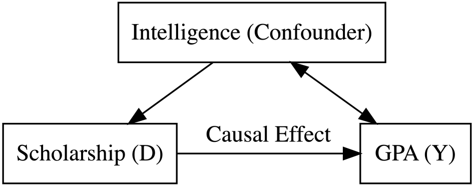
Regression Discontinuity I: Sharp RDD

Lecture 7 - Introduction to Causal Inference

Kevin Li

Our Issue

Issue: We want to find the effect of treatment D on outcome Y, but there is a confounder.



Let us say a random experiment is not possible (we do not control who gets or does not get the scholarship).

- How can we get the treatment D to be exogenous to get the causal effect?

Assignment Cutoff

Let us say that the scholarship is assigned to people who score above 90% on some standardised test.

- 90% is our **assignment threshold**.

Now, compare individuals who scored 89.9% on the exam, and 90.0% on the exam.

- Are these individuals very different from each other in terms of intelligence - no! Only 0.01% seperates them.
- In fact, the people who scored 89.9% are likely on average the same as people who scored 90.0%. There was just some sort of luck/randomness (bad night's sleep, stupid mistake) that caused some to get 90% and others to get 89.9%.

Cutoff and Exogeneity

People just below the cutoff for treatment (89.9%) and just above the cutoff (90.0%) are on average the same, and getting either 89.9% or 90.0% is just some random luck.

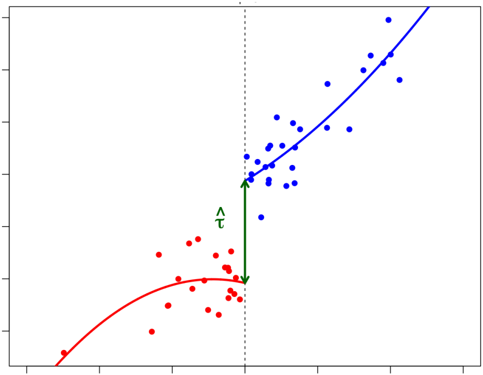
Thus, right above and below the cutoff for receiving the treatment (scholarship), treatment assignment is approximately random/exogenous.

Regression Discontinuity: Compare the people right below and right above the cutoff.

- Since treatment is approximately exogenous right below and above the cutoff, we should be able to get the causal effect.

Graphical Identification

Since individuals slightly above and below the threshold are similar, their academic performance should also be similar.



But if there is a noticable "jump" in academic performance between 89.9% and 90.0%, that must be result of treatment. So the jump is our causal effect $\hat{\tau}$.

Regression Discontinuity

The setup of regression discontinuity is as follows:

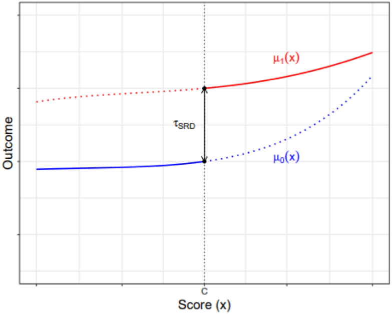
- We have a treatment D and a outcome Y.
- Treatment D is assigned based on some **running** variable X (like test-scores). Some cutoff $X = c$ determines if a unit is treated or not treated.

We will usually create a new adjusted running variable $\tilde{X}$. This is basically the original X variable, but adjusted so the cutoff value for treatment assignment is $\tilde{X} = 0$, anything $\tilde{X} \geq 0$ means treatment is assigned, and $\tilde{X} < 0$ means no treatment.

- We will explore non-compliance (so when not all units follow the cutoff) in lecture 10.

Assumption: Continuous Potential Outcomes

We identify the causal effect in RDD by the "jump" (slide 5)



Assumption: if no treatment occurred, Y would have not "jumped" at the cutoff $X = c$ (basically $Y_i(0)$ is continuous).

- If Y value jump at the cutoff $X = c$ even without treatment, then we do not know if the jump we observe is because of the treatment D, or of some other reason.

Sorting

We know RDD requires the assumption that if no treatment occurred, Y would have not "jumped" at the cutoff $X = c$ (basically $Y_i(0)$ is continuous).

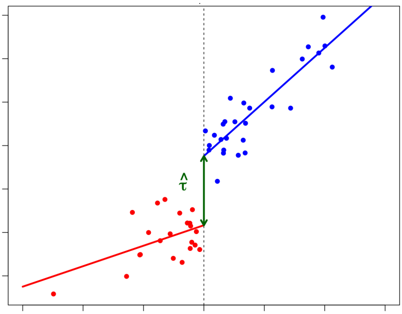
The most common reason this is violated is because **sorting**:

- Sorting is when individuals i can **manipulate** their position above/below the treatment.
- For example, if a welfare programme only allows £20,000 salary or lower individuals, some people at £22,000 might intentionally lower their salary to qualify.
- But these £22,000 salary people might be different from the people at £19,999, which messes up the continuity of Y.

The McCary density test can check if sorting is occuring.

Estimating Causal Effects

We use a **local linear regression** (machine learning) to get best-fit lines on both sides of the cutoff. Then, we estimate the jump.



Local Linear Regression differs from normal linear regression, because we can weight different points differently (Kernel).

- In RDD, we weight points closer to the cutoff more heavily, so we can more accurately estimate the jump

Bandwidth Selection

We are interested in the jump at the cutoff. So, it is not very useful to include data from individuals far from the cutoff.

There is a tradeoff in selecting bandwidths (how much data around the cutoff to use).

- For maximum accuracy, we want as small bandwidth as possible - only consider people with scores 89.9% and 90.0%.
- However, only using as small bandwidth as possible reduces the amount of data/sample size we have, which creates much higher variance/uncertainty.

Solution: Catanneo et al (2020, 2024) propose using the bandwidth that minimises the mean squared error of the predictions of the best-fit lines (since lowest MSE means most accurate fitted lines).

Parametric Estimators

While the most common RDD estimators use the machine learning local linear regression, we can also fit a normal linear regression or polynomial on both sides of the cutoff.

- Linear lines: $Y = \alpha + \tau D + \beta_1 \tilde{X} + \beta_2 \tilde{X} D + \varepsilon$
- Polynomials:
  $Y = \alpha + \tau D + \beta_1 \tilde{X} + \beta_2 \tilde{X}^2 + \beta_3 \tilde{X} D + \beta_4 \tilde{X}^2 D + \varepsilon$

You should not do this unless you have very small sample sizes, in which local linear regression performs worse.

- It can be useful if you are dealing with panel data (fixed effects are easy in these models). But, you can also manually do fixed effects (de-mean) and then run a local linear regression estimator.