

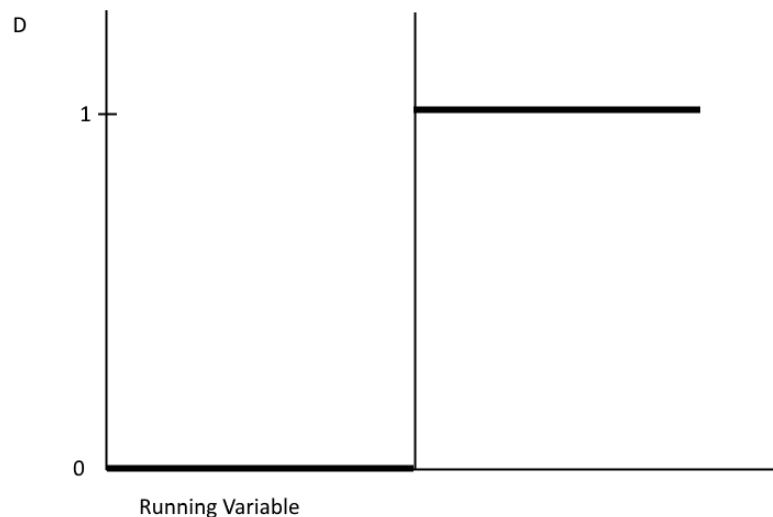
Lecture 5 – Regression Discontinuity

Intuition

Often government programs or other treatments are assigned on the basis of a score on a scale. For instance, the Pell Grant depends on Expected Family Contribution, admission to public 4 year schools in Georgia is based on the SAT, DUIs are based on a set threshold, etc.

It may be surprising that these rules can generate “as if random” variation. The basic intuition is if people can’t pick their score on the scale, which side of the cutoff they fall on is close to random. This intuition is formalized in Cattaneo Frandsen Titiunik (2015).

Let’s say we’re interested in how getting a DUI effects whether someone has a DUI again. When someone is pulled over by the police, they take a breathalyzer test. If the driver’s Blood Alcohol Content is $\geq .08$ they get a DUI, if it’s less than that, the driver does not. This can be seen in the figure below where treatment (D) turns on once someone’s value of the running variable (BAC) moves above a threshold.

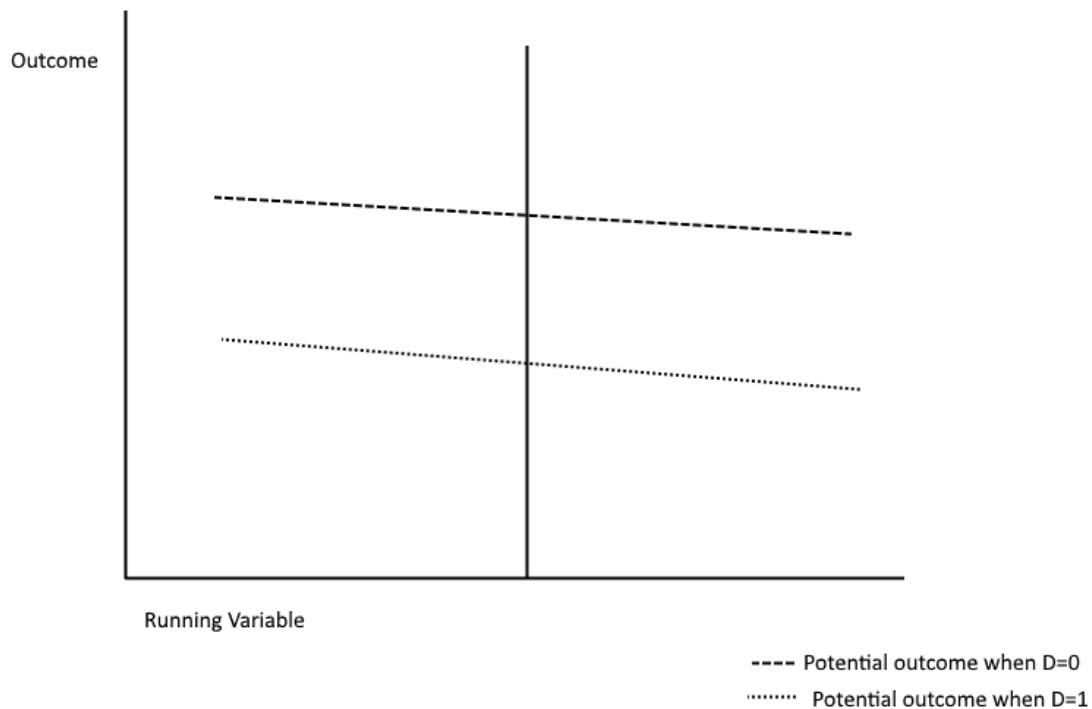


This figure shows a “sharp” regression discontinuity where treatment goes from 0 to 1 (no one is treated to everyone treated). This figure demonstrates that there is a (strong) first stage.

Will drivers between with a BAC of .079 be different than .080? Can they precisely manipulate the value of their BAC? Hence, in expectation, the only difference between someone with .079 and .080 would be that one got a DUI and one did not.

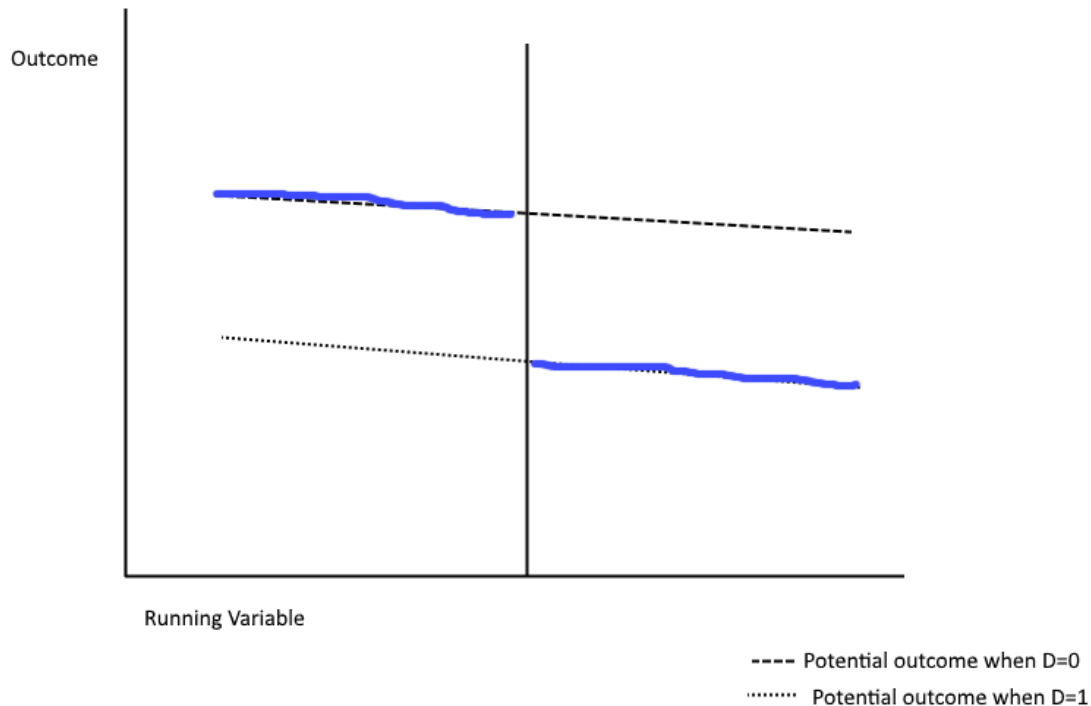
This intuition is formalized in a Regression Discontinuity (RD) design.

The typical formal assumption is not that it is random which side of the cutoff falls on. It is that potential outcomes are smooth through the the cutoff. You can see an example of this below.



We only see one of the dashed lines above. However the above picture illustrates the fundamental assumption of RD.

What happens when there is treatment? See the figure below. Treatment turns on the right side of the figure below and so people's potential outcomes move from D=0 to D=1.



When presenting RD results two figures/tables are key:

1. First Stage. Does treatment change at the threshold?
2. Reduced Form. Do outcomes change at the threshold?

The estimating equation for the first stage is

$$D_i = \theta \cdot 1(R_i \geq 0) + f(R_i) + \mathbf{X}\beta + \varepsilon_i \quad (1)$$

Where D_i is an indicator for treatment, R_i is the running variable recentered such that the cutoff is $=0$, $1(R_i > 0)$ is an indicator for the running variable being above the cutoff, $f()$ is a flexible function, \mathbf{X}_i is a vector of covariates, and ε_i is an idiosyncratic error term. θ reveals how moving just past the cutoff changes treatment.

The estimating equation for the reduced form is

$$Y_i = \gamma \cdot 1(R_i \geq 0) + f(R_i) + \mathbf{X}\alpha + \eta_i \quad (2)$$

Where Y_i is the outcome variable, R_i is the running variable recentered such that

the cutoff is $=0$, $1(R_i > 0)$ is an indicator for the running variable being above the cutoff, $f(\cdot)$ is a flexible function, X_i is a vector of covariates, and η_i is an idiosyncratic error term. γ reveals how moving just past the cutoff changes the outcome.

Fuzzy versus Sharp RD

You may see some discussion of fuzzy versus sharp RD. Sharp RD occurs when the *only* thing that determines treatment is the running variable. Hence estimating equation 1 would yield an estimate of $\theta = 1$

A fuzzy RD is when the jump in treatment is not 1—that is something else determines treatment beyond the value of the running variable. The example in MM is latin school admissions—a cutoff predicts admission but not perfectly. In that case θ will be <1 and will reveal the fraction of people whose treatment is changed as a result of the cutoff. θ is an estimate of the number of compliers.

You can use the intuition we developed in IV to instrument for treatment with the indicator $1(R_i \geq 0)$ for treatment. This means we only use the variation in treatment induced by the cutoff rule. We can recover a LATE from this IV for compliers. You might notice that many of the terms we use for RD are the same as they are for IV (Reduced Form, First Stage), this is because IV “sits on top” of RD.

Instrumental Variables are often used “on top of” identification strategies. For instance, IV is used in imperfect compliance in RCTs, fuzzy RDs (another form of imperfect compliance), and will be used in settings we will discuss in the future.

Assumptions

What assumption is needed?

The assumption is that there is a smooth relationship between the running variable and potential outcomes.

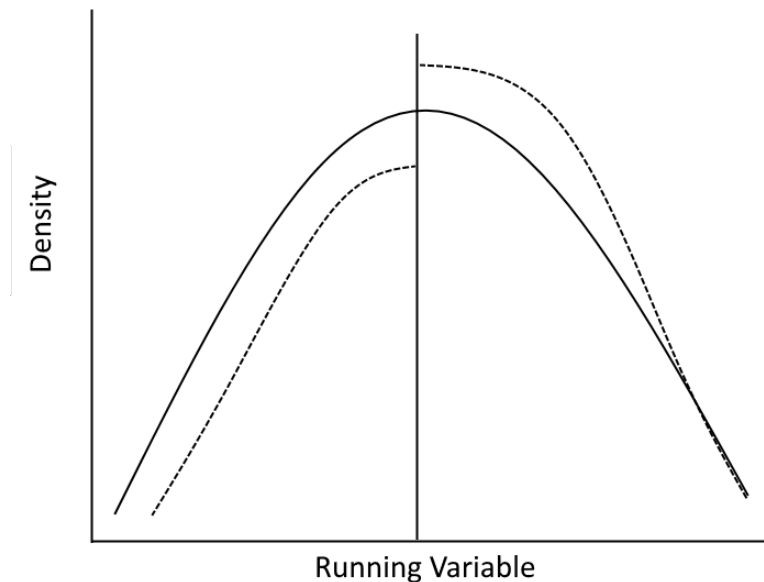
You can think of testable implications of this assumption:

1. People cannot precisely manipulate the value of their running variable. (Why? If they can, we have an issue with selection bias where people can decide to be treated)
2. Observations will look similar on side of the threshold versus the other on predetermined covariates (Why? Further evidence that potential outcomes are smooth)

How do you test this?

To test 1) you check and see if there is bunching in the density of the running variable. If people can precisely manipulate their running variable, you will see more on the treated side than the side without (assuming treatment is a good thing)

The following figure shows two different distributions. The first is the solid line and there is no evidence of manipulation. The second is the dashed line and there appears to be manipulation to get on the right side of the cutoff.



The formal test for this is McCrary (2008) Journal of Econometrics. Basically, it's a test to see if there is extra density on one side of the cutoff. If you have a discrete running variable Frandsen (2017) is the go-to test.

To test 2) you check to see if observed covariates jump discontinuously. You do this by replacing the outcome with predetermined covariates. You then can test jointly whether all of the discontinuities are jointly significant. You can also create an index where you regress the outcome on and see if that index changes. Is this the test you really want to run? (No) We would prefer to run this test on observed AND unobserved confounders. The intuition is very similar to an RCT, we check for balance on observables and need to assume balance on unobservables.

If you run these tests and find that they are failed, you cannot interpret RD as estimating a causal effect.

Global or Local?

There are two general approaches to estimating the relationship between the running variable and the outcome. You can use all available data and estimate the relationship between the running variable and the outcome. This is done typically using a higher order polynomial that varies on each side of the cutoff. This is the "global" approach

The current preferred method is to focus on a bandwidth, or window, on either side of the cutoff. That is, zoom in close to the cutoff and compare people just on one side or the other. This is the “local” approach and has the advantage matching our intuition of comparing people close to the cutoff. It does reduce sample size

In a local bandwidth, you can estimate the relationship $f(R_i)$ non parametrically using a local-linear approximation. This means that the relationship between the outcome and the running variable is linear, if we zoom in close enough. Often this means having a linear function of the running variable and allowing the slope to differ on each side of the cutoff.

You also have to pick the bandwidth somehow. In general, you are trading off the additional power from a bigger bandwidth with the potential bias of using data far away from the cutoff to estimate $f(R_i)$. Econometricians fight about this, but there are several popular data-driven approaches to making this tradeoff. Two worth mentioning are Imbens Kalyanaraman (2012) (IK) Calonico Cattaneo and Titiunik (CCT) approach. See the following site for documentation for CCT:

<https://sites.google.com/site/rdpackages/rdrobust>

If your estimate depends critically on the bandwidth selected, the estimate will be viewed as not very robust. Hence, people usually show estimates using the IK or CCT bandwidth. They then show it doesn’t change as the bandwidth grows or shrinks.

A good reference for best practice is Imbens Lemieux (2008) Journal of Econometrics