\ .\ .\ \mathrm{D}_i\tilde{x}_i{}^p\right\}$$

Example: Effect of cash transfer on consumption

Target transfer to poorest households

Method

- Construct poverty index from 1 to 100 with pre-intervention characteristics
- Households with a score ≤ 50 are poor
- Households with a score >50 are non-poor

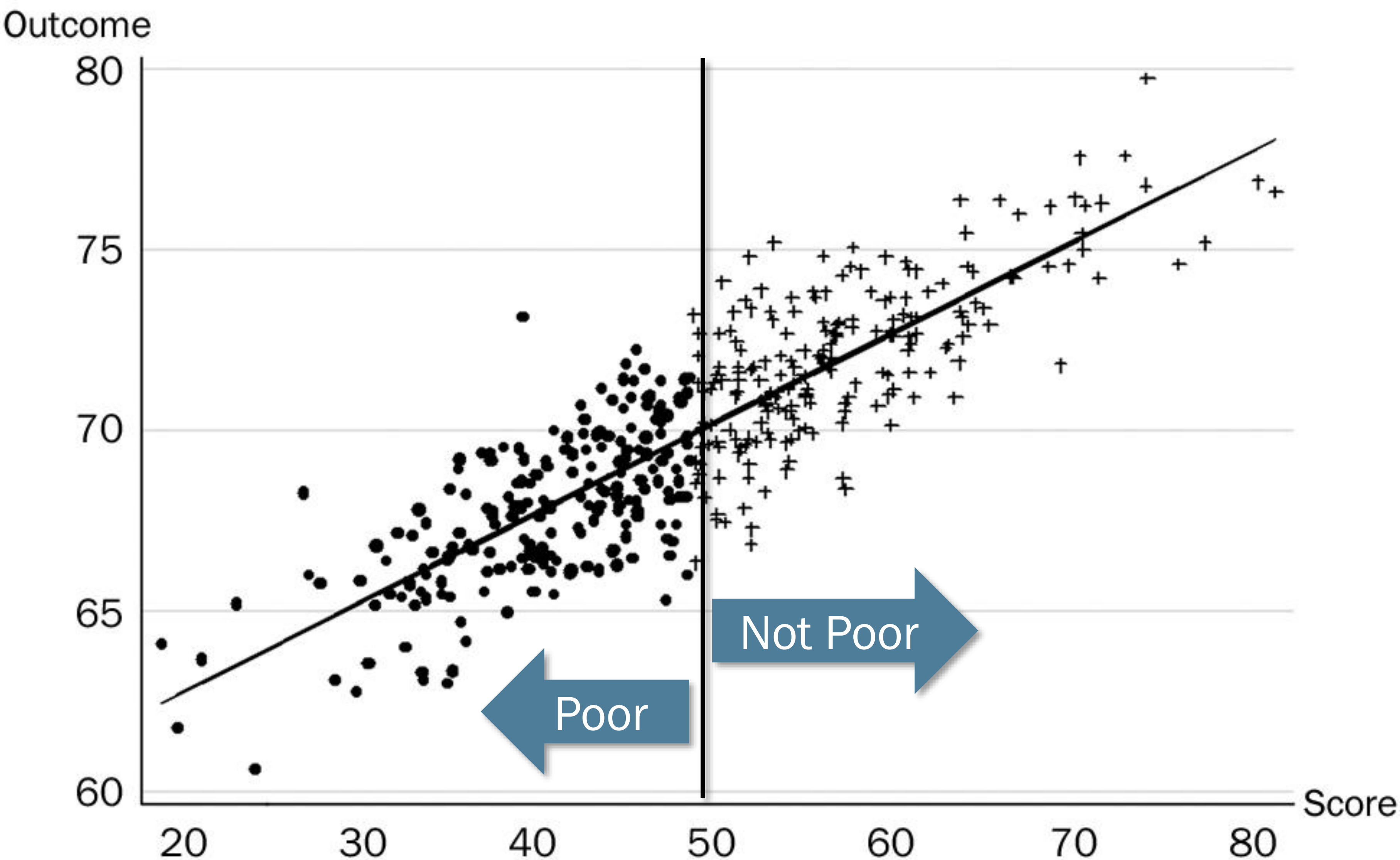
Implementation

Cash transfer to poor households

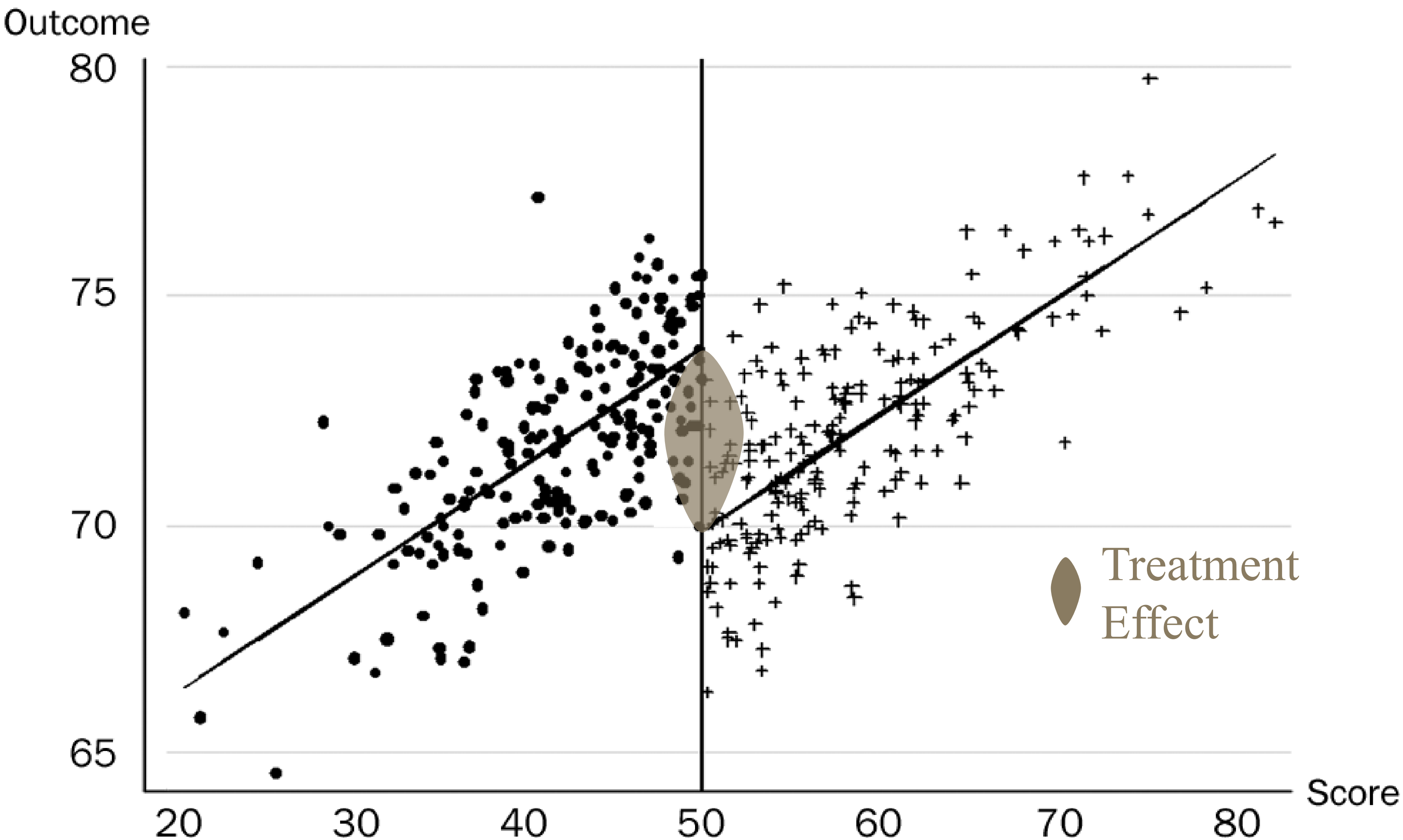
Evaluation

Measure outcomes (i.e. consumption, school attendance rates) before and after transfer, comparing households just above and below the cut-off point.

Regression Discontinuity Design-Baseline



Regression Discontinuity Design-Post Intervention



Identification for sharp discontinuity

$$y_i = \beta_0 + \beta_1 D_i + \delta(score_i) + \varepsilon_i$$

$$D_i = \begin{cases} 1 & \text{If household } i \text{ receives transfer} \\ 0 & \text{If household } i \textit{ does not} \text{ receive transfer} \end{cases}$$

$\delta(score_i)$ = Function that is continuous around the cut-off point

- Assignment rule under sharp discontinuity:

$D_i = 1 \iff score_i \leq 50$

$D_i = 0 \iff score_i > 50$

Identification for fuzzy discontinuity

$$y_i = \beta_0 + \beta_1 D_i + \delta(score_i) + \varepsilon_i$$

$$D_i = \begin{cases} 1 & \text{If household receives transfer} \\ 0 & \text{If household } \textit{does not} \text{ receive transfer} \end{cases}$$

- But
  Treatment depends on whether $score_i$ > or < 50

  &

  other endogenous factors

Identification for fuzzy discontinuity

$$y_i = \beta_0 + \beta_1 D_i + \delta(score_i) + \varepsilon_i$$

IV estimation

- First stage:

$$D_i = \gamma_0 + \gamma_1 \underbrace{I(score_i > 50)}_{\text{Dummy variable}} + \delta(score_i) + \eta_i$$

- Second stage:

$$y_i = \beta_0 + \beta_1 D_i + \underbrace{\delta(score_i)}_{\text{Continuous function}} + \varepsilon_i$$

bandwidth

Two things aren't universally clear:

1. How wide a bandwidth around the cutoff are we looking at?

Estimation and Interpretation:

We're really only confident in our estimate for people that are close to the cutoff. This is a LOCAL AVERAGE TREATMENT EFFECT.

We can confidently say that a school right around the cutoff would improve average test scores by X if they received the treatment, but we're not so confident that already awesome schools would get the same benefit.

Functions

2. Without the program, what shaped function would there be naturally?

What sort of function do we throw in to control for the fact that even if there was no scholarship, smarter kids are likely to earn more later in life?

The solution: SHOW YOUR WORK

In addition to showing your work, another good robustness check is to test for the effects of non-existent programs.

Tips

- It is probably advisable to report results for different specifications (linear v.s. non-linear)

- In robustness checks you also want to show that including higher order polynomials does not substantially affect your findings.

- You also want to show that your results are not affected if you vary the window around the cutoff (standard errors may go up but hopefully the point estimate does not change)

Graphical Analysis

- Outcome Variable
- Treatment Variable
- Other control variables
- Density function

- Also, we can also repeat RD analysis on control variables

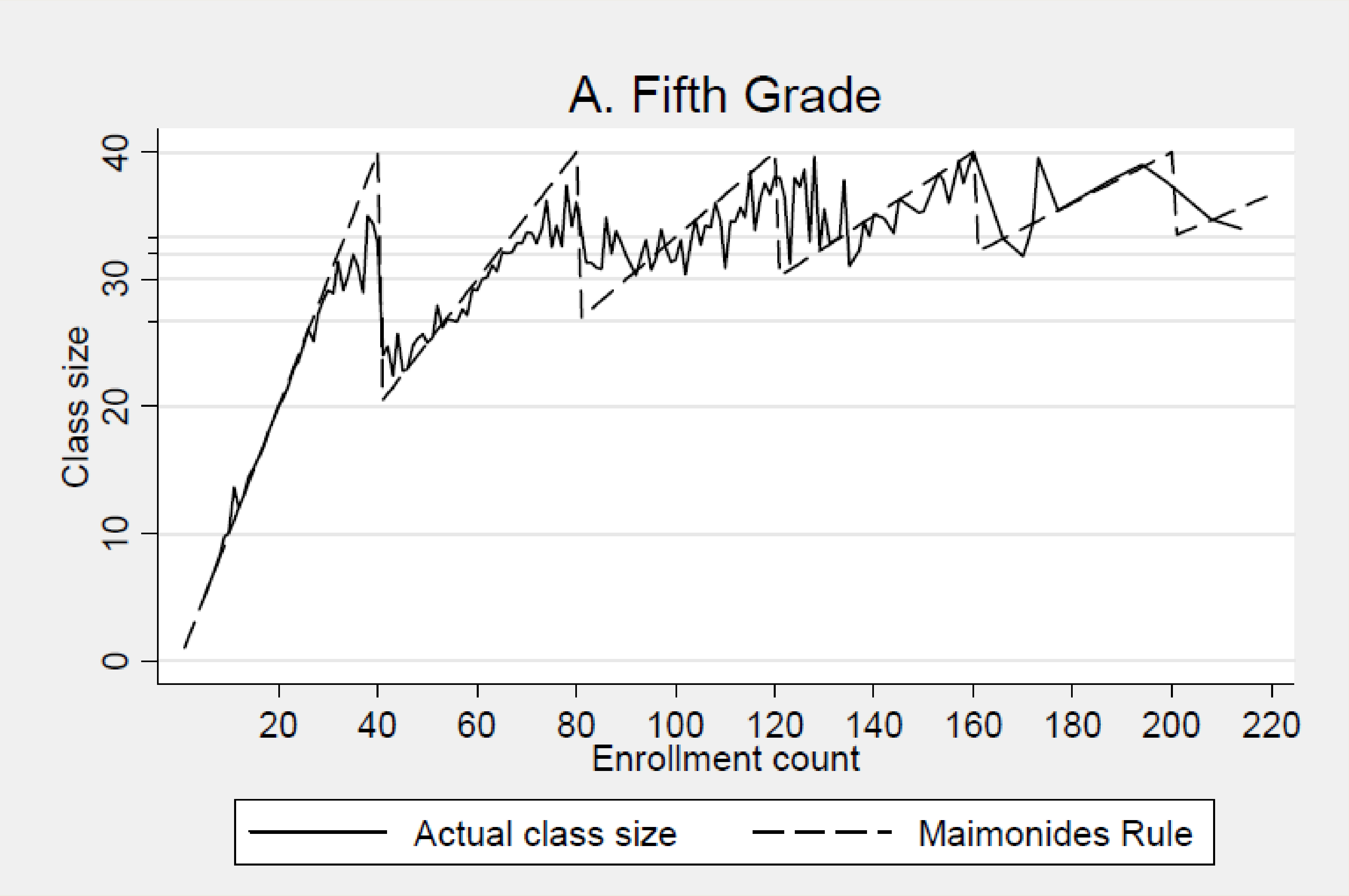
Discontinuity Examples

- School Class Size
  - *Maimonides' Rule--No more than 40 kids in a class in Israel.*
  - *40 kids in school means 40 kids per class. 41 kids means two classes with 20 and 21.*

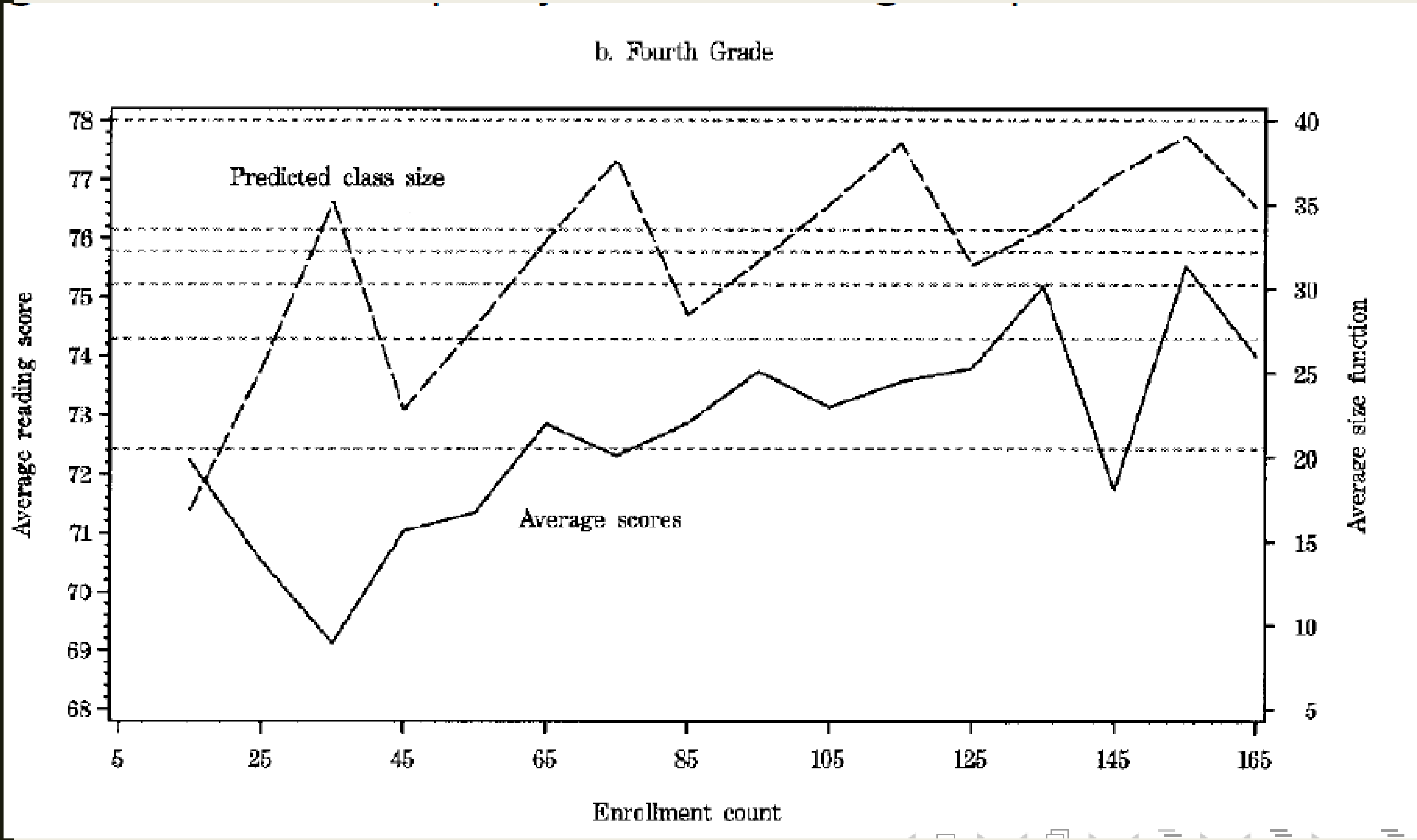
$$m_{sc} = \frac{e_s}{int[\frac{(e_s - 1)}{40}] + 1}$$

(Angrist & Lavy, QJE 1999)

Predicted class size and actual class size: Sharp or Fuzzy?



Enrollment count and average scores



Discontinuity Examples

- Air Pollution and Home Values
  - *The Clean Air Act's National Ambient Air Quality Standards say if the geometric mean concentration of 5 pollutant particulates is 75 micrograms per cubic meter or greater, county is classified as "non-attainment" and are subject to much more stringent regulation.*
  - *The "non-attainment" is associated with large reduction in pollution.*

(Ken Chay, Michael Greenstone, JPE 2005)

A Bandwidth of Randomness

Test scores aren't random, and neither is class size, nor air pollution.

But is a kid in the 94.9th percentile really that different from the 95th percentile kid?

Is a school with 40 kids that different from a school with 41?

Right around the cutoff, there's a good chance things are random.

No Sorting - Observables

But don’t take my word for it. Look at the averages of the observables in your below cutoff group, and the averages of the observables in the above cutoff group. Are they the same? Hopefully, but maybe not.

Do people know about this cutoff? Are they doing some endogenous sorting? When deciding where to live, did good moms look for schools where their kids would be the 41st kid? Did certain types of polluters look for counties where they’d be below the cutoff?

These things can be checked to some degree--look at the average observables above and below the cutoff.

No Sorting - Clumping

In addition to checking the observables on either side of the cutoff, we should check the density of the distribution. Is it unusually low/high right around the cutoff?

If there's some abnormally large portion of people right around the cutoff, it's quite possible that you don't have random assignment.

Conclusion

- Find a threshold
- Look at people just above and just below
- Make sure there's no sorting
- It's only a local effect