Lab #17 - Regression Discontinuity Basics

Econ 224

November 6th, 2018

Introduction

Today we'll go through some nuts-and-bolts of sharp regression discontinuity. I'll begin by providing some details of the RD approach that go beyond the explanation in MM. We'll then look at how to carry out RD analysis in R.

"Sharp" Regression Discontinuity

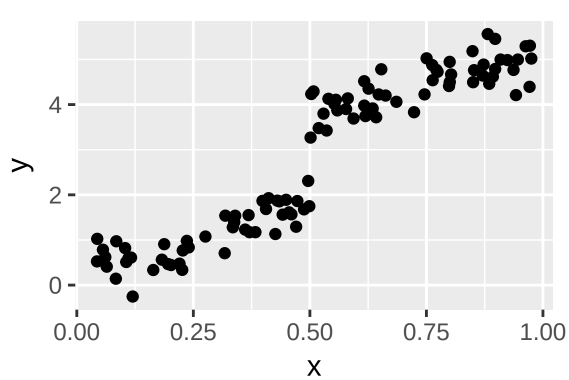
Suppose we are interested in learning the causal effect of a binary treatment D on an outcome Y. In some special settings, whether or not a person is treated is a solely determined by a special covariate x, called the *running variable*

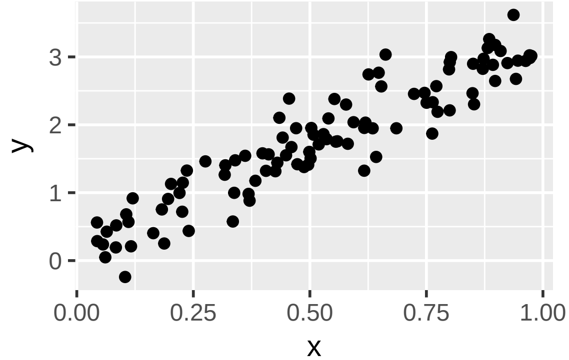
$$D_i = \begin{cases} 0 & \text{if } x_i < c \\ 1 & \text{if } x_i \geq c \end{cases}$$

The preceding expression says that D is a *deterministic function* of x: everyone who has $x \geq c$ is treated, and no one who has $x < c$ is treated. This setting is called a *sharp regression discontinuity design* and it provides us with a powerful tool for causal inference. We'll distinguish this from another kind of regression discontinuity setup called a *fuzzy regression discontinuity design* below. I'll use the shorthand RD to refer to regression discontinuity in these notes.

When we previously used regression to carry out causal inference, the idea was to compare two groups of people who had been *matched* using a set of covariates $\mathbf{x}$. One group was treated and the other was not, but both groups had exactly the same values of $\mathbf{x}$. Under the assumption that treatment is "as good as randomly assigned" after conditioning on $\mathbf{x}$, we could learn the causal effect of treatment by comparing the mean outcomes of the two groups.

Sharp RD is very different since there is no way to carry out matching using the running variable x. This is because everyone who has $x < c$ is untreated while everyone who has $x \geq c$ is treated. Instead of matching people who have the same covariate values, sharp RD *extrapolates* by comparing people with *different* covariate values. The basic idea is very simple: we compare people whose x is close to but slightly *below* the cutoff c to people whose x is close to but slightly *above* the cutoff. Both D and x could affect Y, but since D abruptly switches from 0 to 1 at c, a causal relationship between D and Y should show up as a "jump" in the relationship between x and Y at c. For example, in the left panel there is clearly a jump at $c = 0.5$ and in the right panel there is not:





If the threshold c equals 0.5, then the left figure suggests that D has a substantial causal effect on Y: when D switches from zero to one as x crosses the threshold, the mean of Y jumps from around 2 to around 4. In contrast, the right panel doesn't show evidence of a causal effect of D on Y: when D switches from zero to one, there is no discernible change in the mean of Y. Now we'll be a little more precise about this intuition by thinking about exactly how x and Y are related. The key will be to create a link to potential outcomes, since this is our main tool for thinking about causality. Let's stick with the same x-axis as in the preceding figures: imagine that the running variable x is between zero and one, and that the cutoff c equals 0.5. This means that anyone with $x < 0.5$ is untreated while anyone with $x \geq 0.5$ is treated.

To begin, suppose we had a data for a large random sample of people who all had $x = 0.3$. If we took the mean Y for these people we would get an unbiased estimate of $E[Y_i|x_i = 0.3]$. The crucial point about sharp RD is that treatment is *completely determined* by x. This means that there is *no selection bias* since individuals are not free to choose their treatment status. Since everyone with $x_i = 0.3$ is untreated, we have $E[Y_i|x_i = 0.3] = E[Y_{0i}|x_i = 0.3]$. Note that this is *not* the same thing as $E[Y_{0i}]$ since the running variable x could have a direct effect on Y. In words, a person's potential outcome when untreated could depend on her value of x. For example, $E[Y_{0i}|x_i = 0.3]$ may not equal $E[Y_{0i}|x_i = 0.4]$ even though neither someone with x equal to 0.3 nor someone with x equal to 0.4 is treated. But this is fine, since we know how to use *predictive* modeling tools to estimate a conditional mean function. Here is the key point: since there is no selection into treatment for people with $x < c$, we can use *predictive regression* to estimate $E[Y_i|x_i]$ and this will give us an estimate of $E[Y_{0,i}|x]$ for any x below the threshold.

What about when x is above the threshold? Consider for example, a large group of people with $x = 0.6$ and suppose as above that $c = 0.5$. All of these people are treated since their x exceeds the threshold. If we take the average Y for this group of people, we will obtain an unbiased estimator of $E[Y_i|x_i = 0.6]$. But since this group of people could not possibly select *out* of treatment, there is once again no selection bias and hence $E[Y_i|x_i = 0.6] = E[Y_{1i}|x_i = 0.6]$. This is not the same thing as $E[Y_{1i}]$ since x could affect Y directly. But, again, this doesn't present a problem: since there is no selection out of treatment for people with $x \geq c$, we can use *predictive regression* to estimate $E[Y_i|x_i]$ and this will give us an estimate of $E[Y_{1i}|x_i]$. To summarize the reasoning from this and the preceding paragraph,

$$E[Y_i|x_i] = \begin{cases} E[Y_{0i}|x_i], & \text{if } x_i < c \\ E[Y_{1i}|x_i], & \text{if } x_i \geq c \end{cases}$$

Again, this relationship holds because individuals are *not* free to choose their treatment: everyone with $x \geq c$ is treated and no one with $x < c$ is treated.

There is a key distinction you need to bear in mind: $E[Y_i|x_i]$ includes the effect of *both* D and x while $E[Y_{0i}|x_i]$ and $E[Y_{1i}|x_i]$ hold D *fixed*. The function $E[Y_i|x_i]$ answers the question "what value should we predict for Y for someone who has a covariate value of x_i?" In contrast, the function $E[Y_{0i}|x_i]$ answers the question "what would be the average outcome for a person with covariate value x_i if I randomly assigned her $D = 0$?" Similarly, $E[Y_{1i}|x_i]$ answers the question "what would be the average outcome for a person with covariate value x_i if I randomly assigned her $D = 1$?" If we knew $E[Y_{0i}|x_i]$ and $E[Y_{1i}|x_i]$ for all values of x, then by taking the difference, we could learn how the ATE *varies* across people with different values of x:

$$\text{ATE}(x) = E[Y_{1i}|x_i = x] - E[Y_{0i}|x_i = x] = E[Y_{1i} - Y_{0i}|x_i = x]$$

using the linearity of expectation. The idea of estimating an ATE as a function of some covariate x is called "heterogeneous treatment effects." For example, a treatment may be more effective for younger people than older people. If we had experimental data, we could estimate ATE(x). But in the RD setting we only have *observational data*. Crucially, we never observe $E[Y_{0i}|x_i]$ for anyone with $x_i \geq c$, and we never observe $E[Y_{1i}|x_i]$ for anyone with $x_i < c$.

So how can we proceed? In RD, the key assumption is that $E[Y_{0i}|x_i]$ and $E[Y_{1i}|x_i]$ are *continuous functions* of x. In other words, while we allow for the possibility that people with different values of x will have different potential outcomes, we assume that people with values of x that are *very similar* have will have potential outcomes that are *nearly equal*. In particular, we assume that $\lim_{\Delta \to 0} E[Y_{1i}|x_i = c + \Delta] = E[Y_{1i}|x_i = c]$ and similarly that $\lim_{\Delta \to 0} E[Y_{0i}|x_i = c - \Delta] = E[Y_{0i}|x_i = c]$. But since $E[Y_i|x_i]$ equals $E[Y_{0i}|x_i]$ for x above the threshold and $E[Y_{1i}|x_i]$ for x below the threshold, this implies that

$$\lim_{\Delta \to 0} E[Y_i|x_i = c + \Delta] - E[Y_i|x_i = c - \Delta] = E[Y_{1i} - Y_{0i}|x_i = c] = \text{ATE}(c)$$

So by using predictive regression to estimate $E[Y_i|x_i]$ for x_i *just below* c and comparing it to an estimate of $E[Y_i|x_i]$ for x_i *just above* c, RD allows us to learn the average treatment effect for individuals with $x = c$.

RD Analysis in R

Exercise A

In this exercise you will derive Equation 4.3 from MM – albeit with slightly different notation – by drawing a connection with Exercise A above. Suppose that we want to predict Y using X while allowing a difference slope and intercept for $X > c$ compared to $X \leq c$, in particular:

$$Y_i = \begin{cases} \alpha_0 + \alpha_1 X_i + \epsilon, & \text{for } X_i \leq c \\ \beta_0 + \beta_1 X_i + \epsilon, & \text{for } X_i \leq c \end{cases}$$

1. In terms of $\alpha_0, \alpha_1, \beta_0, \beta_1$, and c, what is the the value of the RD causal effect?
2. Show how to write the two "separate" linear regressions from above as a single "joint" linear regression of the form
$$Y_i = \gamma_0 + \gamma_1 X_i + \gamma_2 D_i + \gamma_3 D_i X_i + \epsilon$$
where D_i is a dummy variable that equals one if $X_i > c$. What is the relationship between the coefficients $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ of the "joint" regression, and the coefficients (α_0, α_1) and (β_0, β_1) of the "separate" regressions?
3. By combining your answers to parts 1 and 2, show how to write the RD causal effect in terms of the parameters of the "joint" regression.
4. Simulate 500 observations from the regression model from part 1. Generate X from a Uniform$(0, 1)$ distribution and and the errors ϵ from a $N(0, \sigma^2)$ distribution. Set the value of the cutoff c equal to 0.5. Choose values of the parameters in your simulation that imply visibly different slopes and intercepts to the left and right of the cutoff c and an error variance that is small enough that these differences are visible in the data.
5. Use the simulated data from part 3 to fit the regression model from part 1. Plot the raw data along with the estimated regression lines to the left and right of the cutoff.
6. If we interpret the model in this exercise as an RD design, which coefficient gives us the estimated causal effect? What is the associated standard error?

1. If we let $f(x) = \alpha_0 + \alpha_1 x$ and $h(x) = \beta_0 + \beta_1 x$, the RD causal effect is $h(c) - f(c) = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)c$.
2. Define $D_i = \mathbf{1}\{X_i > c\}$, which implies that $(1 - D_i) = \mathbf{1}\{X_i \leq c\}$. Using this notation, we can write

$$
\begin{aligned}
Y_i &= (1 - D_i) \times (\alpha_0 + \alpha_1 X_i) + D_i \times (\beta_0 + \beta_1 X_i) + \epsilon_i \\
&= (\alpha_0 + \alpha_1 X_i) + D_i \times [(\beta_0 + \beta_1 X_i) - (\alpha_0 + \alpha_1 X_i)] + \epsilon_i \\
&= \alpha_0 + \alpha_1 X_i + (\beta_0 - \alpha_0) D_i + (\beta_1 - \beta_0) D_i X_i + \epsilon_i
\end{aligned}
$$

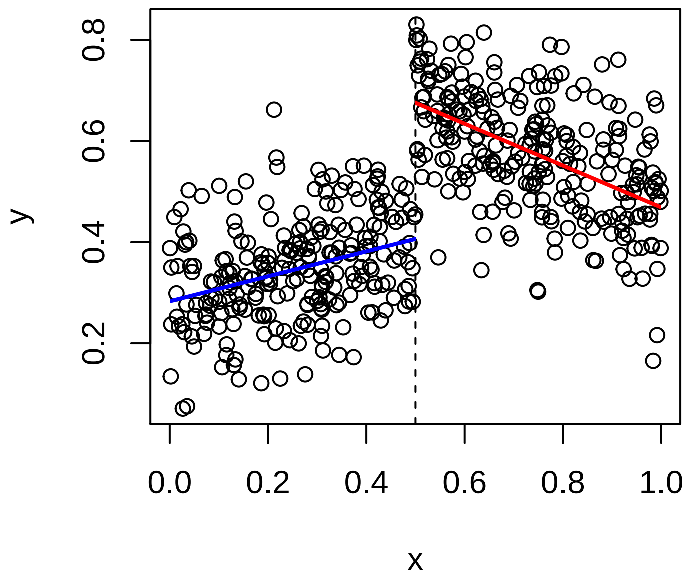
Thus, we see that $\gamma_0 = \alpha_0, \gamma_1 = \alpha_1, \gamma_2 = (\beta_0 - \alpha_0)$, and $\gamma_3 = (\beta_1 - \alpha_1)$.

3. The RD causal effect is $(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)c = \gamma_2 + \gamma_3 c$.

```r
# Parameters of the "separate" regressions
a0 <- 0.3
a1 <- 0.2
b0 <- 0.8
b1 <- -0.3

# Implied parameters of the "joint" regression
g0 <- a0
g1 <- a1
g2 <- b0 - a0
g3 <- b1 - a1

# Simulation draws
set.seed(1234)
n <- 500
x <- runif(n)
cutoff <- 0.5
D <- 1 * (x > cutoff)
epsilon <- rnorm(n, sd = 0.1)
y <- g0 + g1 * x + g2 * D + g3 * D * x + epsilon

# Fit linear regression model
rd <- lm(y ~ x + D + x:D)
summary(rd)
```

```
Call:
lm(formula = y ~ x + D + x:D)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.31021 -0.06761  0.00252  0.06190  0.32662 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.28308    0.01331  21.262  < 2e-16 ***
x            0.24660    0.04579   5.386 1.12e-07 ***
D            0.59943    0.03382  17.725  < 2e-16 ***
x:D         -0.66036    0.06152 -10.734  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09925 on 496 degrees of freedom
Multiple R-squared:  0.6171,	Adjusted R-squared:  0.6148 
F-statistic: 266.4 on 3 and 496 DF,  p-value: < 2.2e-16
```

```r
# Estimated lines before and after cutoff
a0hat <- coef(rd)[1]
a1hat <- coef(rd)[2]
b0hat <- coef(rd)[3] + coef(rd)[1]
b1hat <- coef(rd)[4] + coef(rd)[2]

# Plot raw data
plot(x, y)
abline(v = cutoff, lty = 2)

# Add regression line before the cutoff
clip(min(x), cutoff, min(y), max(y))
abline(a0hat, a1hat, col = 'blue', lwd = 2)
clip(cutoff, max(x), min(y), max(y))
abline(b0hat, b1hat, col = 'red', lwd = 2)
```



Exercise C

Create an example where non-linearity masquerades as a discontinuity. Then have them repeat 2 but with a *quadratic* regression. Maybe talk about the non-parametric version?