ØAMET4100 · Spring 2019
Lecture Note 8

Instructor: Fenella Carpena

March 14, 2019

This lecture note provides a review of experiments and quasi-experiments (Stock & Watson, Chapter 13). This lecture note is not intended to be a comprehensive review of lecture or the textbook, since there is a lot more material than we have time to cover. However, I have tried to focus on the concepts which I believe are necessary to be successful in our class.

1 Experiments

1.1 Overview

Experiments occur when a treatment (or policy) is randomly assigned to investigate its causal effect. In an experiment, some entities (e.g., individuals) are assigned to a "treatment group" which receives the treatment, while the rest are assigned to a "control group" which does not receive the treatment.

Why might we care about experiments in an econometric course? I can think of several reasons not only for the purpose of academic research but also for the interests of businesses or the private sector.

- Randomized experiments are the "gold standard" for estimating causal effects. Experiments can overcome threats to internal validity of observations studies, and they provide a conceptual benchmark to judge estimates of causal effects made with observational data. Note that while experiments can be a powerful tool for estimating causal effects, they also have their own threat to internal and external validity.
- Experiments can be very influential. For example, experiments in the medical fields (i.e., medical trials) are often discussed in the news. Hence, it is important to understand the strengths and weaknesses of experiments as well as the threats to the validity of an experiment.
- Experiments help us understand quasi-experiments (or natural experiments), which occur when the treatment wasn't explicitly randomized but because of external circumstances is "as-if" it was randomized.
- Many companies are also using experiments to improve their business. On example is A/B testing: one version of a company's website ("Version A") might have red "Buy Now" button, while a second version ("Version B") might have a blue "Buy Now" button. Experiments can be carried out to investigate which version of the website leads to higher sales.

Example. Suppose we want to investigate the following question: Is a new drug effective in lowering cholesterol? To conduct the experiment, we might implement the following steps.

Step 1 Get study participants (this would be our sample)

Step 2 Get "baseline" data of the participants (e.g., age, gender, cholesterol level). The term "baseline" here means we are measuring these variables *before* the treatment begins.

Step 3 Randomly assign participants to the treatment group or the control group. The treatment group will get the drug, while the control group will get a placebo.

Step 4 Participants take the drug or the placebo (depending on which group they were assigned to).

Step 5 Get "endline" data. In our case, it would be important to measure cholesterol levels at endline because of our research question. Note that the term "endline" here means we are measuring the variable *after* the treatment has been completed.

1.2 Analysis

How do we analyze experimental data? We can use the differences estimator or the difference estimator with additional regressors. We now discuss each of these approaches.

1.2.1 Differences Estimator

Regress the outcome Y on a treatment dummy. Specifically, estimate

$$Y_i = \beta_0 + \beta_1 T_i + u_i \tag{1}$$

where in our example described above on cholesterol, Y_i is person i's cholesterol level measured at endline, and T_i is a dummy variable equal to one of the person was assigned to the control group. How do we interpret the coefficients here?

- $\widehat{\beta}_1$ is the unbiased estimator for the average causal effect of the drug on cholesterol levels. We are getting an unbiased estimate because the treatment T_i was randomly assigned, which means that $E(u_i|T_i) = 0$ (that is, the first least squares assumption holds.
- $\widehat{\beta}_0$ is the sample average Y_i of the control group, which we can write as $\overline{Y}^C$.
- $\widehat{\beta}_0 + \widehat{\beta}_1$ is the sample average Y_i of the treatment group, which we can write as $\overline{Y}^T$.

Using the above, we can see that $\widehat{\beta}_1 = (\widehat{\beta}_0 + \widehat{\beta}_1) - \widehat{\beta}_0 = \overline{Y}^T - \overline{Y}^C$. Thus, taking the difference in means $\overline{Y}^T - \overline{Y}^C$ gives us the same estimate as β_1 in the regression. Note that $\overline{Y}^C$ tells us what the average outcome would among those people in the treatment group had they not been treated; without the control group, we would not have an idea about what would have happened to the treatment group because there are no parallel universes.

1.2.2 Difference Estimator with Additional Regressors

We could augment equation (1) by adding control variables (denoted below as W's), which are variables measured at baseline or those that are constant over time (e.g., gender).

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 W_{1i} + \beta_{1+r} W_{ri} + u_i \tag{2}$$

Why might we include additional regressors?

1. Adding controls allow us to get a more precise estimate of β_1. Specifically, if we have controls, the variance of $\widehat{u}_i$ falls, so $SE(\widehat{\beta}_1)$ falls as well.
2. Adding the W's is necessary if randomization was based on covariates. Suppose that randomization was conducted in such a way that old people had an 80% probability of getting the treatment, but young people had a 40% of getting the treatment. Because young and old people have different probabilities of being in the treatment group, the variable T_i is correlated with u_i in equation (1) since u_i contains age. Thus, we need to include *age* as a control variable to get an unbiased estimate of β_1.

1.3 Threats to Internal Validity of Experiments

Failure to randomize. In the example experiment described above, this means that Step 3 fails. This happens when something else other than random assignment was done (for example, maybe we set $T_i = 1$ for the sickest patients). This leads to biased estimates of the causal effect. Using baseline data, we can test whether the randomization was successful by regressing T_i on the W's. Then, we can test the joint hypothesis that the coefficients on the W's are all zero. Rejecting the null hypothesis would provide evidence that randomization was successful.

Failure to follow treatment protocol. In the example experiment described above, this means that Step 4 fails. For example, some control group participants might have gotten the actual drug instead of the placebo; this is called "partial compliance." There are many reasons why this can happen, such as human error. Additionally, partial compliance can happen if study participants complain and asked to be put in one group versus another, and this would lead to a biased estimate of the causal effect. How can we use econometric tools to solve it? We could use instrumental variables were the instrument Z_i corresponds to what was randomly assigned to person i, and X_i corresponds to what person i actually got.

Attrition. This happens when study participants drop out of the study after they have been randomized. This can happen because study participants don't want to participate anymore, which affects Step 4 in the example experiment above, or because we are not able to contact study participants during endline data collection (e.g., they may have changed addresses, etc.), which affects Step 5 in the above example. If attrition is random, it doesn't affect bias. If attrition is not random, then we would get a biased estimate of the causal effect. There is no solution to attrition other than trying our best to keep track of participants.

Experimental Effects. Participants change their behavior because they know they're being watched as part of the study. This is called Hawthorne effects, and leads to biased estimates. One potential approach that can mitigate this issue is to use a "double-blind" protocol, which means that neither the study participants nor the experimenter know who is in the treatment or the control group.

1.4 Threats to External Validity

Non-representative sample. Participants in our experiment might not be representative of the population of interest. For example, the population of interest might be all adults in Norway, but our study participants consist of patients from hospitals in Oslo. To get a representative sample instead, we can take a simple random sample from the population of interest.

Non-representative program. The treatment or policy studied differs from the treatment or policy that will actually be implemented on a broader scale. For instance, suppose we were interested in studying the effect of small class size (a new government program) on children's test scores. The program we study lasts only for the length of the experiment, while the actual program that the government is considering implementing throughout the country might be available for longer periods of time.

General equilibrium effects. The environment changes because of the treatment. Turning a small, temporary program into a widespread, permanent program might change the economic environment sufficiently that the result from the experiment cannot be generalized.

2 Quasi-Experiments

2.1 Overview

In a quasi-experiment, also called a natural experiment, randomness is introduced by variations in individual circumstances that make it appear *as if* the treatment is randomly assigned. In a true experiment, the treatment and control groups are explicitly chosen. In a quasi-experiment, the treatment and control groups arise because of laws, circumstances, accidents of nature, etc. such that some entities are treated and others are not.

2.2 Analysis

How do we analyze quasi-experiments? The exact estimation approach depends on the set-up of the quasi-experiment, but we will discuss four approaches: differences-in-differences, differences-in-differences with additional regressors, instrumental variables, and regression discontinuity.

2.2.1 Differences-in-Differences

Note that as a shorthand for "differences-in-differences," I will also use "diff-in-diff" or "DD." Also, while this method is generally used for analyzing quasi-experiments, it can be used for analyzing experiments as well. Furthermore, diff-in-diff requires panel data (i.e., data on the same set of entities for at least two time periods).

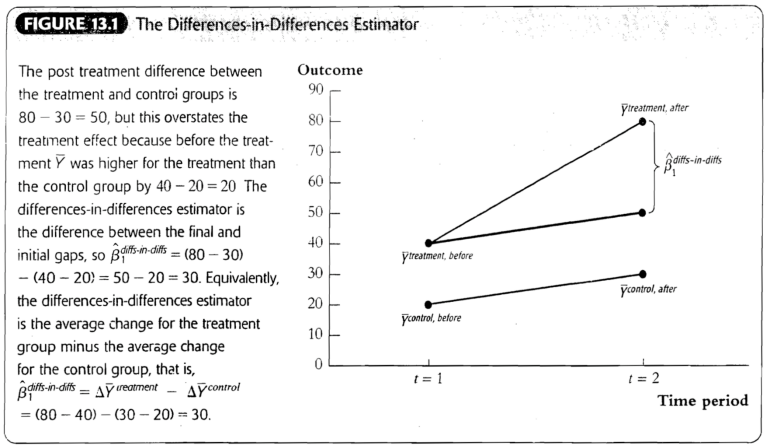
To understand DD, let's consider the following example. What is the causal effect of increasing the minimum wage on employment? This example comes from a very famous paper by Card and Krueger (1994). We have panel data with two years (1991 and 1993) of fast food restaurants in two states, New Jersey (NJ) and Pennsylvania (PA). NJ raised the minimum wage in 1992 from \$4.25 to \$5.05. In PA, the minimum wage did not change during this period, and it remained at \$4.25. Our "treatment group"

then is NJ, and our "control group" is PA.

Suppose that using these data, we have the following sample means (i.e., the sample average number of employees across fast food restaurants for each state and year), where "before" is 1991 and "after" is 1993. We can use these sample means to find the DD estimator, as follows.

State	After	Before	Difference
NJ	80	40	40
PA	30	20	10
Difference	50	20	30

The differences-in-differences estimator is equal to 30. A graphical representation of differences-in-differences can be found in Figure 13.1 in the textbook (which I reproduce below), and it also gives a very good explanation of the differences-in-differences.



Differences-in-differences can also be estimated using a regression. The regression is given by

$$emp_{it} = \beta_0 + \beta_1 NJ_i + \beta_2 POST_t + \beta_3 NJ_i * POST_t + u_{it} \tag{3}$$

where i denotes a fast food restaurant and t denotes years. The outcome variable emp_{it} is the number of employees in restaurant i at time t. The variable NJ_i is a dummy variable equal to one if restaurant i is in New Jersey, the variable $POST_t$ is a dummy variable equal to one if the year is 1993 (i.e., post or after the minimum wage change).

The coefficient of interest in equation (3) is β_3, which gives us the differences-in-differences estimator. Note that $\widehat{\beta}_3 = (\overline{emp}^{NJ,93} - \overline{emp}^{NJ,91}) - (\overline{emp}^{PA,93} - \overline{emp}^{PA,91})$, which is exactly the same as the difference in sample means as shown in the table above.

An important underlying assumption of DD is called the parallel trends assumption, which we can see in Figure 13.1 above. This assumption holds that if there was no minimum wage change, PA and NJ would have trended in the same way. Another way to say this is that in the absence of a minimum wage change, the PA trend (i.e., increasing by 10, from 20 to 30) would have been what we would have expected for NJ (i.e., increasing by 10, from 40 to 50). This assumption cannot be tested directly because there are is only one reality (i.e., there is no other world that exists where we could see what would have happened to NJ). However, one way to probe into this assumption is to look at periods before 1991. For example, if PA and NJ had similar trends between 1981-1991, then we can be more confident that PA and NJ would have had similar trends in 1993 had there not been a minimum wage change in NJ.

2.2.2 Differences-in-differences with additional regressors

We can augment the regression in equation (3) by adding control variables (denoted as W's), as follows.

$$emp_{it} = \beta_0 + \beta_1 NJ_i + \beta_2 POST_t + \beta_3 NJ_i * POST_t + \beta_4 W_{1it} + \ldots + \beta_{3+r} W_{rit} + u_{it} \tag{4}$$

Why might we add control variables? Similar to Section 1.2.2, we might do so if our quasi-experiment is such that the treatment is "as-if" randomly received conditional on the W's.

2.2.3 Instrumental Variables

We can use the IV approach if: (1) the quasi-experiment yields an instrument Z that influences receipt of the treatment, (2) data are available on both Z_i and the treatment actually received X_i, and (3) Z_i is as-if randomly assigned. If so, then Z_i would be a valid instrument for X_i.

Consider the following example, based on a paper by Angrist (1990). What is the causal effect of serving in the military affect labor market earnings? The regression we want to estimate is $earnings_i = \beta_0 + \beta_1 military_i + u_i$. A natural experiment to answer this question arises from the lottery system during the Vietnam War. Specifically, during the war, there was a draft, where priority for the draft was determined by a national lottery: the days of the year were randomly ordered 1 through 365, and those with birthdates ordered first were drafted before those with birthdates ordered second, etc. In short, people who got a low lottery number were more likely to have served in the military than those who got a high lottery number. Thus, the lottery can be used as an instrument (Z_i) for whether a person served in the military (X_i).

2.2.4 Regression Discontinuity

Let's look at the following example, based on Thistlethwaite and Campbell (1960). What is the effect of college scholarships (i.e., a grant or a stipend) on students' earnings later in life? To get the scholarship (i.e., the "treatment"), students must take an exam, and those who score 75 or higher on the exam get the scholarship.

The intuition for regression discontinuity is to compare the wages of those who were just above the cutoff of 75 to those who were just below the cutoff. For example, consider people who got a grade of 74 and 76 on the exam. Being just below or just above the cutoff can be thought of as random because those who got a grade of 74 just had a little bit of bad luck and didn't get the scholarship, while those who got a grade of 76 just got a little bit of good luck and got the scholarship.

There are two types of regression discontinuity (RD): sharp and fuzzy. In a sharp RD, the cutoff is strictly followed. In a fuzzy RD, the cutoff influences the probability of treatment, but is not the sole determinant.

Sharp RD. Because the cutoff is strictly followed in a sharp RD, all students who got a grade of 75 or higher got the scholarship, and those who got a grade lower than 75 did not get the scholarship. To estimate the causal effect of the scholarship, we would implement the following regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i \tag{5}$$

where X_i is a dummy variable equal to one if person i got the scholarship, and W_i is the exam grade.

Fuzzy RD. In a fuzzy RD, getting an exam grade ≥ 75 or < 75 influences the probability of passing, but is not the sole determinant (e.g., if in addition to the exam, one also needs to be above 18 years old to get the scholarship). In this case, we can use the cutoff rule as an instrument for X_i (the dummy variable for whether the person got the scholarship). Specifically, let Z_i be a dummy equal to one if the exam grade is ≥ 75, and zero otherwise. We can then use Z_i as an instrument for X_i to estimate regression equation (5).