Regression Discontinuity Designs

Statistic Modeling & Causal Inference | Oswald & Ramirez-Ruiz

Agenda

- Lecture Review
  - Basic idea behind RDD
  - Continuity of potential outcomes
  - Falsification Checks

- RDD in R

Example case

- Effects of alcohol consumption **(treatment)** on mortality **(outcome)**
  - Carpenter & Dobkin (2009)
- **Running variable:** Age
- **Cut-off:** Minimum Drinking Age

Core Idea

Drinking

Drinking Age

- **Treatment** assigned according to a **rule** based on another variable (**running or forcing variable**)

Age

- Treated and control units may differ in their potential outcomes based on the forcing variable (non-random selection into treatment)

Old & drinking

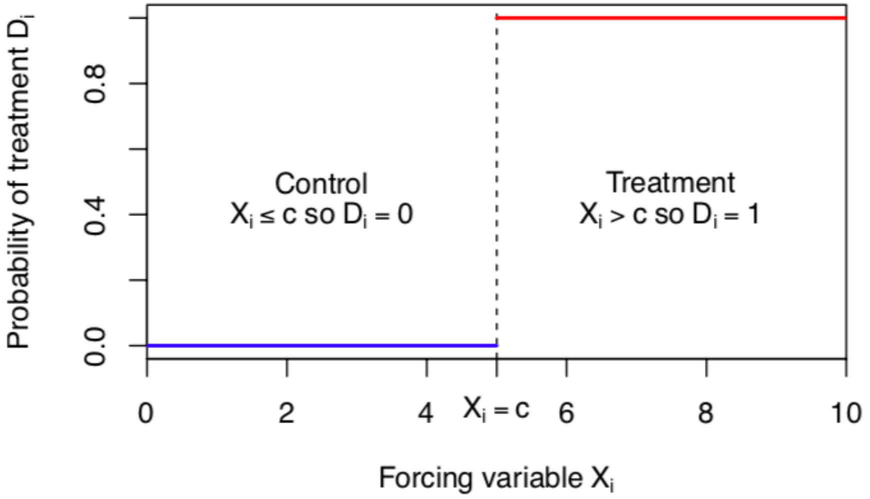
- However, whether units end up just below or just above the threshold can be assumed as a matter of chance (local randomization)
- Units around the cutoff are assumed to be similar in every way except the treatment assignment

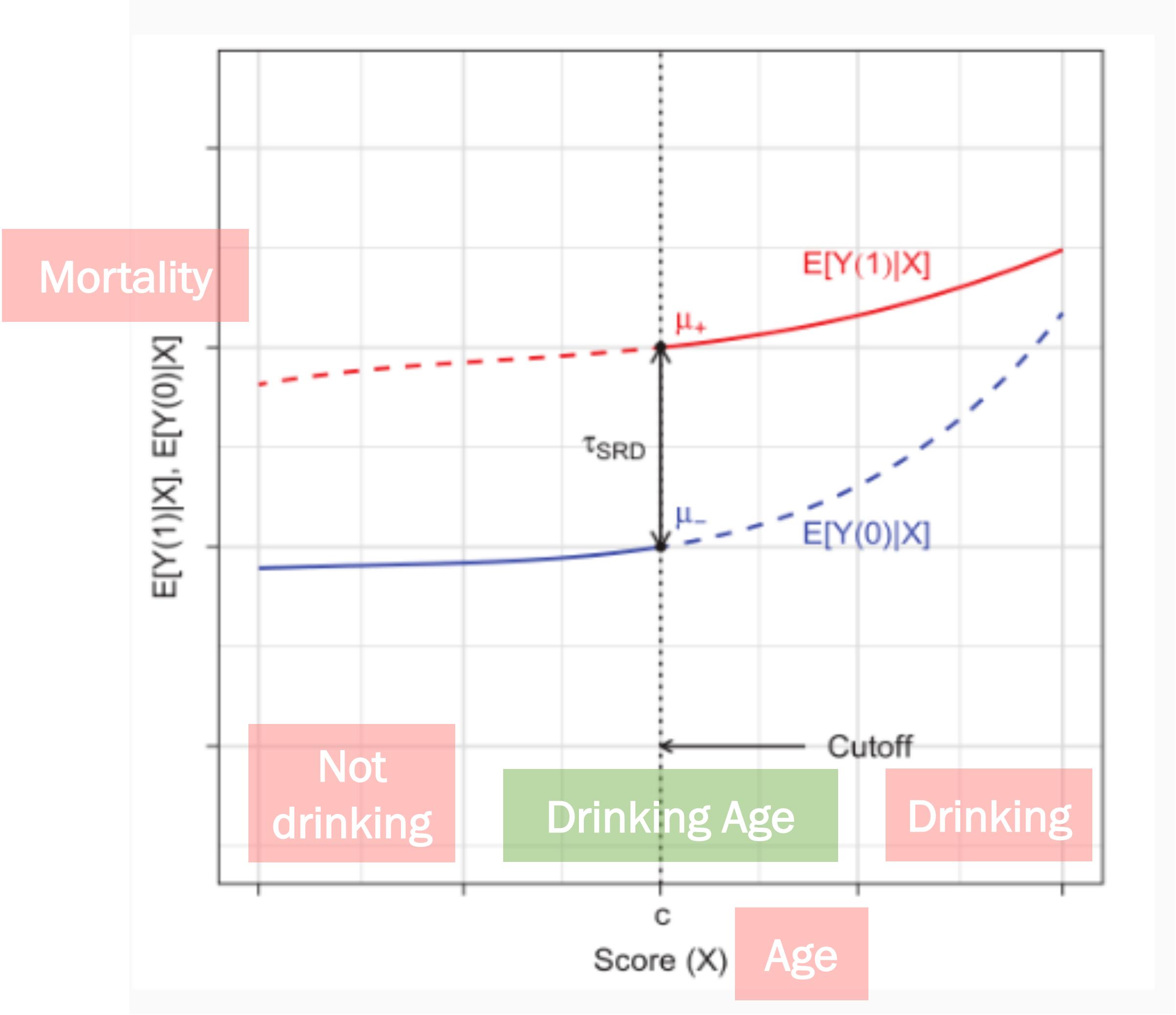
18 – 2 days /
18 + 2 days

- (Local) treatment effect can be determined by comparing cases on both sides of the cut-off

Sharp RDD

- Forcing variable (X) **perfectly determines** which side of the cut-off people are (treatment or control)
- We can only estimate the effect at a **single point**: the cutoff or threshold





Key Assumption

- Continuity of average potential outcomes (on both sides of the cut-off)

  → units on one side of the threshold have essentially the same potential outcomes from those just on the other side

- This allows us to do a tiny bit of extrapolation and estimate

  LATE at the threshold

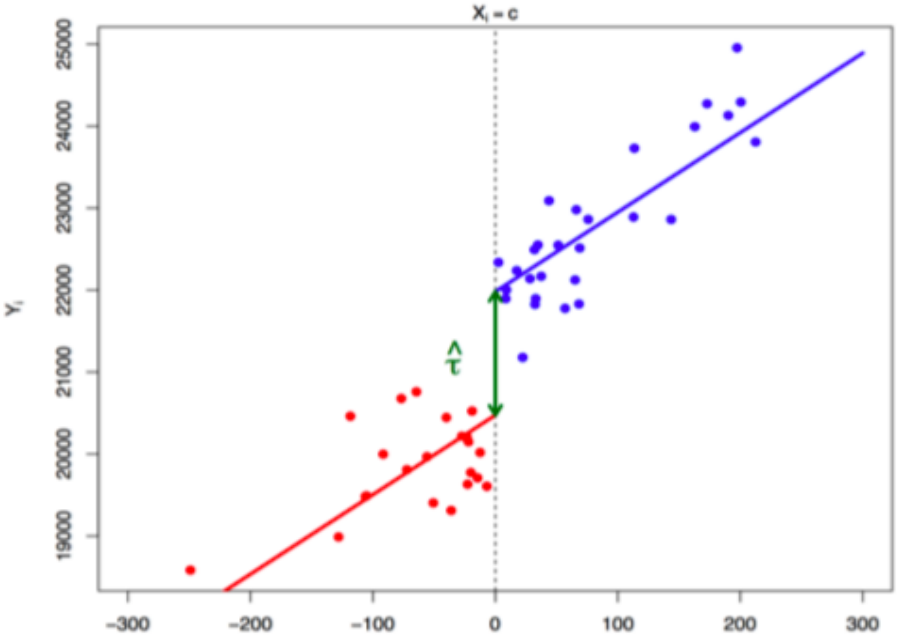
- BUT: this assumption can easily be violated:
  - For example, by some other variable driving differences at the cut-off

Estimating LATE (local polynomial approach)

- Decide which **model** is the most appropriate given the nature of the data: linear with a common slope, linear with different slopes, or nonlinear.
- Choose a **kernel** function for weighting the observations close to cutoff. (common practice: triangular)
- Choose a window or **bandwidth** (h) around the threshold (c) to create a “discontinuity sample.”
- The narrower the better, but can you afford losing many observations? (bias-variance tradeoff)
- Recode **forcing variable** X to deviations from threshold (centered on 0).
- Fit the (WLS) regression model for the observations, within the window, **above** the cutoff.
- Fit the (WLS) regression model for the observations, within the window, **below** the cutoff.
- **The local average treatment effect is the difference between the two intercepts at the cutoff.** 💡

Linear with a Common Slope

- Assumptions:
  - Potential outcomes under treatment and under control are linear in X
  - Treatment effect does not depend on the value of X_i. The effect is constant along X_i.



- In this case, we regress the observed outcome Y_i on D_i + centered X_i.

Mortality

</> 18

Age

Model is $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \epsilon_i$

<18

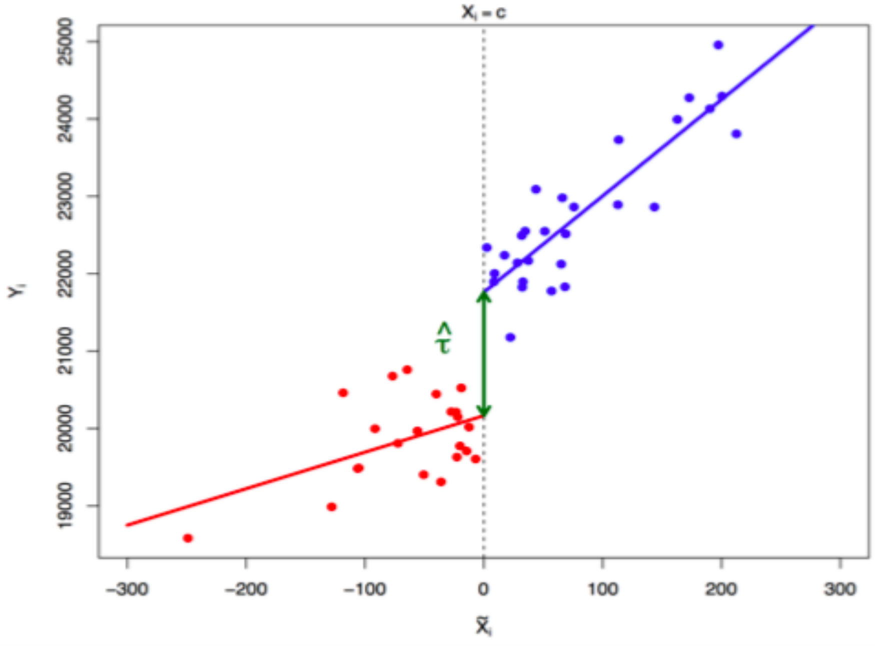
>18

$$\begin{cases} E[Y_{0i}|X_i] = \beta_0 + \beta_1 * X_i \\ E[Y_{1i}|X_i] = \beta_0 + \tau + \beta_1 * X_i \end{cases}$$

Linear with Different Slopes

- Assumptions:
  - Potential outcomes under treatment and under control are linear in X
  - **Treatment effect can vary for different values of X_i.**

- In this case, we regress the observed outcome Y_i on the interaction **D_i * X_i.**



Mortality

</> 18

Age

Interaction

Model is $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \phi D_i X_i + \epsilon_i$

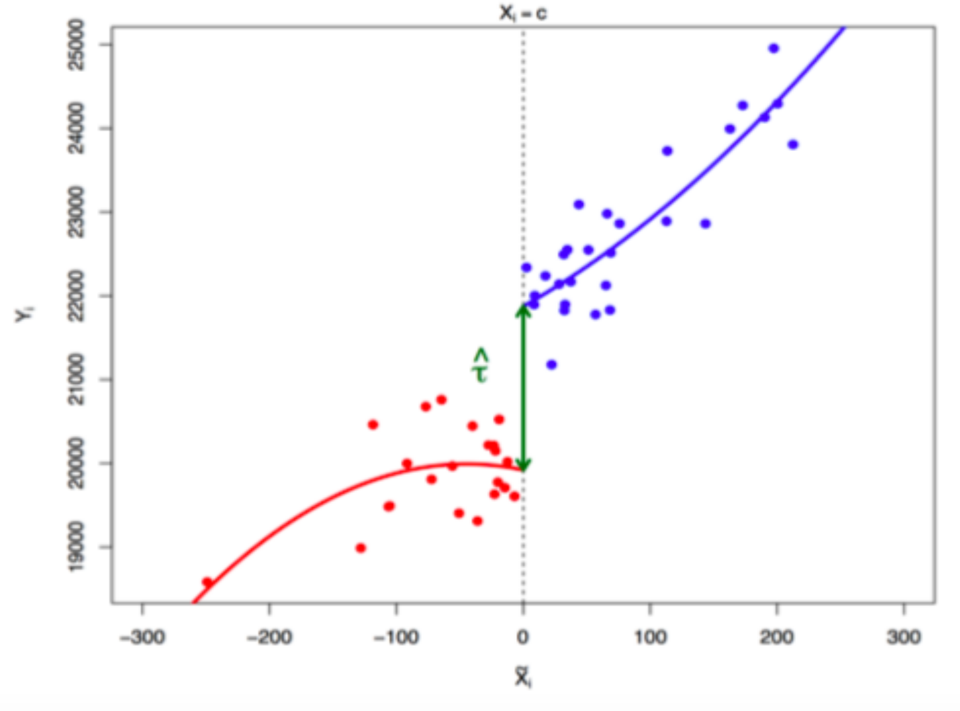
<18

>18

$$\begin{cases} E[Y_{0i}|X_i] = \beta_0 + \beta_1 * X_i \\ E[Y_{1i}|X_i] = \beta_0 + \tau + (\beta_1 + \phi) * X_i \end{cases}$$

Non-linear

- Assumptions:
  - **Potential outcomes are allowed to be non-linear in X but must be correlty specified**
  - Treatment effect can vary for different values of X_i.
- Model can include quadratic, cubic, etc. terms in Xi and their interactions with Di in the equation.



! Be cautious about high-order polynomials: they are difficult to fit, make lots of assumptions about the data, and are sensitive to outliers.

Polynomial

New Interaction

Model: $Y_i = \beta_0 + \tau D_i + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i D_i + \beta_4 X_i^2 D_i + \epsilon_i$

And how do I specify my model? 😟

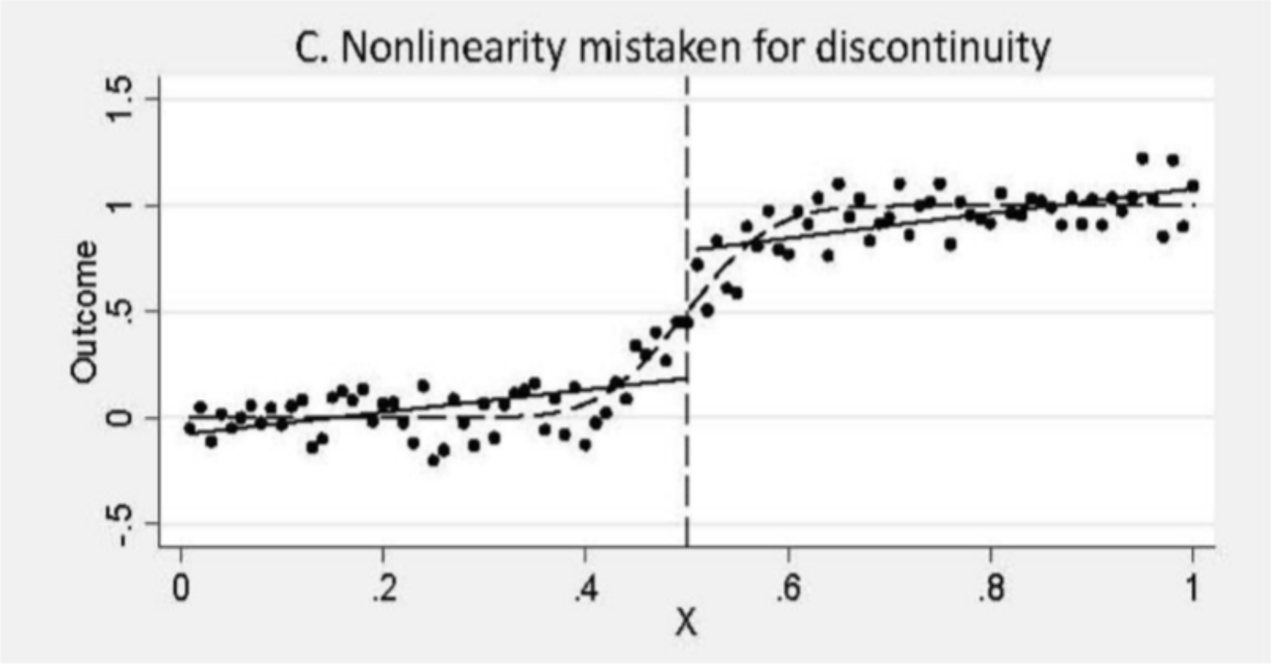
- Model specification is a trade-off between **bias** and **variance**
  - If you choose nonlinear, you might reduce variance because you can pick up every sensitivity in the data, but estimates will be biased due to following “noise.”
- Standard practice: Try and **compare different specifications** to show robustness
  - Ideally you are looking for similar results across different models.
- Always start with a visual inspection: see scatterplot and run a local regression (such as LOWESS) to guide choice
- Remember each model corresponds to a particular set of assumptions about the POs.

Falsification Checks

Sensitivity:

Are results sensitive to alternative specifications?

- Nonlinear relation ≠ discontinuity
- If units start curving up near lower threshold and down near upper, it might just be non-linearity vs. a discontinuity jump.

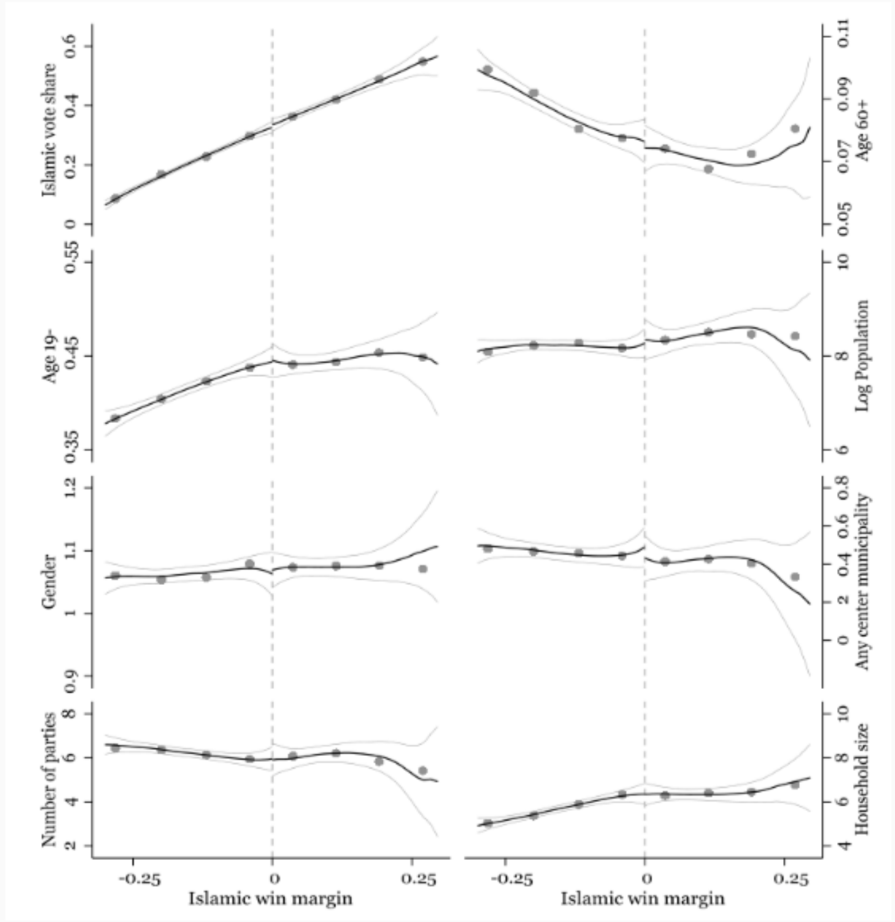


Falsification Checks

Balance checks:

Does any covariate Zi jump at the threshold?

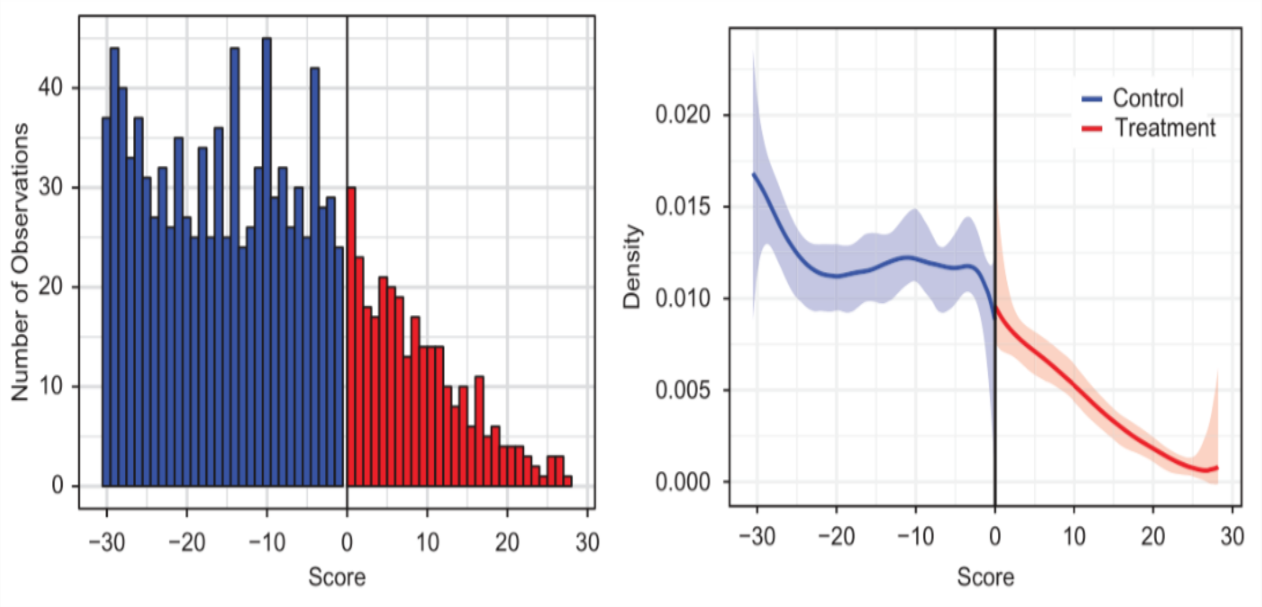
- Aiming for a scenario where individuals are pretty much identical except for treatment ‘assignment’.
- We should only see a jump in Y, not on other **pre-treatment or post-treatment** (not affected by treatment) variables.



Falsification Checks

Sorting:

- Do units sort around the threshold? Is there a jump in number of observations around the cut-off?
  - Sometimes there is an incentive to end up above or below a threshold. An agent's behavior can invalidate the continuity assumption. Local randomization would not hold.



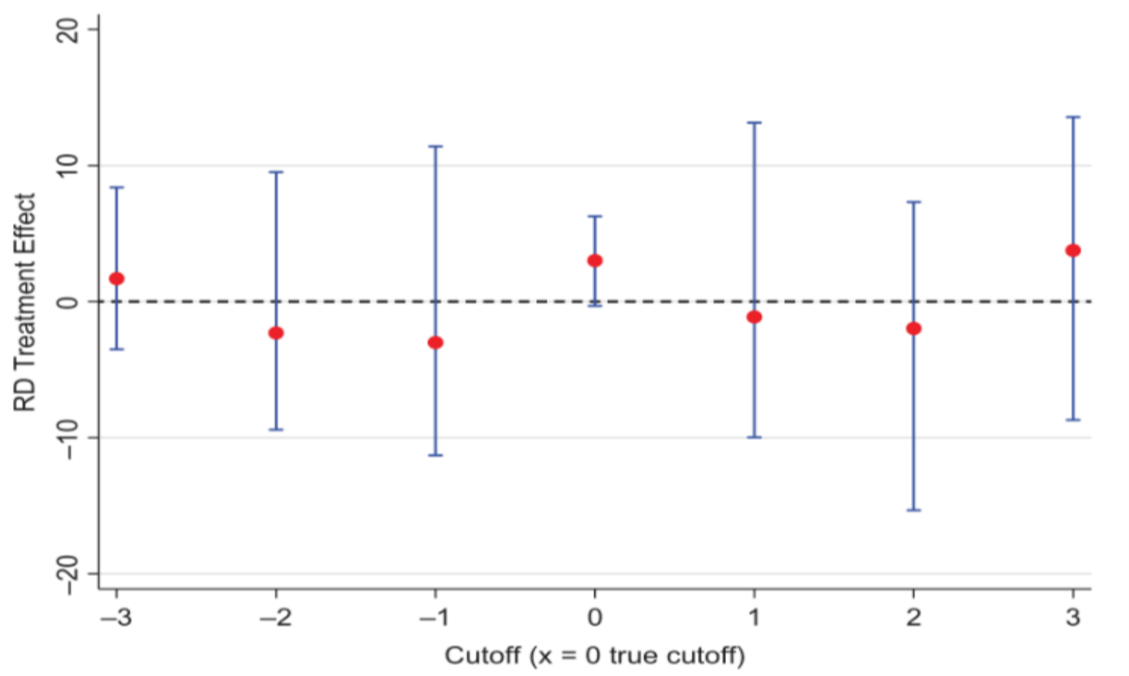
(Cattaneo et al., 2019)

Falsification Checks

Artificial cut-off values:

Do jumps occur at placebo thresholds?

- If they do, this could mean something else is going on that could challenge our research design.



(Cattaneo et al., 2019)

Falsification Checks

Sensitivity to cases near cutoff:

Do results change if we exclude cases near the threshold?

- Remember the different weights in the kernel definition.
- If self selection into treatment took place, the units closest to the cutoff would be the most likely units to engage in it.

Sensitivity to bandwidth choice:

Do results change if we specify the bandwidth differently?

Further Resources

For any coding issues – Stackoverflow

Hertie’s Data Science Lab – Research Consulting