Statistics II

Week 6: **Regression Discontinuity Designs**

Content for Today

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

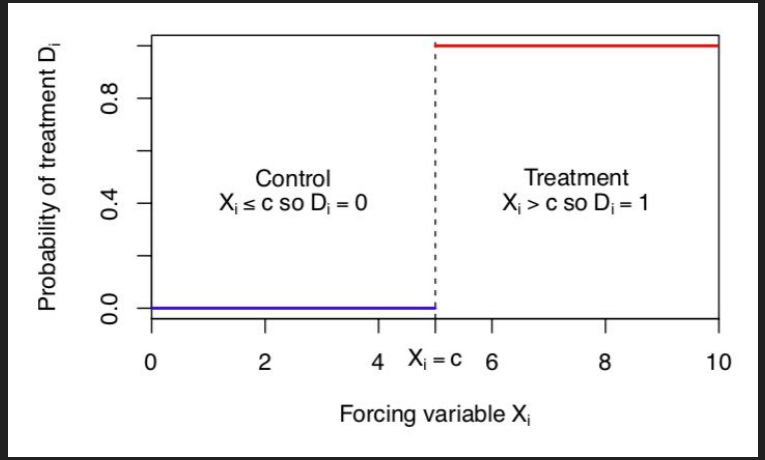
Lecture Review

Core Idea

- Treatment is assigned according to a rule based on another variable (called the **forcing** or **running** variable).
- Treated and untreated units may differ in their potential outcomes (non-random selection into treatment) based on the forcing variable. There might be other factors that determine the forcing variable outcome.
- However, whether units end up just below or just above the threshold is assumed to be a matter of chance (**local randomization**).
- Treatment effect is determined by comparing those just on either side of the cut-off.

Sharp RDD

- In sharp RDD, our forcing variable **X perfectly determines** which side of the cut-off you 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.



Key Assumption

- **Continuity of average potential outcomes**: basically, units on one side of the threshold need to be essentially the same as units on the other side. Average potential outcomes should be continuous on both sides of the cut-off.
- 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

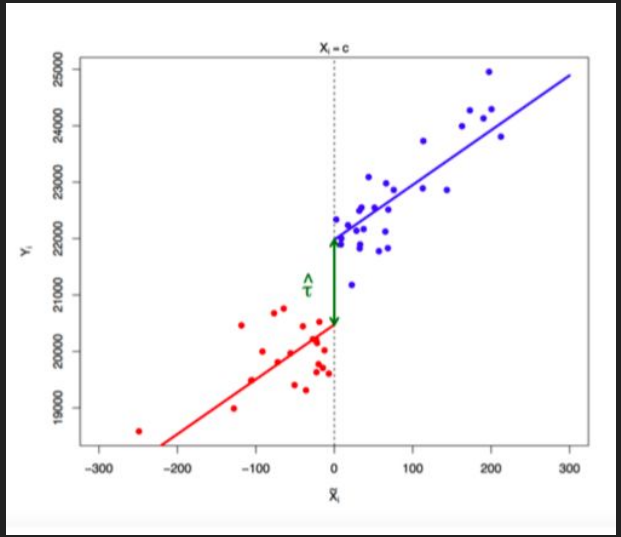
- **Choose a window** around the threshold c to create a “discontinuity sample.”
  - The narrower the better, but can you afford losing many observations?
- **Recode forcing variable** X to deviations from threshold (centered on 0).
- **Decide which model** is the most appropriate given the nature of the data: linear with a common slope, linear with different slopes, or non-linear.

Linear with a Common Slope

Assumptions:

- PO under control is linear in X
- Treatment effect does not depend on the value of Xi

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

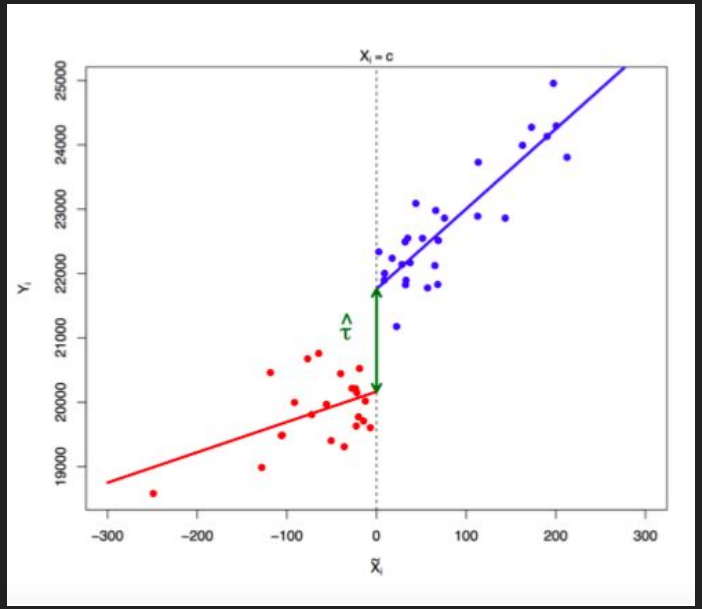


Linear with a Different Slope

Assumptions:

- POs for treatment and control groups are both linear in X
- But we now allow treatment effect to vary with Xi

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



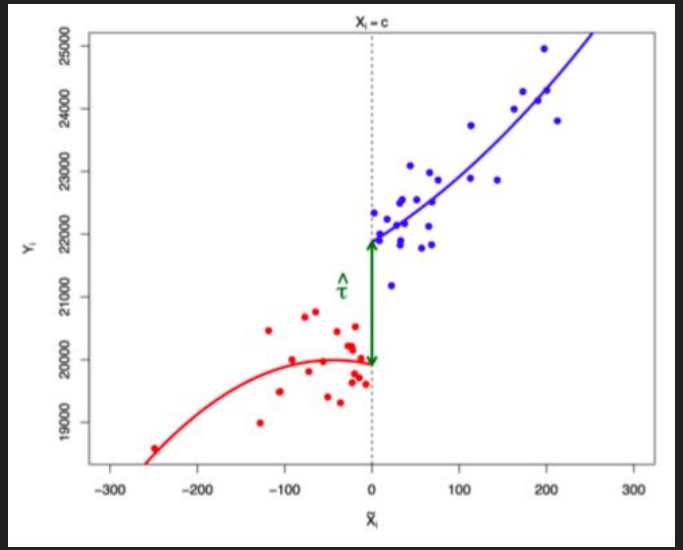
Non-linear

Assumptions:

- POs 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.

- 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**
  - When you go non-linear 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 looking for similar results across different models.
- Do **local regression** (such as LOWESS) to guide choice

Falsification Checks

1. **Sensitivity**: Are results sensitive to alternative specifications?
    a. Nonlinear relation ≠ discontinuity
    b. If units start curving up near lower threshold and down near upper, it might just be non-linearity vs. a discontinuity jump.
2. **Balance checks**: Does any covariate Zi jump at the threshold?
    a. Aiming for scenario where we are comparing individuals that are pretty much identical except for what side of the cut off they ended up on. Only want to see a jump in Y, no other variables.
3. Do jumps occur at **placebo thresholds** c∗?
    a. If yes, this could mean something else is going on that could challenge our research design.
4. **Sorting**: Do units sort around the threshold?
    a. Sometimes there is an incentive to end up above or below a threshold. An agent's behavior can invalidate the continuity assumption

Questions?