Executive Summary for Lecture Set #3, part 2

The first part of this summary talked about the basic intuition for difference-in-differences estimates, but it never discussed how (or why) regressions were involved.

Lesson #1: Why do we use regressions in this context? Because they enable us to perform hypothesis tests. In fact, we'll need to test two major hypotheses in this setting:

(1) Are the treatment and control group **"similar" in the "pre" period**? Remember that we need the control group to serve as a counterfactual for the treatment group. But this is only possible if the two groups are similar to each other before we enact any changes to the treatment group.
(2) Did the change actually have an impact on our Y? Specifically: did the treatment group exhibit a change in Y that was **significantly different** than the change in Y exhibited by the control group?

But **how** do we test these ideas?

Lesson #2: The method (or "how") we test the hypotheses in lesson #1 is with a regression that looks quite strange. Suppose that we have a dependent variable that we care about (we'll call it Y), and two dummy variables:

(i) "Post" is equal to one for any observation (from either the treatment or the control group) collected in the "post" period, and zero if it was collected in the "pre" period.
(ii) "Treatment" is equal to one for any observation collected from our treatment group at any time, and zero if the observation was collected from our control group at any time.

Now run the following regression:

$$Y = b_0 + b_1(Post) + b_2(Treatment) + b_3(Treatment * Post)$$

I know it looks like a mess, and the intuition behind this regression seems impenetrable, but we'll go through this very slowly and carefully in class. Here is the upshot that we'll rely upon to test the ideas in Lesson #1:

(1) The coefficient b_2 measures the average difference in Y, in the "pre" period, between the treatment and control groups. In plain language: it tells us if the two groups have a similar average Y in the "pre" period. And we can formally test this with the null hypothesis that this coefficient is equal to zero – because if we reject this hypothesis, then we can reject that the two groups are the same in the "pre" period (which means that the control group isn't a good counterfactual in our model). By contrast, if we fail to reject this hypothesis, it means that we can't reject that the two groups are basically the same in the "pre" period. So, we hope that we fail to reject this hypothesis (and we pay very close attention to the result).
(2) The coefficient b_3 measures the difference-in-differences effect! And again, we can formally test if this coefficient is zero. If we reject that it's zero, then we reject that our change had no impact on the treatment group, relative to the control group – there was a meaningful causal effect in this case. But if we fail to reject that it's zero, then we can't reject the idea that our change had no impact on the treatment group, relative to the control group – the change yielded no meaningful, causal effects.