Recitation 2: Merging Counterfactuals and Regressions

Colin Gray

14.03/14.003, Fall 2016

Outline

Outline

So what's the deal with all these subscripts?

- Subscript i denotes a unit of observation (individual, household, zip code, state, etc)
- Define Y_{i1} as the outcome Y *if* unit i received "treatment", and Y_{i0} as the outcome Y *if* unit i did *not* receive treatment.
- This has nothing to do with whether unit i actually did receive the treatment! It's a *counterfactual*.

What we can see, and what we can't see.

- It turns out that expectations and averages are really mathematically convenient. We sometimes use other properties of the data (e.g. medians).
- What does $E[Y_{i0}|X_i = 1]$ mean?
- What about $E[Y_{i0}|X_i = 0]$, $E[Y_{i1}|X_i = 0]$, or $E[Y_{i1}|X_i = 1]$? Which can we estimate from the data with no further assumptions?

The problem with "naive" observation.

- We often want to know $E[Y_{i1} - Y_{i0}|X_i = 1]$. Why?
- One natural first try is to estimate expectations $E[]$ by calculating averages, and then plugging them into the following expression:

$$\begin{aligned}
&E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0] \\
&= \big(E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 1]\big) \\
&+ \big(E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]\big)
\end{aligned}$$

How to get around bias.

- We can *randomize*. What does that do?

$$E[Y_{i0}|X_i = 1] = E[Y_{i0}|X_i = 0]$$
$$\Rightarrow E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0] = E[Y_{i1} - Y_{i0}|X_i = 1]$$

- When we can't randomize, we look for a *control group* in which we believe that $E[Y_{i0}|X_i = 1] = E[Y_{i0}|X_i = 0]$.

Outline

When we can't observe everything about a group.

- It's often infeasible to randomize and unreasonable to think that our control group and treatment group are really that similar, so we probably don't believe that

$$E[Y_{i0}|X_i = 0] = E[Y_{i1}|X_i = 0]$$

- What if we instead assume that our treatment and control groups don't have the same $E[Y_{i0}]$, but do follow the same counterfactual *trends*?
- Now we add the time dimension, so Y_{i0} turns into Y_{it0}. We express our new assumption as:

$$E[Y_{i10}|X_i = 1] - E[Y_{i00}|X_i = 1] = E[Y_{i10}|X_i = 0] - E[Y_{i00}|X_i = 0]$$

DD to the rescue!

- Now we want the effect of the treatment in period 1: $E[Y_{i11}|X_i = 1] - E[Y_{i10}|X_i = 1]$
- Check this out...

$$
\begin{aligned}
& E[Y_{i11}|X_i = 1] - E[Y_{i00}|X_i = 1] \; - \\
& E[Y_{i10}|X_i = 0] - E[Y_{i00}|X_i = 0] \\
= \;\; & E[Y_{i11}|X_i = 1] - E[Y_{i00}|X_i = 1] \; - \\
& E[Y_{i10}|X_i = 1] - E[Y_{i00}|X_i = 1] \\
= \; & E[Y_{i11}|X_i = 1] - E[Y_{i10}|X_i = 1]
\end{aligned}
$$

Woah! What just happened?!?!

- We call this a *difference-in-differences (DD)* estimator.
- To actually estimate this, we can replace every $E[]$ we see with a sample average. However, a more flexible method is to use a *regression*.

Outline

Operationalizing counterfactuals with a regression

- We just went over how a clever choice of which expectations we estimate allows us to uncover treatment effects under reasonably weak assumptions.
- Now I'll show you that we shouldn't actually be calculating a bunch of expectations in turn... Regressions can do the same thing with some *very* useful perks!

Let's consider an example.

- A couple years ago, there was some hearty debate over whether the Affordable Care Act (ACA) would lower the wages of employees at small businesses, since small firms now needed to purchase health insurance for workers while large firms (usually) already offered health insurance.
- Suppose we wanted to run a regression of employees at small firms (100-500 employees) vs. employees at large firms (500+ employees) in 2014:

$$Earn_{i1} = \beta_0 + \beta_1 Small_{i1} + e_{i1}$$

- Let's work through this example using data I downloaded from the Current Population Survey

What a single-variable regression is doing.

- Let Y stand for earnings and X stand for treatment (being at a small firm). Let's drop t until we get to the DD again, since we're considering only the post-ACA world (2014, otherwise known as $t = 1$).
- Regressions mechanically give us an estimate of $\beta_1 = E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0]$. If you need proof, check out what happens when I compare a t-test (which explicitly compares the difference in these sample means) against a regression.
- A t-test...

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.
0	63,033	57398.29	244.2421
1	17,913	51914.83	463.4656
combined	80,946	56184.82	216.2307
diff		5483.459	520.5349

earnings	Coef.	Std. Err.	t

What's the problem here?

- So the β_1 in our regression is giving us $E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0]$. Let's imagine why that might *not* be the treatment effect we are looking for...
- Recall our term for bias:

$$E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]$$

¡3-¿ What if we think that college-educated workers often work at large firms, and that college-educated workers make more money regardless of the regulations surrounding firms of different sizes?

What it means to "control" for a variable.

- Unlike comparing sample means directly, a regression also gives us a natural way to take out the effect of these *omitted variables* and therefore reduce the bias in our estimate of the treatment effect.
- Suppose we estimate the effect of college on the likelihood of working at a small firm (X):

smallfirm	Coef.	Std. Err.	t
college	-.0430479	.0029512	-14.59
_cons	.2394288	.0019154	125.00

- Then we took the movements in X that seem unrelated to college, and regress *that* variable on earnings:

earnings	Coef.	Std. Err.	t
smallfirm_~1	-3308.268	521.4461	-6.34
_cons	56184.82	216.1783	259.90

What it means to “control” for a variable (2).

- This measures the relationship between working at a small firm and earnings without the effects of education. To do this quickly, we can actually put all of these variables right into the same regression:

earnings	Coef.	Std. Err.	t
smallfirm	-3308.269	499.3756	-6.62
college	35716.2	419.8433	85.07
_cons	41872.14	297.2388	140.87

- Now we understand how to make direct comparisons of means with a regression, and how regression allows us to easily include omitted variables to reduce bias!

Outline

The Grand Finale

- Of course, you probably think that workers at small and large firms have different earnings for a whole host of reasons, not just regulations and not just college educations. We can't just randomize where people apply for jobs, and we can't see all these differences between our groups. So...
- TIME FOR A DD!

The Grand Finale

- Consider the following regression. Now the observation is at the person-time (so (i, t)) level:

$$Earn_{it} = \beta_0 + \beta_1 Small_i + \beta_2 Post2014_{it} + \beta_3 Small_i Post2014_{it} + u_{it}$$

- Or in our alternative notation:

$$Y_{it} = \beta_0 + \beta_1 X_i + \beta_2 1[t = 1]_{it} + \beta_3 X_i 1[t = 1]_{it} + u_{it}$$

- FYI, the notation $1[condition]$ stands for a variable that equals 1 when the *condition* is true and 0 otherwise. Sometimes we call this a *dummy variable*.

The Grand Finale

- For now, just trust me that $E[u_{it}|X_i] = 0$ when we fit this kind of regression (but not necessarily all kinds of regressions!).
- Using our previous notation, $E[Y_{i11}|X_i = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
- $E[Y_{i00}|X_i = 1] = \beta_0 + \beta_1$
- $E[Y_{i10}|X_i = 0] = \beta_0 + \beta_2$
- $E[Y_{i00}|X_i = 0] = \beta_0$
- Therefore our DD estimator is

$$
\begin{aligned}
&E[Y_{i11}|X_i = 1] - E[Y_{i00}|X_i = 1] \; - \\
&E[Y_{i10}|X_i = 0] - E[Y_{i00}|X_i = 0] \\
&= \beta_3
\end{aligned}
$$

The Grand Finale

- So for our purposes, a regression does all computational work for us! Plus, it lets us do a whole bunch of other cool things that we'll continue to see throughout the semester.
- When we do the DD in our example, we see no significant effects on wages for small business workers after the ACA (although this design is far from perfect).

earnings	Coef.	Std. Err.	t
smallfirm	-3169.461	305.6749	-10.37
aca	2547.216	265.8766	9.58
aca_small	$\boxed{-141.663 \quad 565.0023}$		-0.25
college	35346.34	217.3792	162.60
age	587.5353	9.885164	59.44
_cons	13975.27	462.7783	30.20

A visual interpretation