Recitation 9 - Quasi-Experiments

Nathaniel Mark

November 27, 2017

1 Example

Say we want to analyze the effect of being in medicare on health care costs. That is, do people in medicare spend more on health care than those in private health insurance?
Simplest model:

$$Cost_i = \beta_0 + \beta_1 Medicare_i + \epsilon_i$$

Where $Medicare_i$ is =1 if person i is in medicare and 0 otherwise. We want to find an estimate of the causal parameter β_1. What are the problems we have if we estimate this model above?
Many. Biggest is OVB– being in medicare is correlated with a lot of things that cause higher health care costs, including age and disability status.

How do we fix this problem? We have already learned some methods, such as IV, fixed effects, adding controls, etc. This week, we discuss special ways you can do so when you have quasi-experimental data.

2 Intro

What is a Quasi-Experiment?
Quasi-experimental data are data where some entities are treated because of some outside random shock instead of the fact that they choose the treatment.
This is not the same as being randomly chosen (as in a randomized controlled experiment), because those who are treated are often similar in some way. This similarity is what differentiates "as if" random and truly random.

3 Topic One: Difference Estimators and Diff-in-Diff Estimators

Difference estimators are an estimation of the simple, 1 regressor OLS model.

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\beta_0 + \beta_1 = E[Y_i | X_i = 1]$$

$$\beta_0 = E[Y_i|X_i = 0]$$

So, if our OLS conditions hold, then

$$\hat{\beta}_1 \to E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$$

which is our average treatment effect. Recall that another way to estimate this is simply by estimating each conditional mean by its simply average estimator:

$$\hat{\beta}_1 = \bar{Y}^{treated} - \bar{Y}^{untreated}$$

But, omitted variable bias is a problem here. How do we fix it? One way is to include controls. That would give us the **difference estimator with additional regressors**. Another way is to use fixed effects methods. If we use time fixed effects under the condition where T=2 and a dummy for if treated (I call it T, your book calls it G), we call this a Difference-in-Difference Estimator:

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 T_i + \gamma_t + u_{i,t}$$

We estimate using the dummy variable method:

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 T_i + \gamma D_t + u_{i,t}$$

$$X_{i,t} = T_i D_t$$

where $D_t = 1$ when t=2 and 0 otherwise. Now,

$$\beta_0 + \beta_1 + \beta_2 + \gamma = E[Y_{i,t}|treated, t = 2]$$

$$\beta_0 + \gamma = E[Y_{i,t}|untreated, t = 2]$$

$$\beta_0 + \beta_2 = E[Y_{i,t}|treated, t = 1]$$

$$\beta_0 = E[Y_{i,t}|untreated, t = 1]$$

Note that with these estimators the treatment is only possible in the second time period, so the treated population is untreated when t=1. In our example, a usage of this type of estimator would be if a certain state expanded medicare.
It is easy to see that if all our OLS assumptions apply, then

$$\hat{\beta_1} \to \beta_1 = ((\beta_0 + \beta_1 + \beta_2 + \gamma) - ((\beta_0 + \beta_2)) - ((\beta_0 + \gamma) - (\beta_0))$$

$$= (E[Y_{i,t}|treated, t = 2] - E[Y_{i,t}|treated, t = 1]) - (E[Y_{i,t}|untreated, t = 2] - E[Y_{i,t}|untreated, t = 1])$$

It is clear then that another way to estimate this is simply by estimating each conditional mean by its simple average estimator:

$$\hat{\beta}_1 = (\bar{Y}^{treated,t=2} - \bar{Y}^{treated,t=1}) - (\bar{Y}^{untreated,t=2} - \bar{Y}^{untreated,t=1})$$

But we still may have OVB! How would we fix it? Now that we are in the regression format, simply conduct normal OLS techniques to the above regression: Add fixed effects to remove OVB from variables that change over entity but not time, estimate using instrumental variables, add control or omitted variables as new covariates...

4 Topic 2: Regression Discontinuity Designs

SHARP RDD:

We think our true world looks like:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

There is some other variable, W_i and a "threshold" such that
$P(X_i = 1) = 1$ IF $W_i \geq$ threshold and
$P(X_i = 1) = 0$ IF $W_i \leq$ threshold.

Then β_1 can be estimated by the average of Y_is for which W_i is just above the threshold and the average of Y_is for which W_i is just below the threshold.

FUZZY RDD:

The only difference between fuzzy and sharp RDD is that the probabilities no longer have to be exactly 1 or 0 anymore. All we need is:
There is some other variable, W_i and a "threshold" such that
$P(X_i = 1) = f(W_i)$ IF $W_i \geq$ threshold and
$P(X_i = 1) = g(W_i)$ IF $W_i \leq$ threshold
AND $g(threshold) < f(threshold)$.

Interpretation of RDD:
Note that $\hat{\beta}_1^{RDD}$ estimates the effect on the expected increase in Y_i from being treated *for those close to the threshold.*

Using our example:
Y_i is Cost or person i.
X_i is the Medicare dummy.
Age_i is W_i.
Threshold is 65.
This problem in (almost) sharp, but there are likely some exceptions.

5 Topic 3: Heterogeneous Effects and Potential Outcomes Framework

In all of our previous models, we wrote it out as if the effect of covariates are the same for all members of the population. Of course, this is not true. The potential outcomes framework is a way to analyze the fact that treatments can have different effects on different people.
The framework:
X_i is the treatment dummy (=1 if treated)
Y_i is the *observed* outcome.
$Y_i(0)$ is the outcome that would be observed if i is untreated.
$Y_i(1)$ is the outcome that would be observed if i is treated.

We can write the model in two different, equivalent ways:
OLS: $Y_i = \beta_{0,i} + \beta_{1,i}X_i + \epsilon_i$
Definitional: $Y_i = X_iY_i(1) + (1 - X_i)Y_i(0)$

Notes:
1) Both $Y_i(0)$ and $Y_i(1)$ are never observed for the same individual, but they both exist.
2) X_i is not independent of Y_i as long as $\beta_{1,i} \neq 0$.
3) X_i *may* not be independent of $Y_i(0)$ and $Y_i(1)$. Think what if those with high costs choose to be treated?
3) If X_i is independent of $Y_i(0)$ and $Y_i(1)$, this buys us a lot in estimation power, as $E[Y_i(1)|X_i = 1] = E[Y_i(1)]$ and more generally, $P[Y_i(1) = y|X_i = 1] = P[Y_i(1) = y]$. It is feasible to estimate the left hand side, so this allows us to estimate the (usually infeasible) left hand side.

Model Results in the Potential outcomes framework:

OLS:
Assumptions:
1) X_i is randomly assigned (X_i is independent of $Y_i(0)$ and $Y_i(1)$)
2) LSAs 2-4 (LSA1 is implied by the above)
THEN
$\hat{\beta}_{1,OLS} \rightarrow ATE$
Where $ATE = E[\beta_{1,i}]$

IV:
Assumptions:
1) X_i *may* not be independent of $Y_i(0)$ and $Y_i(1)$.
2) IV assumptions, ie valid instruments...
THEN
$\hat{\beta}_{1,IV} \rightarrow LATE$
Where $LATE = \frac{E[\beta_{1,i}\pi_{1,i}]}{E[\pi_{1,i}]}]$ ($\pi_{1,i}$ is the effect of the instrument on the treatment).

LATE=ATE IF

1) Treatment has the same effect on the outcome for everyone ($\beta_{i,1} = \beta_1 \forall i$).

2) Instrument has the same effect on the Treatment for everyone ($\pi_{i,1} = \pi_1 \forall i$).

OR

3) $Corr(\pi_{i,1}, \beta_{i,1}) = 0$