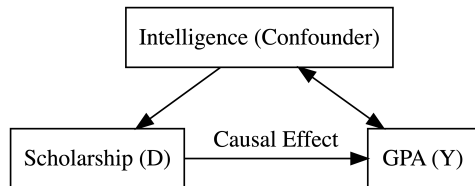# Regression Discontinuity I: Sharp RDD

Lecture 8 - Introduction to Causal Inference

Kevin Li

Setup

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

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

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

Cutoff and Exogneity

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%.

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

▶ Since D at cutoff is exogenous, we can find treatment effect.

Graphical Identification

Since individuals slightly above and below the threshold are similar, their academic performance should also be similar.
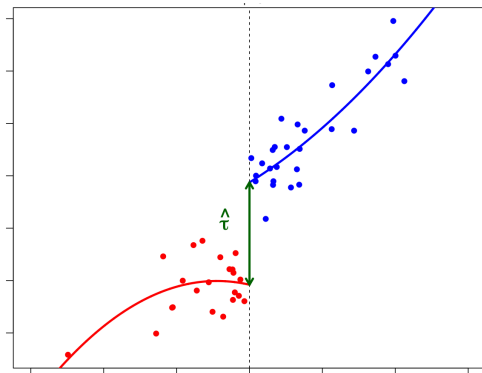
But if there is a noticable "jump" in academic performance between 89.9% and 90.0%, that must be result of treatment. So the jump is our causal effect $\hat{\tau}$.

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) with fuzzy discontinuity in lecture 9.

Assumption: Continuous Potential Outcomes

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.
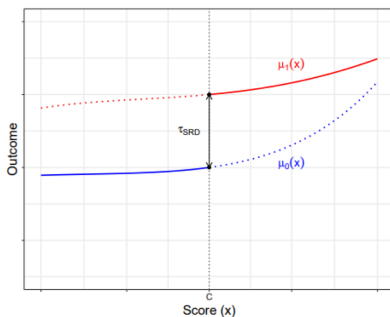
(We also assume $Y_i(1)$ is continuous for same reason)

Sorting

We know RDD requires the assumption that the potential outcomes are 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. Cattaneo et al (2024) also have a test for sorting.

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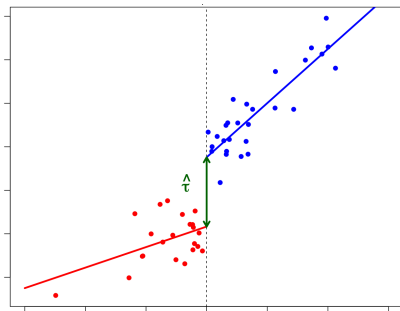
Local Linear Regression differs from normal linear regression, because we can weight different points differently (Kernel).

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

Local Average Treatment Effect

What is the estimated $\hat{\tau}$ we get from looking at the jump/discontinuity? It is the **local average treatment effect** (LATE)

Recall the LATE is the average treatment effect for a set of individuals i who meet a set of criteria.

▶ In RDD, the estimated $\hat{\tau}$ is the average treatment effect for the units **exactly at the cutoff** $X = c$.

▶ We can usually generalise the LATE to individuals nearby the cutoff $X = c$ with no issues.

▶ However, the estimated LATE cannot be generalised to all individuals in our study - it is possible people far from the cutoff $X = c$ will have drastically different treatment effects.

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 Estimator

Generally, we prefer the local regression (machine learning) approach unless we have very mimimal data. But we can fit normal linear regression lines on both sides of the cutoff as follows:

$$Y_i = \alpha + \tau D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i D_i + \varepsilon_i$$

▶ The estimate of τ is our causal estimate of the LATE.

▶ β_1 is the slope of the line fit for values $\tilde{X}_i < 0$ (below the cutoff, the side with no treatment).

▶ $\beta_1 + \beta_2$ is the slope of the line fit for values $\tilde{X}_i > 0$ (above the cutoff, the side with treatment).

We can also fit polynomials in a similar way (for quadratic: two terms $\tilde{X}$ and $\tilde{X}^2$, then the same two terms interacted with D).