Statistics II

Week 7: **Regression Discontinuity Designs**

1. Review of core concepts from lecture
2. RDD in R

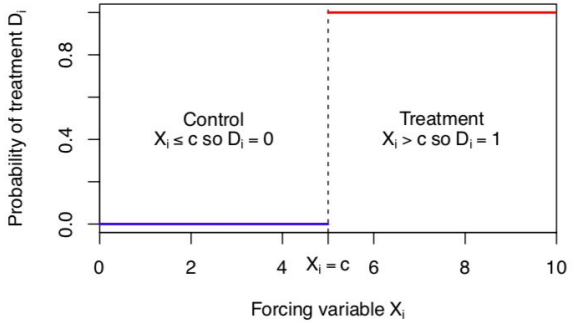
Lecture Review

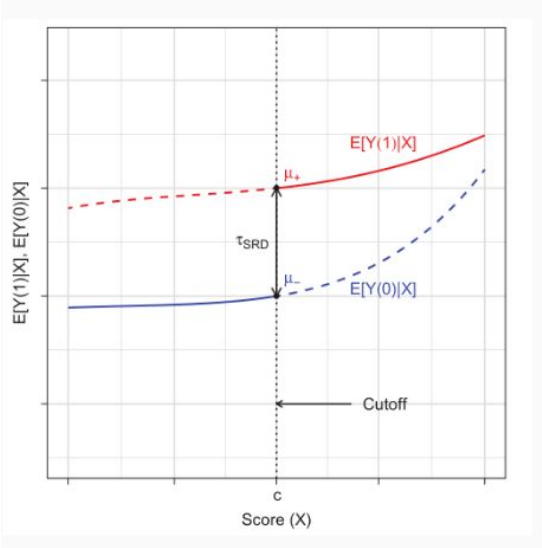
Core Idea

- In some circumstances, treatment is assigned according to a rule based on another variable (called the **forcing** or **running** variable).
    - Eg: legally drinking for ages above 21, or winning an election with more than 50% of vote share.
- Treated and untreated units may differ in their potential outcomes based on the forcing variable (non-random selection into treatment).
- However, whether units end up *just* below or *just* above the threshold, can be assumed as matter of chance (**local randomization**). Units around the cutoff might be similar in every way except for treatment assignment.
- Treatment effect is determined by comparing those just on either side of the cut-off.

Sharp RDD

- In sharp RDD, the forcing variable (X) **perfectly determines** which side of the cut-off people are on (treatment or control).
    - For example, being over or under the age of 21 (in the US) determines whether or not you are eligible to legally buy alcohol.
- We can **only** estimate the effect at a single point: the **cutoff** or threshold.





Key Assumption

- **Continuity of average potential outcomes**: Average potential outcomes should be continuous on both sides of the cut-off. Units on one side of the threshold need to be essentially the same as units on the other side.
- The continuity assumption allows us to do a tiny bit of extrapolation and estimate **LATE at the threshold**.
- However, this assumption can easily be violated: It could be that the potential outcomes are actually not continuous and there is some other variable driving differences at the cutoff point.
    - For example, you may be incentivized to report your income just below a threshold for government support - this sorting violates our assumption.

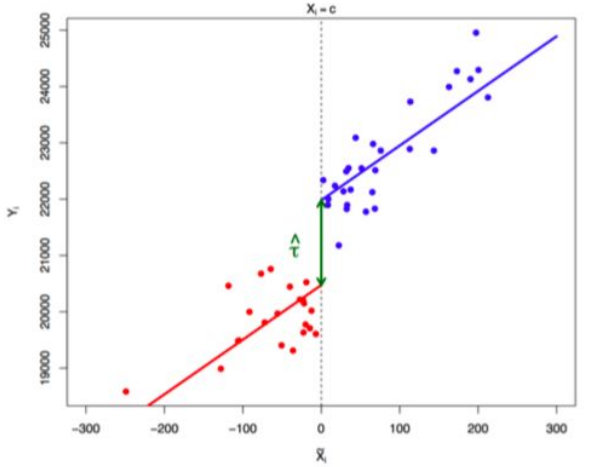
Estimating LATE (local polynomial approach)

- **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 Xi. The effect is constant along Xi.



In this case, we just regress the observed outcome Yi on **Di + centered Xi**.

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

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

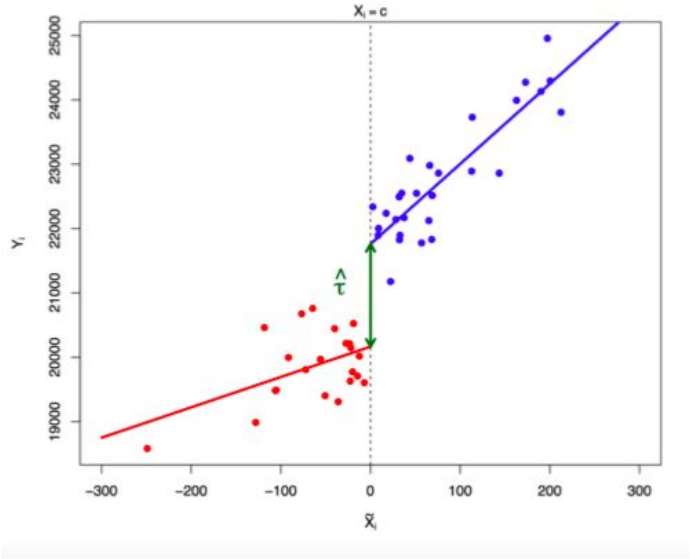
Linear with a Different Slope

Assumptions:

- Potential outcomes for treatment and control groups are both linear in X
- But treatment effect can **vary** in Xi.

We regress Yi on **the interaction Di*Xi.**

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



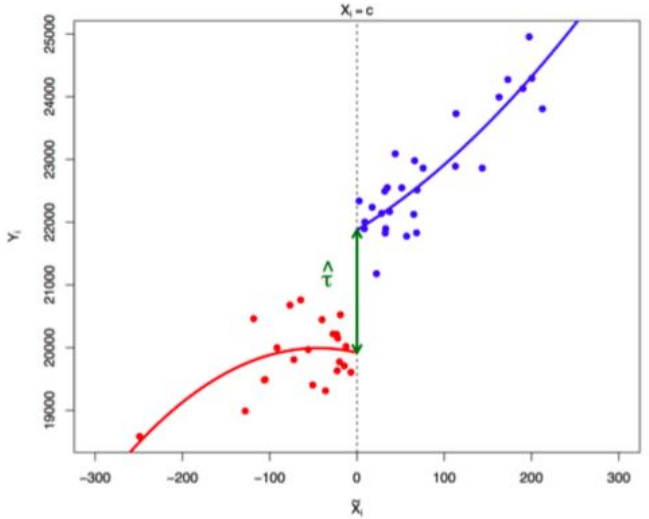
$$\begin{cases} E[Y_{0i}|X_i] = \beta_0 + \beta_1 * X_i \\ E[Y_{1i}|X_i] = \beta_0 + \tau + (\beta_1 + \phi) * X_i \end{cases}$$

Non-linear

Assumptions:

- Potential outcomes are now allowed to be non-linear in x, but must be correctly specified.
- Treatment effect is allowed to vary across Xi

Can include **quadratic, cubic**, etc. terms in Xi and their interactions with Di in the equation.



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

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

How to choose a model specification?

- 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,

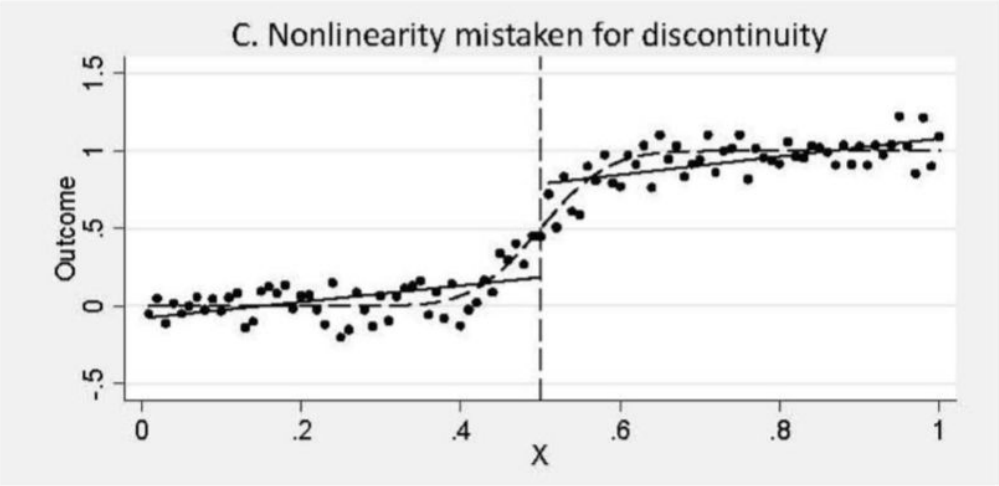
Remember each model corresponds to a particular set of assumptions about the POs.

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.

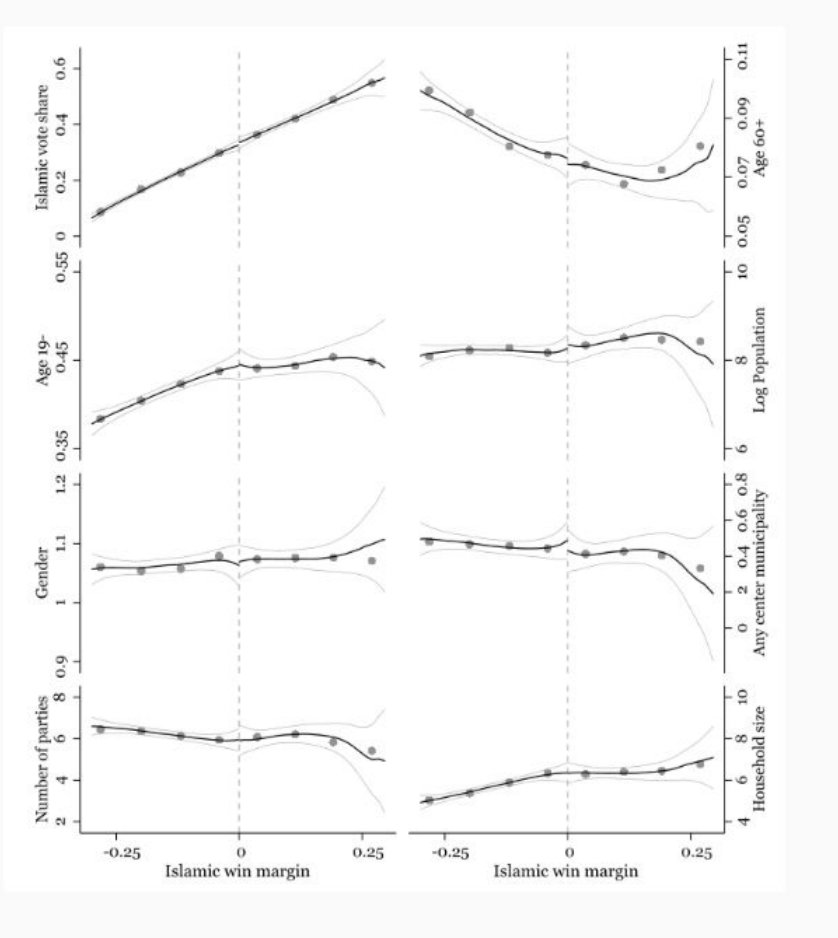


Falsification Checks

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.

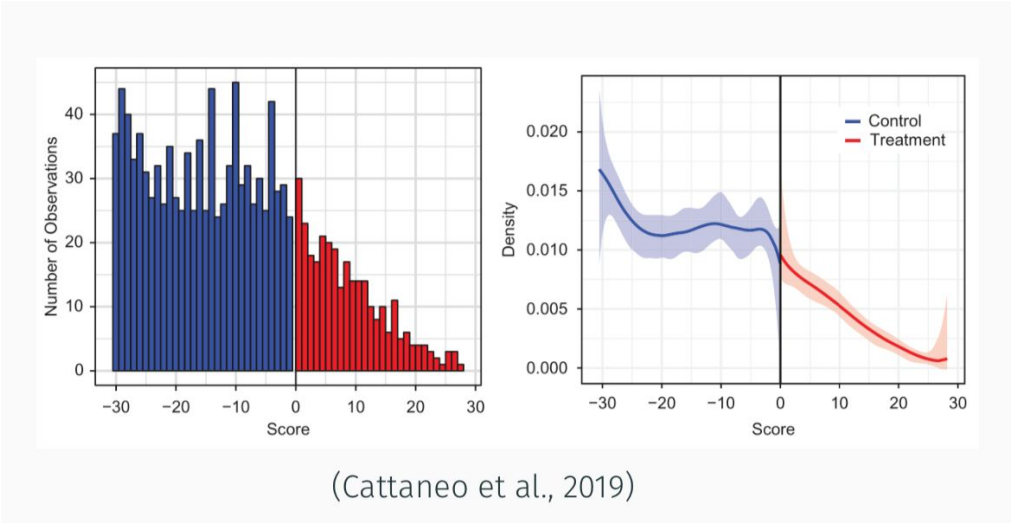


Falsification Checks

Sorting:

Do units sort around the threshold? Is there a jump in number of observations around c?

- 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.



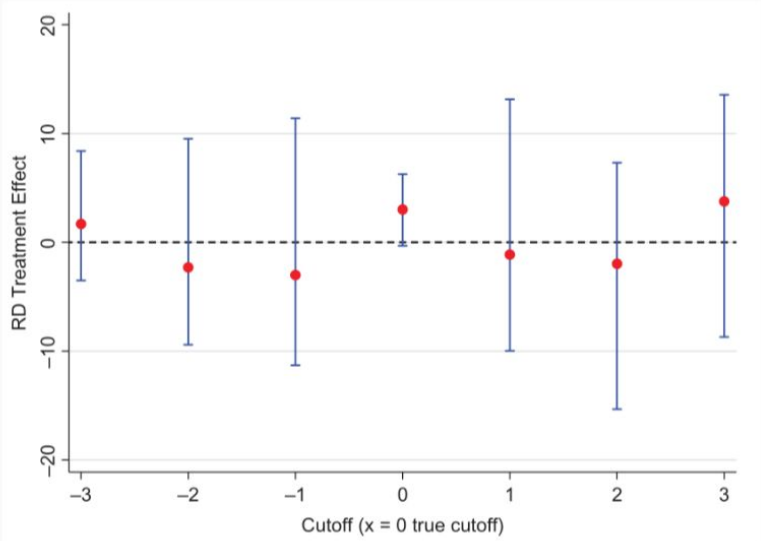
(Cattaneo et al., 2019)

Falsification Checks

Artificial cutoff values:

Do jumps occur at placebo thresholds c∗?

- 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?

Questions?