Regression discontinuity

Outline

- Regression Discontinuity
  - Clean
  - Fuzzy
  - Jacobs/Lefgren example

Regression discontinuity

- Is a special case of an observational study design in which assignment to treatment is decided solely based on values of (typically) one measured variable, *X*
- In essence then *there is only one confounding covariate*
- The odd feature is that all observations with values of *X* on one side of a specified cutoff value, *c*, are given the treatment and all observations with values on the other side are denied the treatment
- The authors of one of your readings (Cook, Campbell and Shadish) purport that this study can only be implemented prospectively, but many would disagree

Regression discontinuity: more to think about

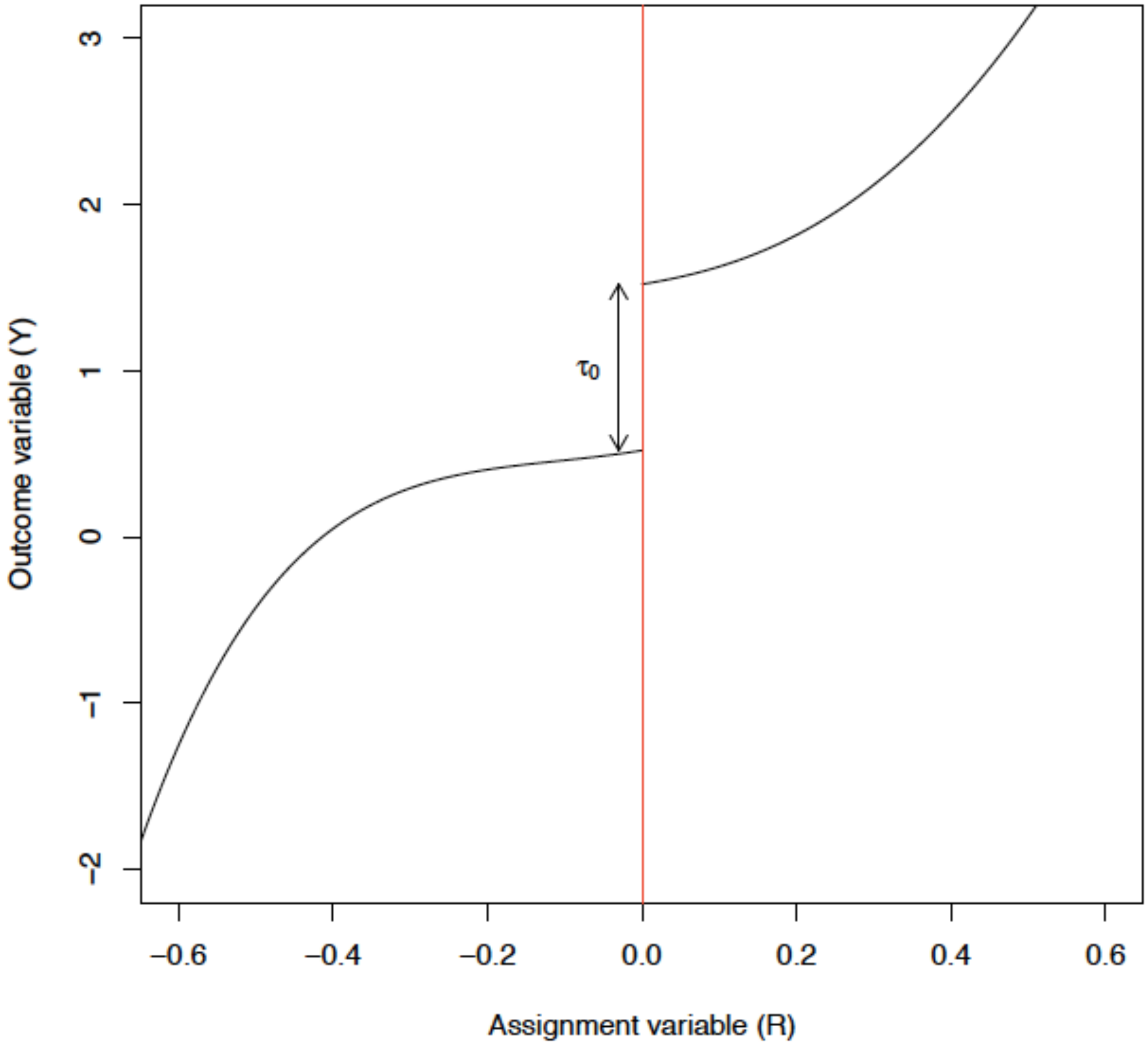
- At first glance it would appear that this is just the type of situation we talked about when first discussion observational studies… that is, what can we do if there is only *one* confounding covariate
- In that discussion we talked about stratifying or otherwise conditioning on this variable so that we compared treatment and controls who looked the same in terms of it
- In this situation however (in its purest form) <u>we have *no* observations with the same value of X that have different values of the treatment variable Z!</u>

Clean RDD

"Clean" regression discontinuity

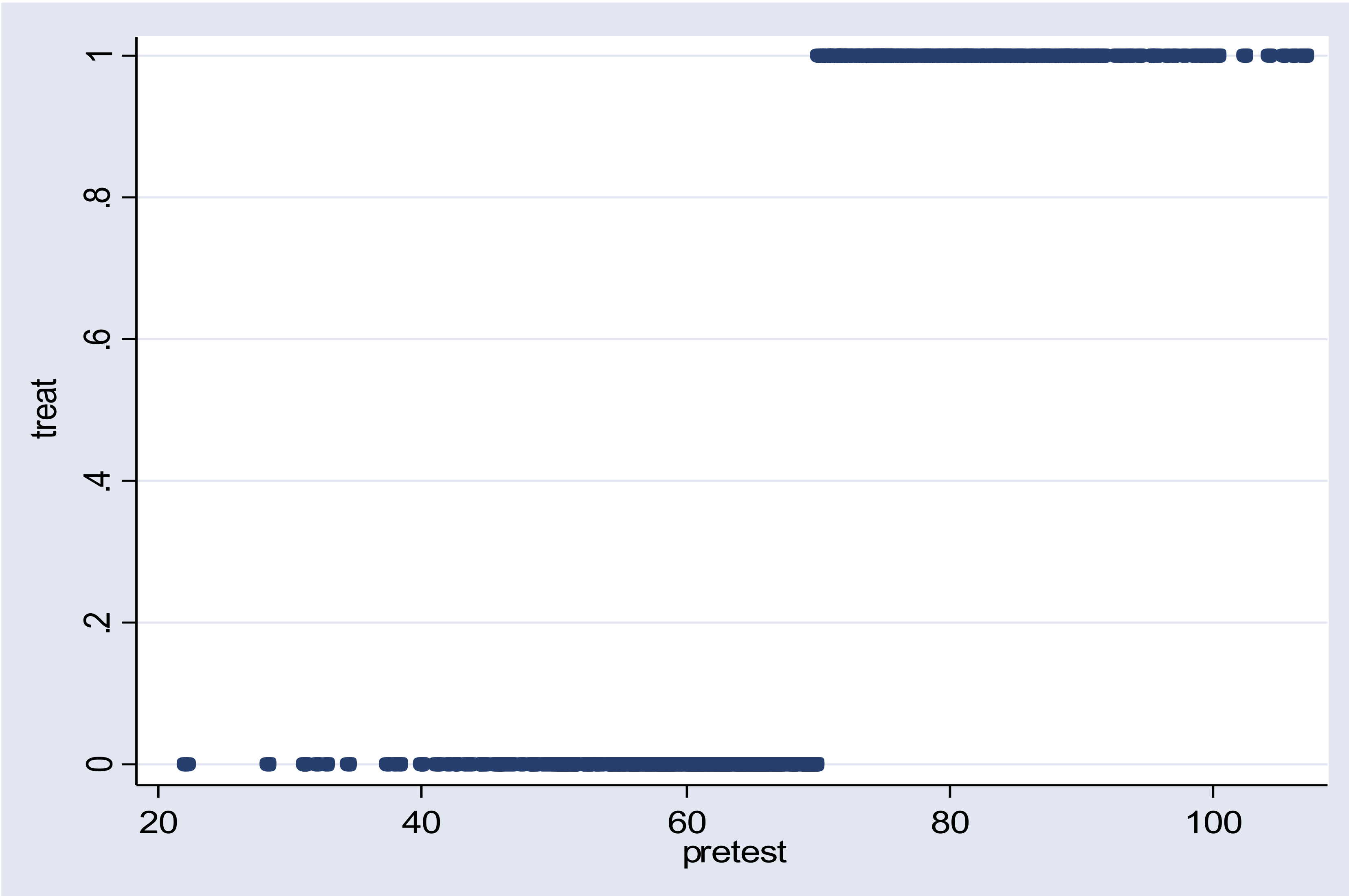
- In its purest (cleanest) form, regression discontinuity is implemented as a prospective quasi-experimental design
- For instance, suppose only students scoring above cutoff c on a qualifications test are allowed into a special math enrichment program
- Then *if the program were effective* we might see the following

Illustration of RD

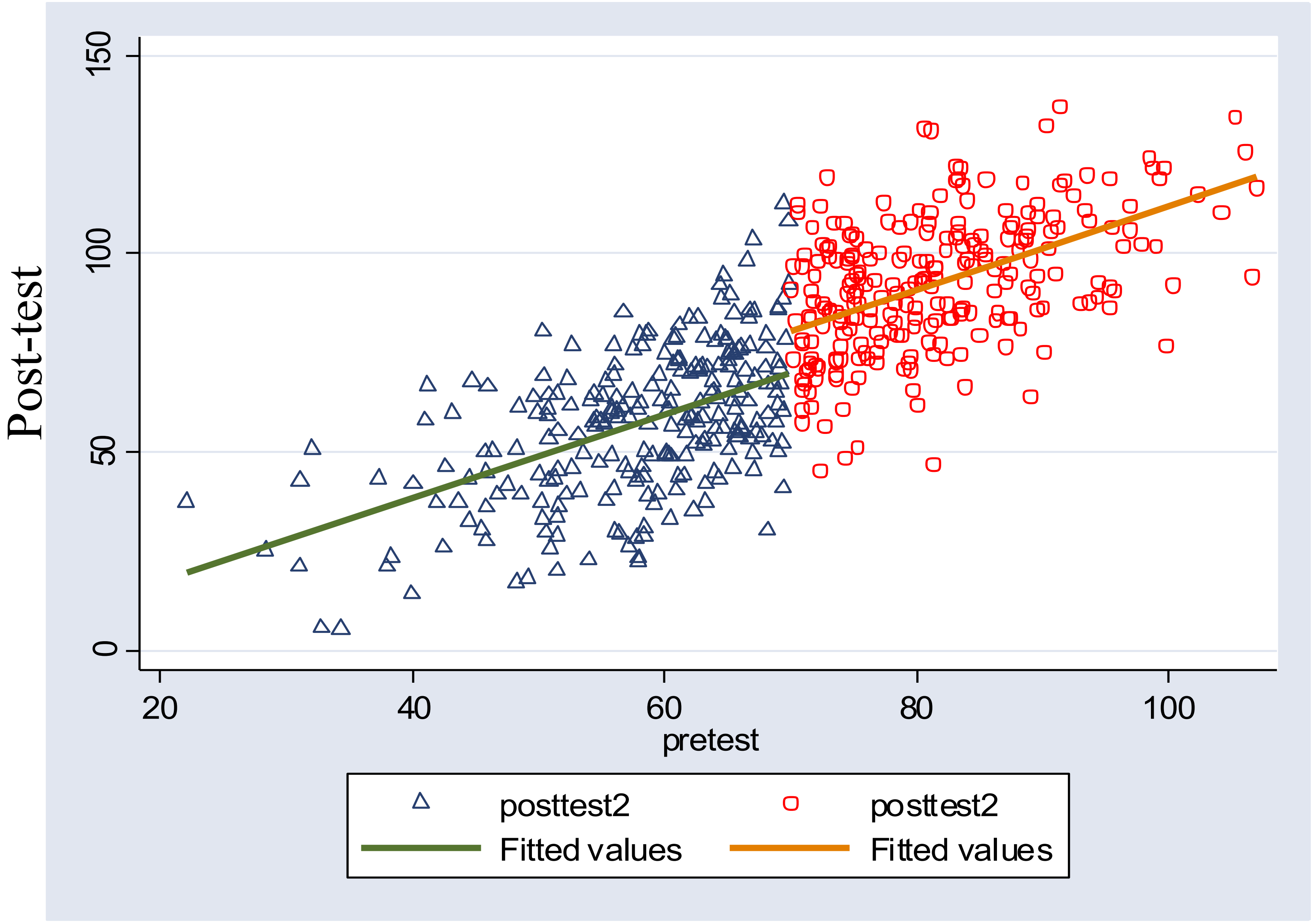


courtesy Cattaneo, 201

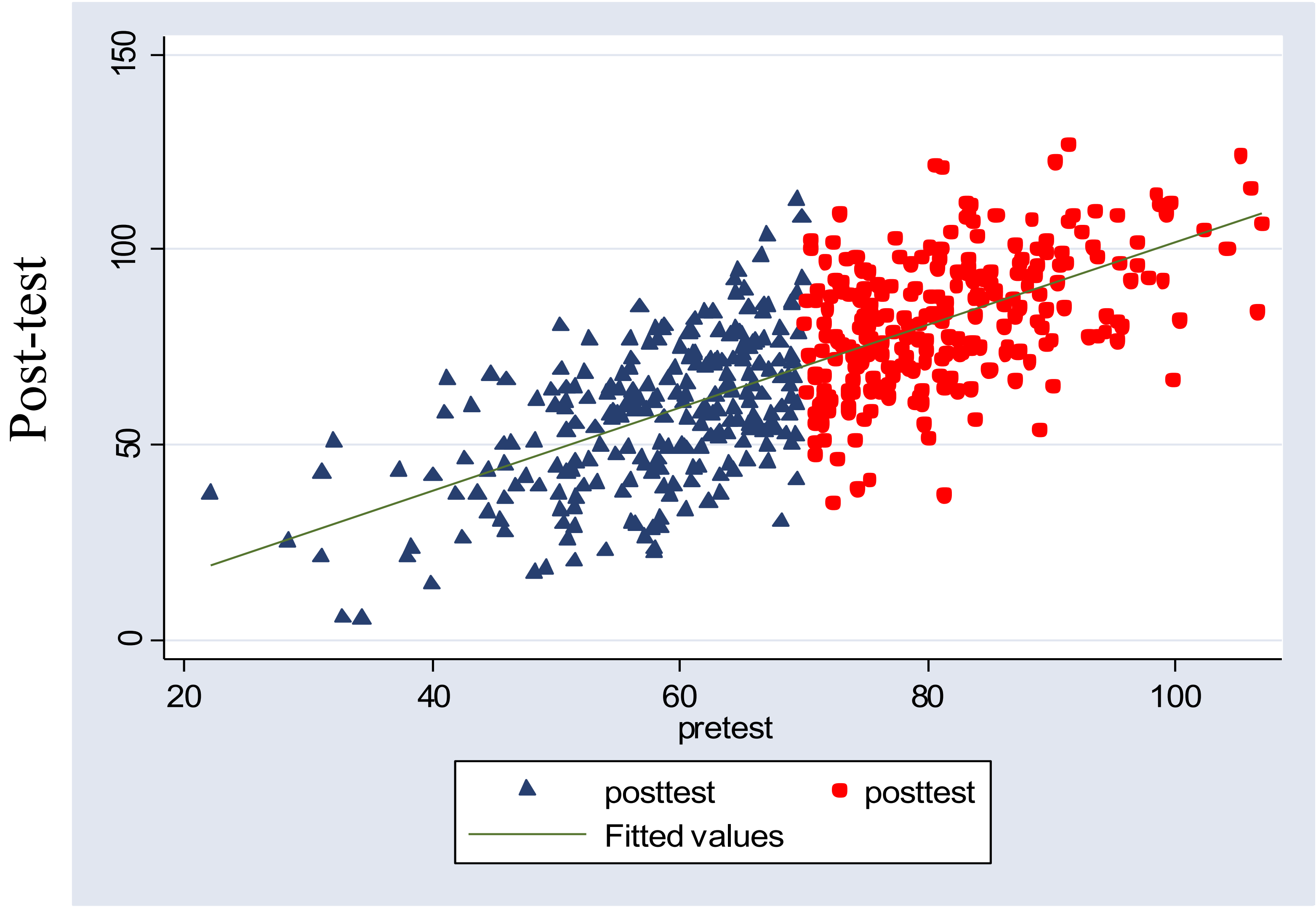
Relationship between pretest and Pr(treatment)



Regression discontinuity: effect



Regression discontinuity: no effect

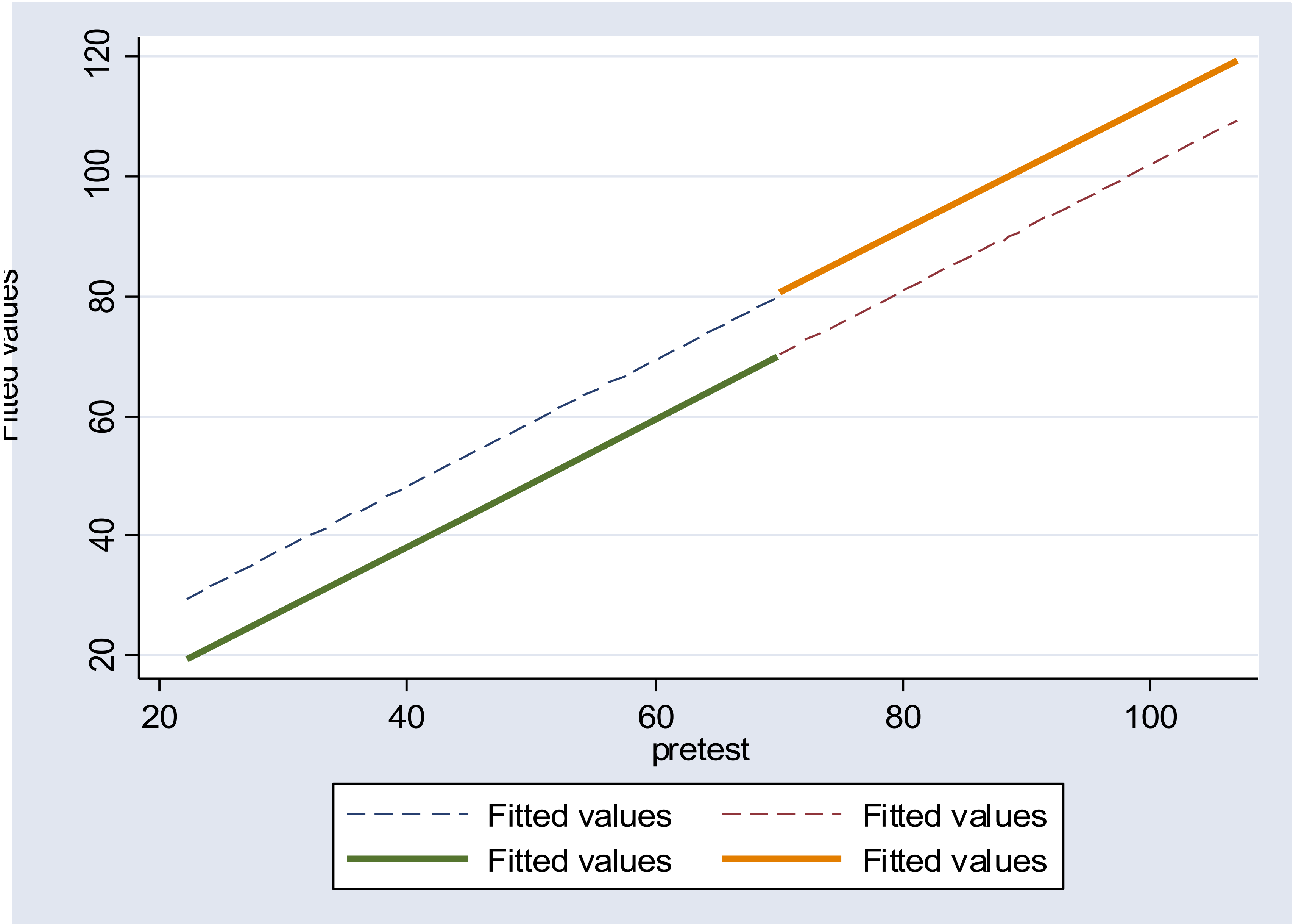


Regression discontinuity: analyses

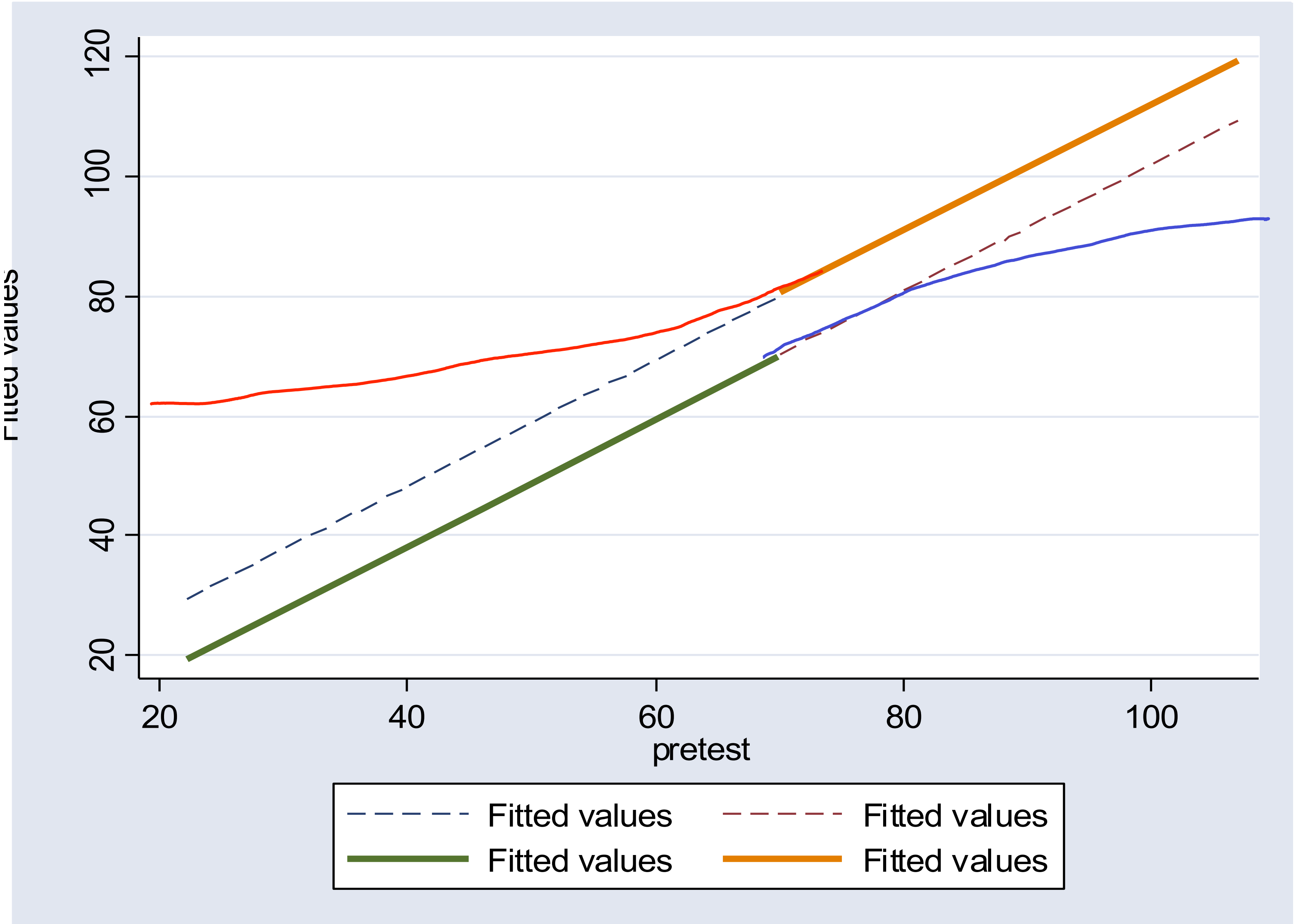
- Results are typically analyzed by regressing the observed outcome on the treatment indicator and (some functions of) the assignment mechanism
- As a simple example

  regress *outcome z x*
- It often makes sense to include an interaction between z and x these as well
- If the model is specified correctly this can yield unbiased estimates of the true treatment effect
- Note here that this type of model implicitly creates a counterfactual for each unit based *entirely* on model extrapolation

Regression discontinuity: analyses



Regression discontinuity: analyses



Regression discontinuity: assumptions

The basic assumptions are typically

- The treatment assignment is ignorable given X within a narrow interval of X around the threshold, c

$$Y(0), Y(1) \perp Z \mid X, (X > c - a \;\&\; X < c + a)$$

  (the length of the interval is 2a)
- We can model E[Y(1) | X, Z=0] and E[Y(0) |X, Z=1] in that interval
- The larger the interval the less plausible the assumptions

!!! Pop Quiz !!!

- How do our assumptions change if our estimand is the effect of the treatment on the treated:

$$E[Y(1) - Y(0) \mid Z=1]$$

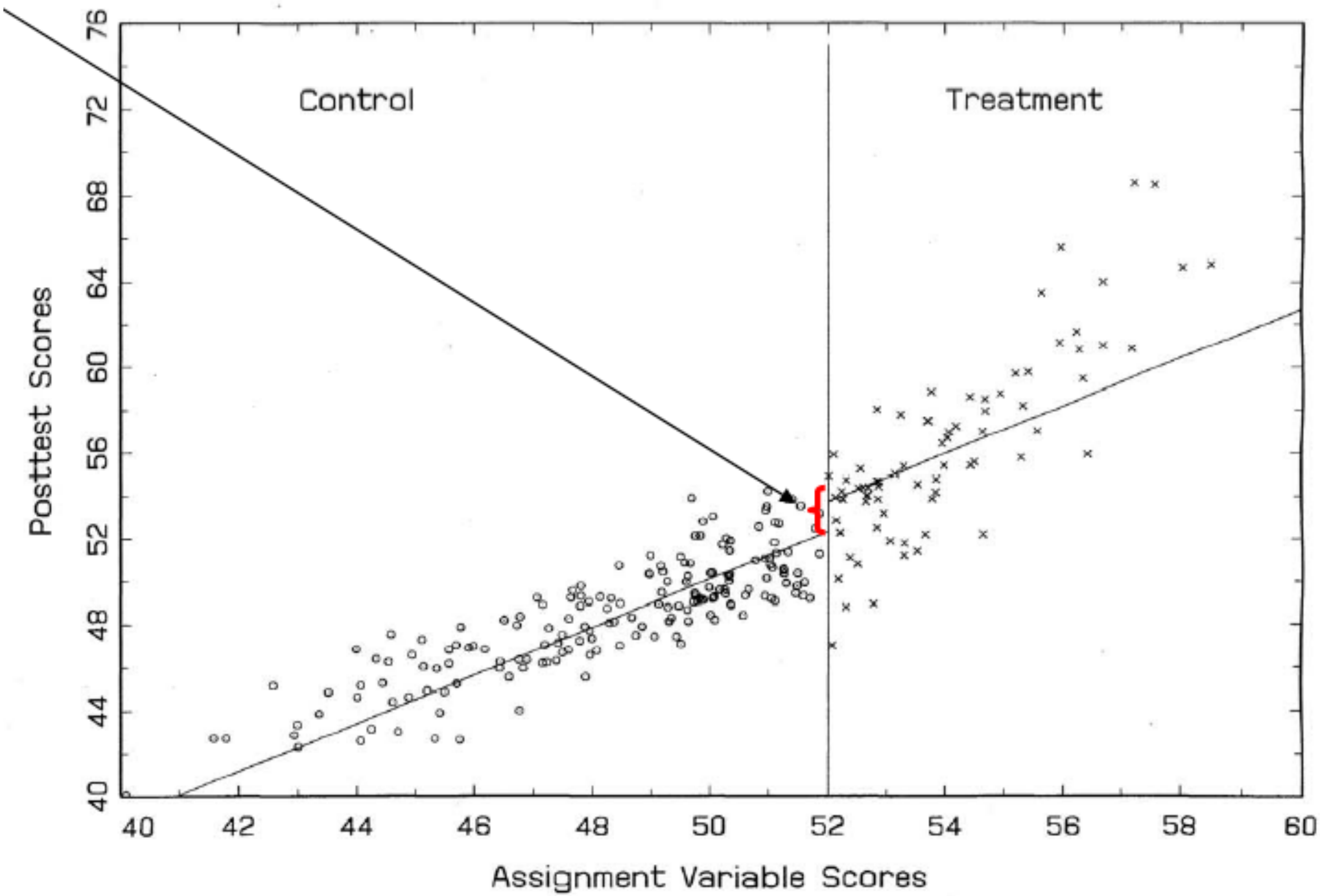
?

Regression discontinuity: current popular usage

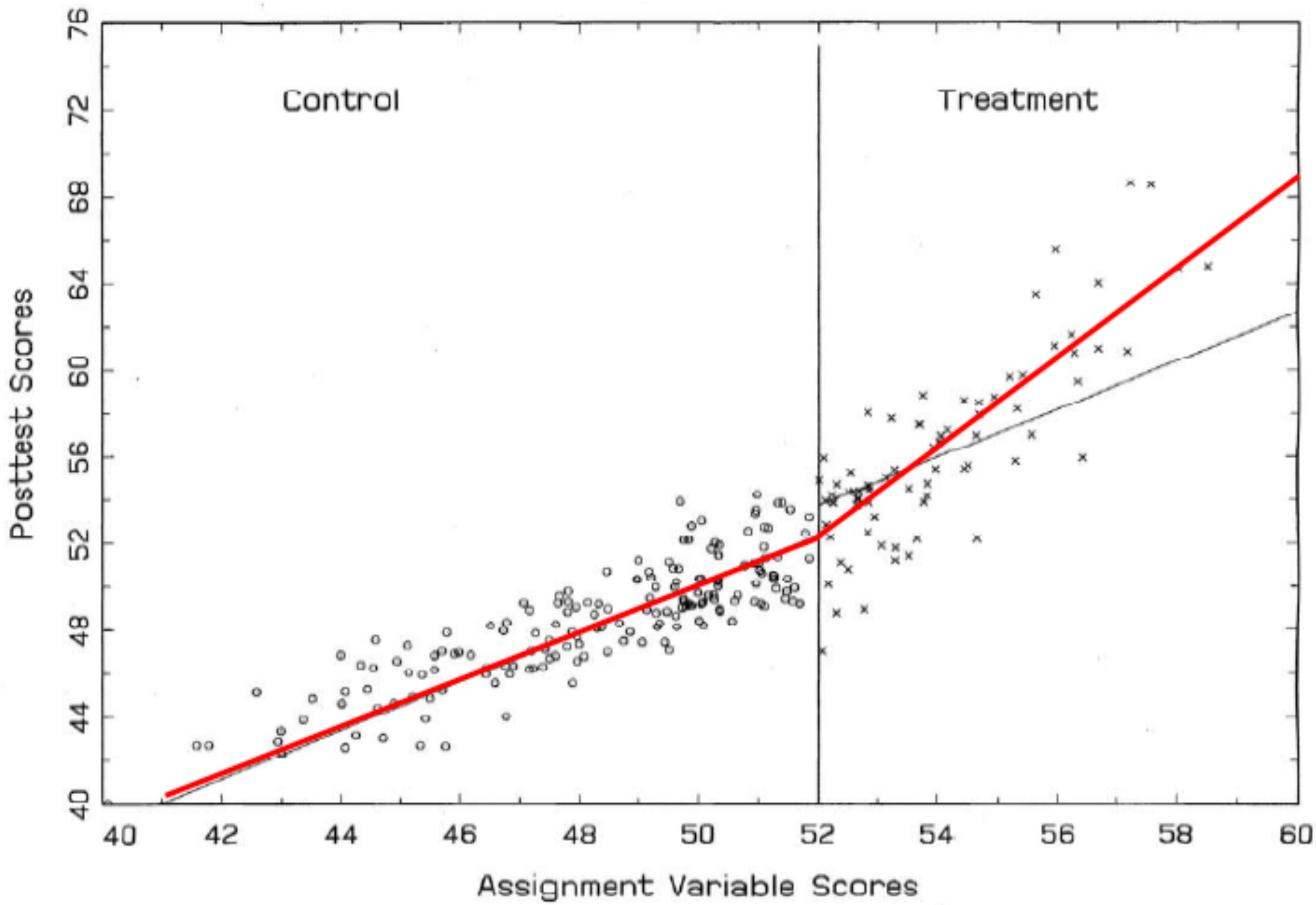
- Regression discontinuity is an increasingly popular favorite among labor economists and other social scientists as a *retrospectively* applied technique for observational studies
- They look for a situation in which a continuous variable was used to assign units to a policy or intervention based on a sharp cutoff
- For example EPA policies that sanction firms exceeding pollution thresholds, social service programs in which income determines eligibility, educational interventions where admission is based on test scores
- Generally then treatment and control units within some caliper width on either side of the cutoff are compared (sometimes using a model, sometimes not)
- This approach is less model dependent when it relies on the hope that the assignment variable is unrelated to the outcome so that assignment is somewhat random with respect to it

Importance of parametric form of the model

Anything wrong with the modeling here?

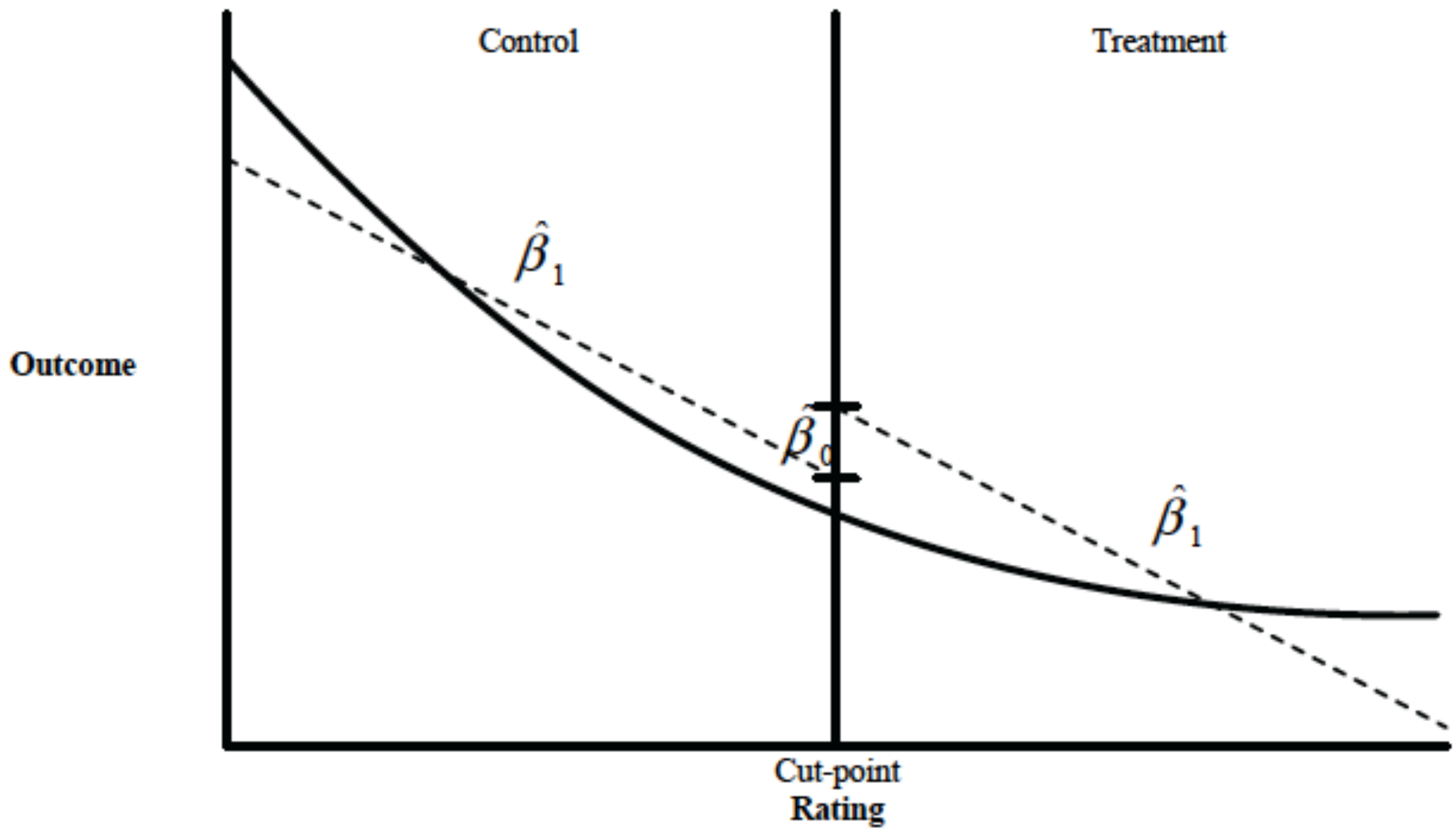


This (in red) better?
What's your interpretation of the effect?



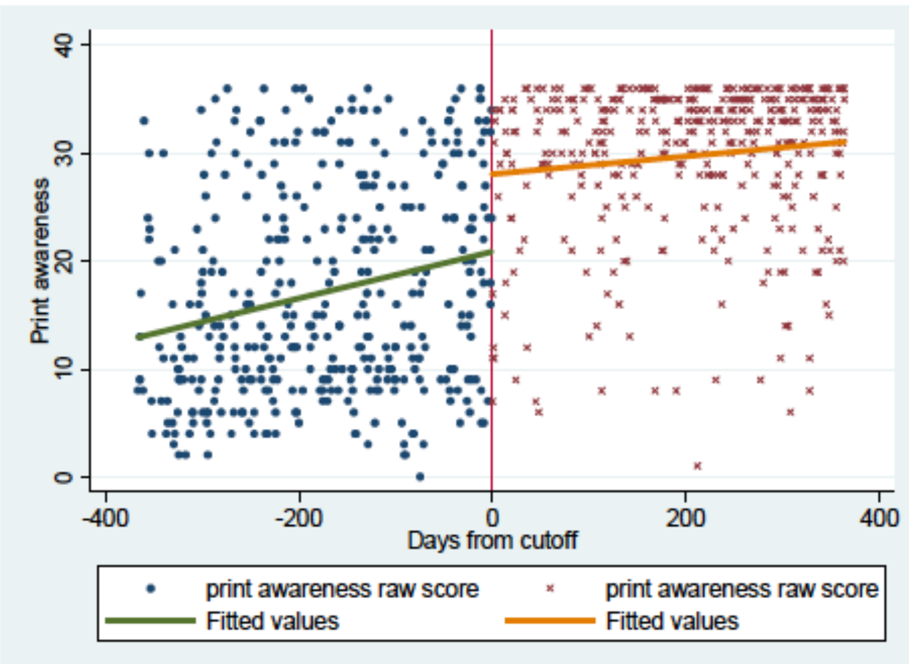
What about other types of non-linearities? Hypothetical

Regression Discontinuity Estimation with an Incorrect Functional Form

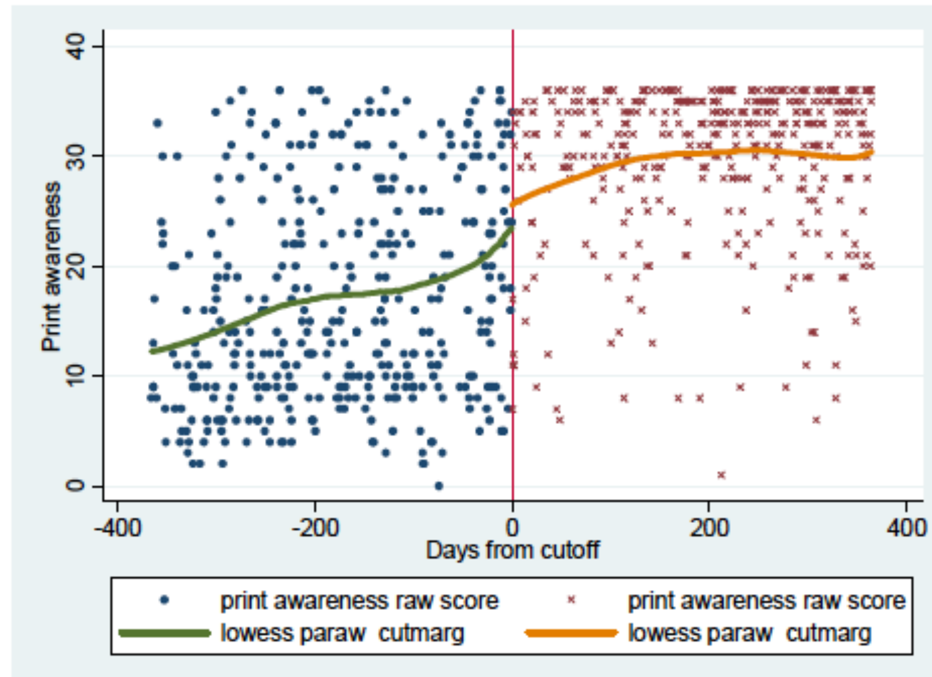


OTE: The solid curve denotes a true relationship that descends at a decreasing rate. The dashed lines represent a
nple linear regression fit to data generated by the curve

Non-linearities in a real example?



Linear plot



Lowess plot

Example: Effect of Oklahoma Early Child Care Program on Print Awareness

How to decide which model to use?

- PLOT YOUR DATA!
- Common advice is/was to try a bunch of polynomial models and do tests and/or print all results. Not a good idea!
  - gives undue weight to points far from the bandwidth
  - sensitivity to degree of polynomial
  - CIs can be too narrow with higher order polynomials, increasing the likelihood of false rejection

See Gelman and Imbens 2017 (or for a summary https://blogs.worldbank.org/impactevaluations/curves-all-wrong-places-gelman-and-imbens-why-not-use-higher-order-polynomials-rd)

- Better idea is to fit a *local* rather than a global model and to avoid higher order polynomials
- But then how to choose bandwidth?

Choosing bandwidth for local (linear or non-linear) regression

- Can choose in an ad-hoc way based on a combination of
  - plots
  - sample size considerations
  - choice of estimand
- and/or can present estimates for different bandwidths
- Can choose using cross-validation

Leave one out cross-validation to choose bandwidth

1) Select bandwidth to try, h_1
2) Select observation A closest to (but lower than) the threshold
3) Fit a regression of Y_A on X_A on all observations with X values in (X_A-h_1,X_A) (not including observation A) and use to predict Y_A
4) Repeat for observation B (with closest smaller X value) and for all other observations to the left of threshold.
5) Repeat for all observations to the right of threshold
6) Use predictions to calculate the cross-validation criterion

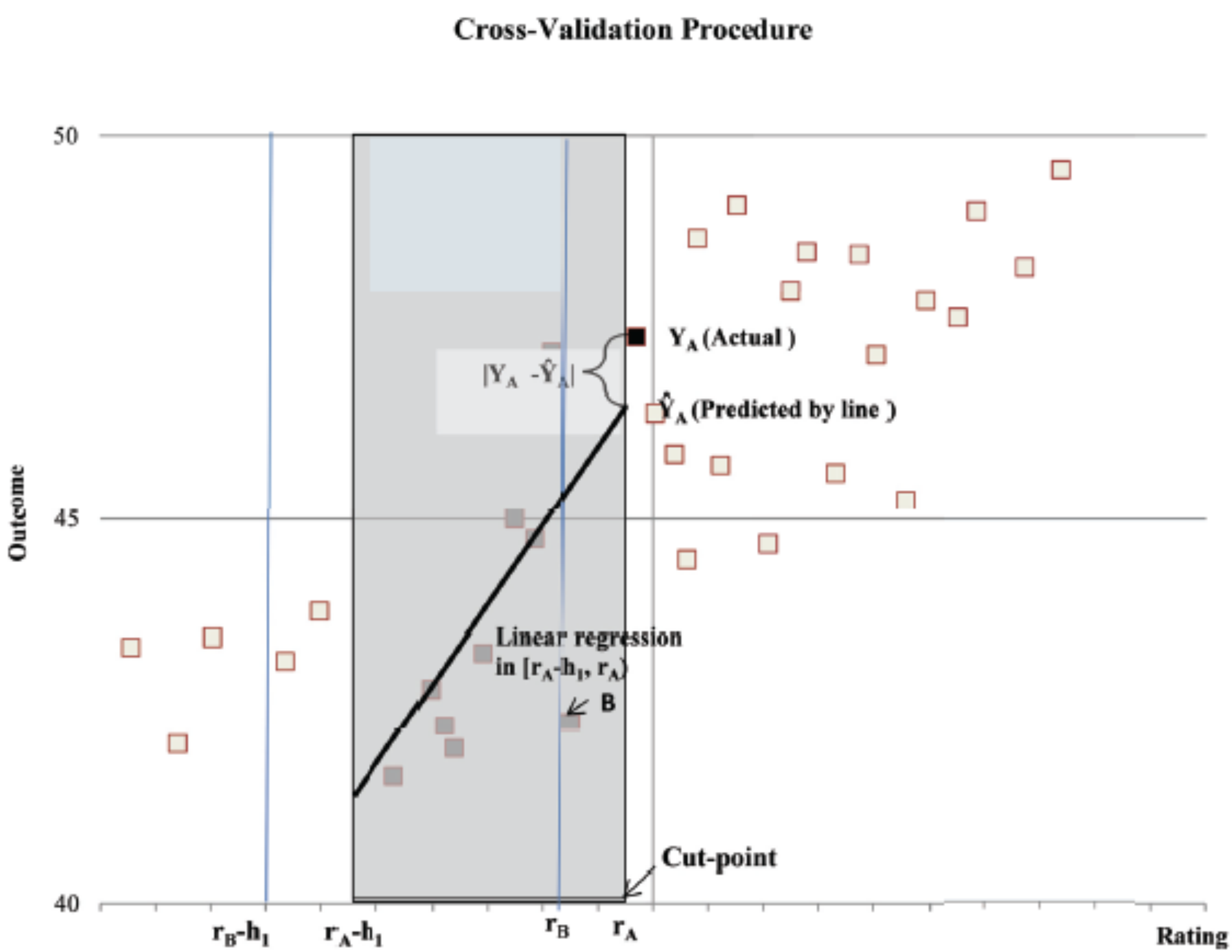
$$CV(h_1) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

7) Repeat (1)-(6) for different bandwidths
8) Choose the bandwidth that minimizes the cross-validation criterion

See Ludwig and Miller 2005 or Imbens and Lemieux 2008

More thoughts on cross-validation and local regression

- Cross-validation procedure just described is not optimized for the effect at the threshold
- More recent work by Imbens and Kalyanaraman (2012) and Calonico, Cattaneo, and Titiunik (2014) derives data-driven optimal bandwidth selection for that specific estimand
- Generally the advice is to “stay local” and keep polynomials to quadratic at most


Cross-Validation Procedure
50
45
40
Outcome
Rating
r_B-h_1
r_A-h_1
r_B
r_A
Y_A (Actual)
$|Y_A - \hat{Y}_A|$
$\hat{Y}_A$ (Predicted by line)
Linear regression
in $[r_A$-h_1, $r_A)$
B
Cut-point

Global parametric vs local non-parametric strategies

- MDRC report summarizes the choice as

"The parametric approach tries to pick the right model to fit a given data set, while the nonparametric approach tries to pick the right data set to fit a given model."

Fuzzy RDD

Fuzzy regression discontinuity

- Often these studies demonstrate "fuzzy" (versus "clean") regression discontinuity, where some units on the wrong side of the cutoff get treated or vice-versa
- Suggestions for dealing with these include dropping all crossovers (not good) or using IV (will discuss more next, can be reasonable)
- Assumptions in this new context are generally rephrased as requiring that unobserved confounding covariates vary continuously with the assignment variable over the threshold in question (may not hold if individuals or administrators can affect assignment independently of the assignment variable)

Tips to boost confidence in assumptions

- Perform density tests near cutoff: distribution of assignment variable should be similar on either side
- Plot data to make sure local-randomization is plausible.
- Check sensitivity to model specification
- Perform placebo tests on pre-determined: there should not be a RD treatment effect for pre-determined covariates or variables that could not plausibly be affected by the treatment
- Perform placebo tests on outcomes: there should not be a treatment effect on the outcome at values other than the cutoff.

RD and IV

Regression discontinuity and IV

- What's the connection?
- Recall that we can conceptualize a "clean" IV design like a randomized experiment (if we look only within a narrow interval)
- In the same way a fuzzy RD design can be conceptualized as a randomized experiment with noncompliance

Regression discontinuity and IV (cont)

- Thus the variable that is instrument is the indicator for being above or below the cutoff
- The treatment is what the individual actually did/ received
- And the outcome is the outcome
- A critical covariate to include in the model is the running variable (aka forcing variable) which is the continuous variable that determines the threshold.
- How do we conceptualize always takers and never takers in this scenario?

Regression discontinuity and IV

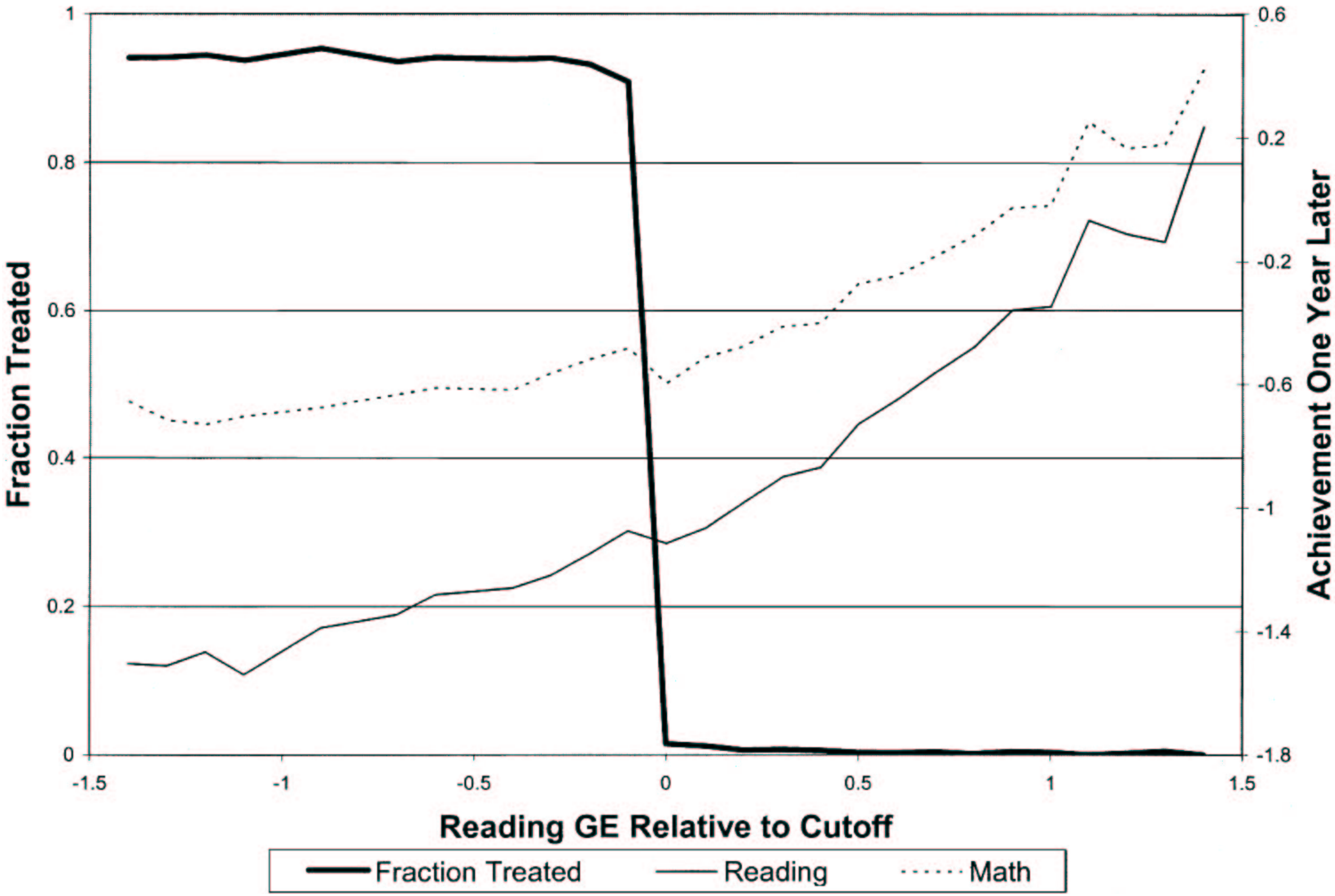
- When using IV to estimate treatment effects in an RD design we still have to satisfy the classic IV assumptions, notably
  - Ignorability (this we'd need anyway for a clean RD design) which will depend on the width of the interval around the cutoff that we are using
  - Exclusion: would those who would e.g. get the program no matter what side of the threshold they fell on remain unaffected by which side they fell on? How about those who would never get the program?

Example: the effect of holding kids back in grade

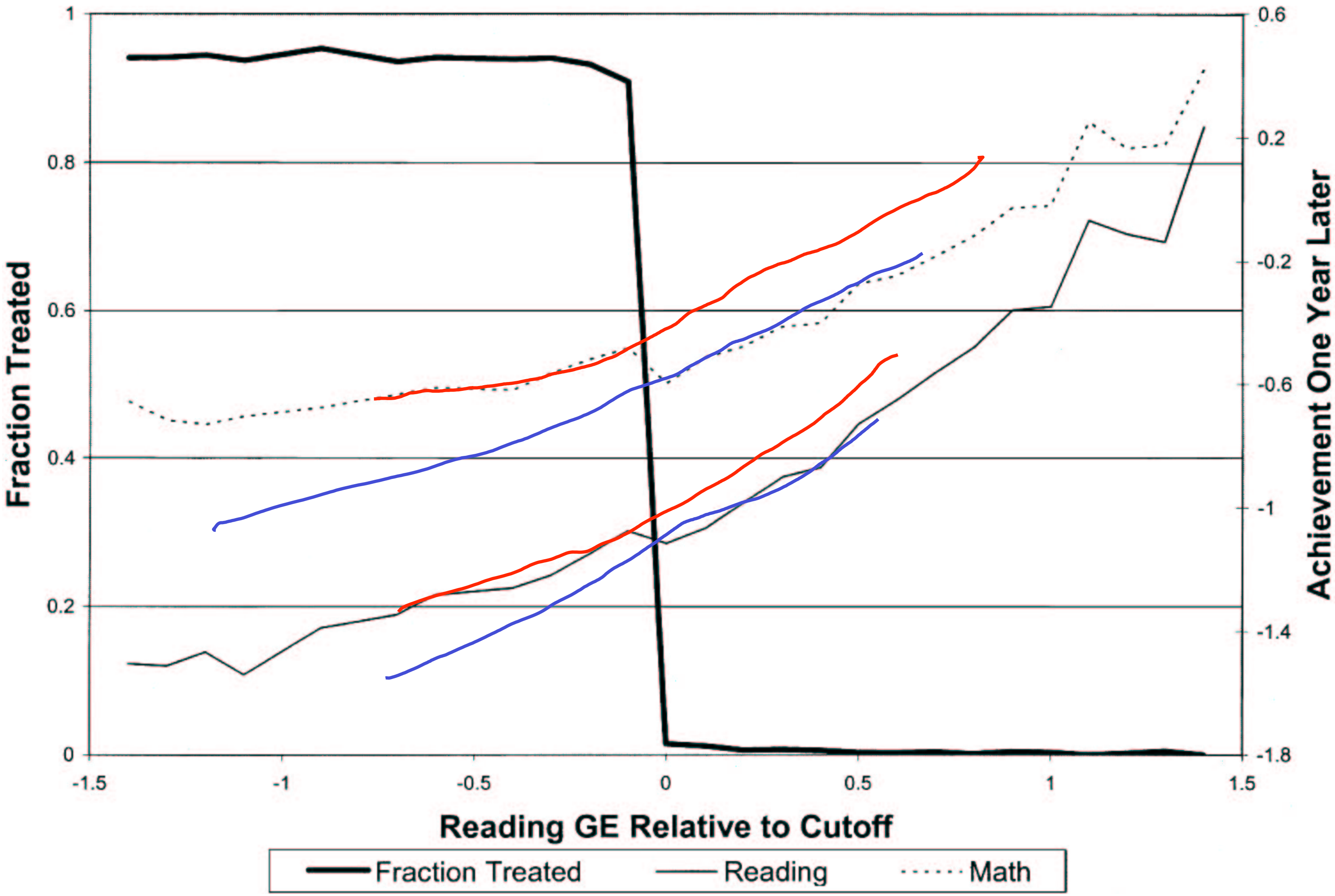
Regression discontinuity: example

- Jacob and Lefgren (2004) present an example of a retrospectively applied regression discontinuity design
- Policy from 1997-1999: In June 3rd and 6th grade children were tested in reading and math, if they failed either they had to go to summer school and retake test in August. If they failed in August they were retained.
- Here the treatment, D, was grade retention or grade retention + summer school
- The assignment variable, Z, was the score on a given high stakes test that decided (mostly) who was retained and who was allowed to progress to the next grade

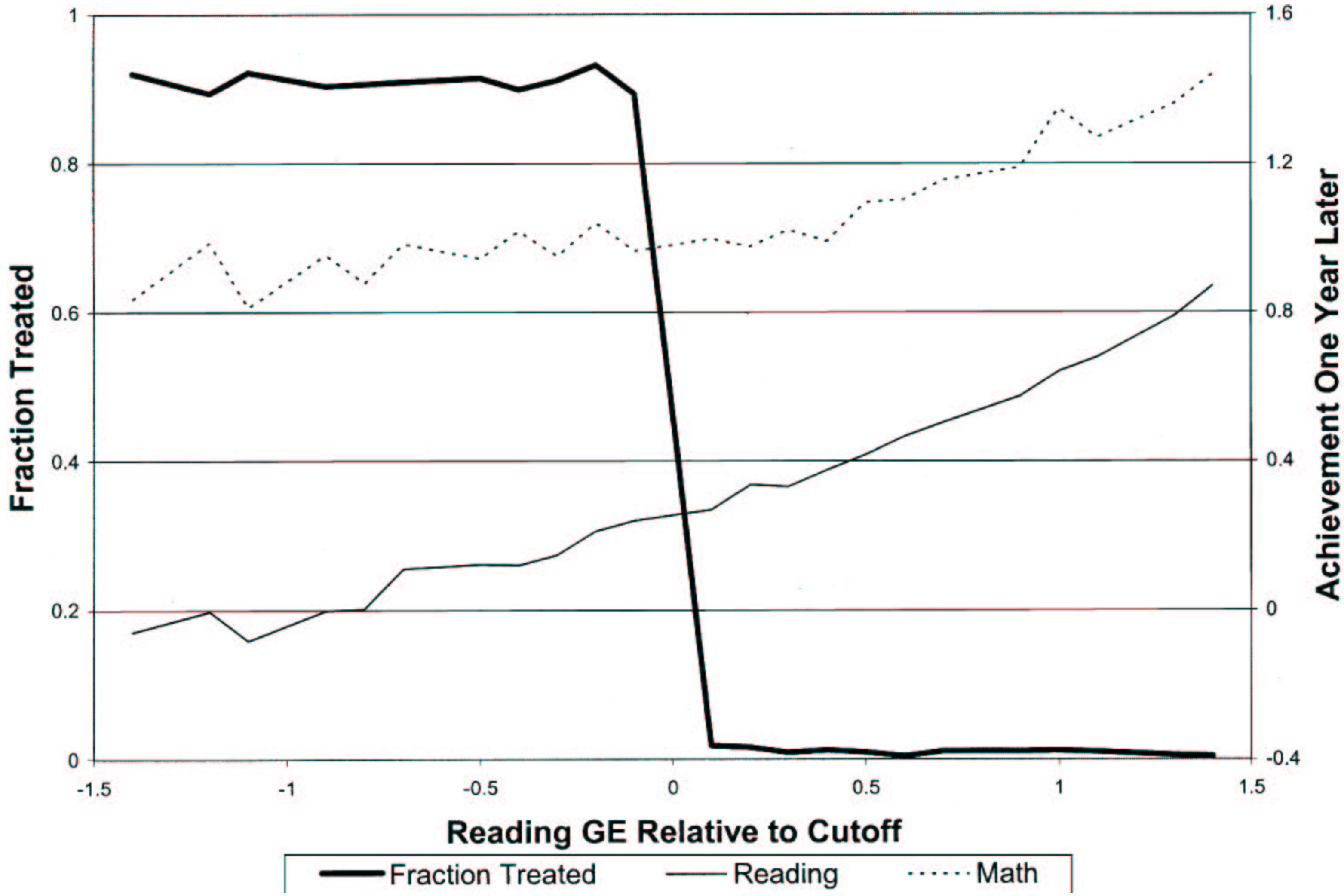
Post-test scores against Z for 3rd graders
Treatment: attending summer school & being retained



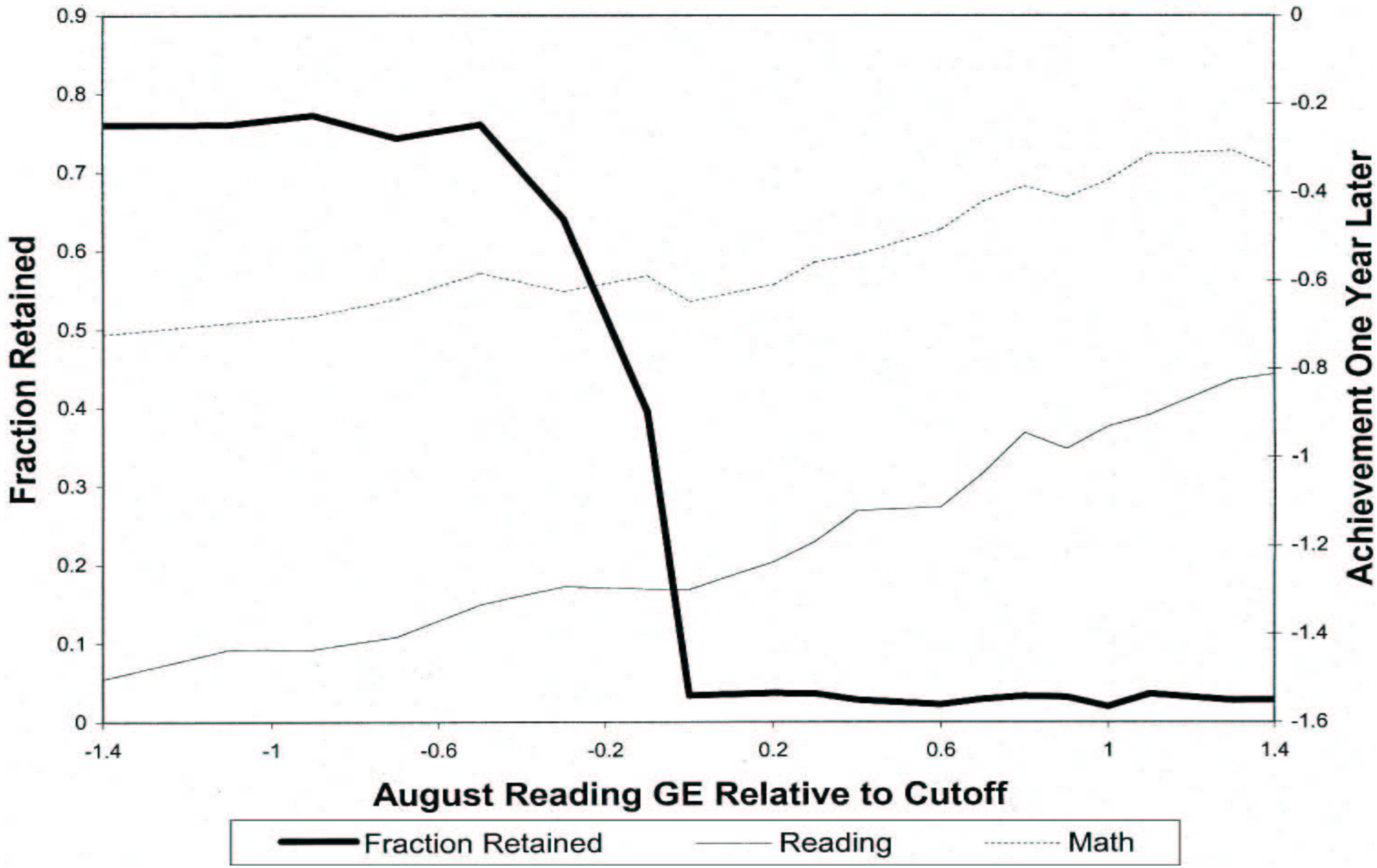
Post-test scores against Z for 3rd graders
Treatment: attending summer school & being retained



Post-test scores against Z for 6th graders
Treatment: attending summer school & being retained

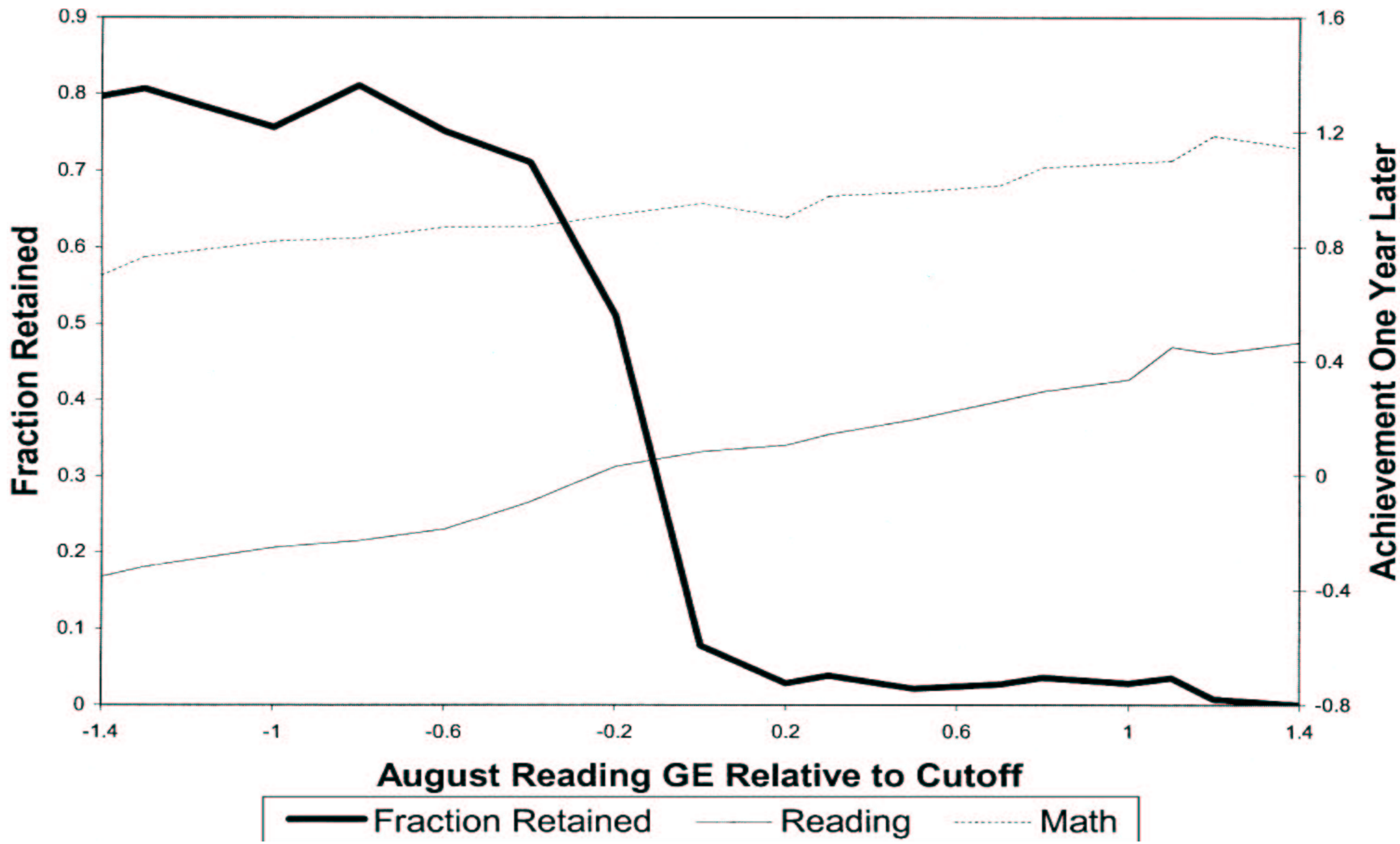


Post-test scores against Z for 3rd graders who went to summer school
Treatment: being retained



Estimated for those who failed June reading but not math

Post-test scores against Z for 6th graders who went to summer school
Treatment: being retained



Estimated for those who failed June reading but not math

Estimation

- They used a DID IV approach. So think of using at the cutoff versus just below cutoff as the instrument and “was retained” or “went to summer school and was retained” as the treatment where the outcome was the change scores (difference between post-test and pre-test scores)
- Who are the always takers and never takers in this conceptualization?

Conclusions of Jacob & Lefgren paper

- JL estimate a positive effect of combined summer school and grade retention on both reading and math scores (about 20% of a year's learning after 1 year) which drop in magnitude (by about 25-40%) but persist in statistical significance after 2 years
- Unlike most studies of grade retention, JL estimate a *positive* effect of retention alone on test scores for children retained in the 3rd grade for both reading and math scores for the first year only (41% and 31% of annual gains in reading and math respectively). Drop substantially by year 2 however
- There were no significant positive effects for 6th graders but some significant negative effects in year 2 which authors attribute to different level of importance of tests for kids in each grade (for kids who were retained it represents a high stakes test)

Comments on JL

- The effect of grade retention here is specifically for students who have already attended summer school
- Use of IV requires ignorability to be satisfied; conditioning on (continuous) pre-test makes more plausible
- Use of IV requires exclusion to be satisfied; do we believe that students who will be promoted no matter what side of the threshold their test scores land on will not be affected by being at least initially assessed (or not) as someone who should be held back?
- Sensitivity analyses do not alter primary conclusions