EC220 Introduction to Econometrics

Dr Canh T. Dang

Lecture Notes

Cedric Tan

September 2019

Causal Inference and Randomised Experiments

We begin Econometrics by understanding the difference between causality and correlation. We look into the different types of causality that can be applied to our econometric examples. We look at potential outcomes, counterfactuals and the issue of selection bias. Further, experiments are lightly covered based on observable and unobservable characteristics. We also look into the motivation for randomised experiments and how they help us reduce bias or conflict within our inferences.   [1]

Concept

Concept.   [2]

Contents

1   Introduction to Econometrics: Michaelmas Term

The initial focus of MT is applied econometrics, particularly causal questions such as "what-if" questions. Mislabelling causality as correlation can be a critical error that people make when analysing data. The course intends to teach you how to analyse data and answer economic questions using "econometrics" and data.

Examples of what-ifs are below:

- What happens to a country if it withdraws from a trade agreement?

- What is the effect of parents' education on children's education?

- What is the impact on your health if you go to a hospital?

1.1   Causality

A causes B: A contributes (or influences) to the occurrence of event B. The *cause* A is partly responsible for the *effect* B, and the effect B is partly dependent on the cause A. A can be necessary for the occurrence of B, but A can simply lead to fluctuations in B, this is still a causal relationship.

So we can take two definitions for causality:

- A is a necessary condition for B to occur

- A can cause fluctuations in B

Labels are also necessary for the structure of causation:

- Event A is called Treatment

- Event B is called the Outcome

- A third variable that causes the two events to happen is called a Confounder

We can have reverse causality - A causes B but also B causes A. An example would be *Umbrellas* and *Rain* where bringing umbrellas is caused by the possibility of rain. Sometimes timing helps to establish causality - as *Rain* happens before *Umbrellas*, *Umbrellas* cannot cause *Rain*.

Reasons that A and B are correlated:

1. A causes B (direct causation)

2. B causes A (reverse causation)

3. A and B are consequences of a common cause but do not cause each other (confounder)

4. A causes B and B causes A (bidirectional causation)

5. A causes C which then causes B (indirect causation)

6. No connection between A and B, the correlation is pure coincidence

All statistical techniques only establish associations, causation requires interpretation.
Correlation: the extent to which A and B tend to decrease and increase at the same time.

Causation can occur without correlation, here is an example for medicine:
Illness (A) can cause death (B), but nowadays healthcare (C) can eliminate the correlation between common illness and death.

1.2   Why Causality?

This is the economist's comparative advantage, the ability to infer causality from correlation.

Examples of causality are listed below with classifications like the above.

- Direct Causation:

- Reverse Causation:

- Confounder Problem:

- Bidirectional Causation:

- Indirect Causation

- Pure Coincidence:

2   Counterfactuals

Insured or not insured. Here we will see the use of counterfactuals to infer causality.

2.1   Health Insurance

What is the effect of health insurance on health? Health insurnace started as a voluntary scheme. The Obama "Affordable Care Act": attempted to compel the whole of America to buy health insurance. Similarly, is the NHS a good policy?

Question: What is the effect of health insurance on health expenditures and on health outcomes?

2.1.1   Health Insurance: Causal Questions

We want to compare:

- The health of someone with insurance Y_i^T

- The health of the same person without insurance Y_i^C

Though this is not a realistic possibility simply because we cannot have the exact same person simultaneously adopting both approaches. Although the data might show that there is a correlation between Health Insurance and Health, there could be other factors. For example, the United States spends more of its GDP on health care than do other developed nations, yet Americans are surprisingly unhealthy. However, America is also unusual in that is has no universal health insurance scheme. Perhaps there is a causal connection here.

- The causal effect of insurance: heaving health insurance may lead to better health because of better health care

- The reverse causal effect: the less healthy are more likely to buy insurance

- Confounder effect: the more educated tend to buy insurance more often and they know how to live healthier (possibly)

- Pure coincidence: another possibility

Many of the working, prime-age poor, however, have long been uninsured. In fact, many uninsured Americans have chosen not to participate in an employer-provided insurance plans. Using the National Health Interview Survey (NHIS), that rates health from poor (1) to excellent (5), we can gauge, with characteristics and analysis of whether or not someone has health insurance or not, the possibility of a causal effect.

Terminology to take into account when looking into this case study:

- Health insurance coverage for individual i is described by a binary random variable; we call this the *treatment*

$$D_i = 0, 1$$

- The outcome of interest, a measure of health status denoted by Y_i and is the index proposed by the NHIS i.e. on a scale from 1 to 5

- Potential outcomes: describes what would have happened to someone under the scenarios where they had or had not been insured.

- Those with insurance are called the *treatment group*

- Those without insurance are called the *control group*; a good control group reveals the fate of the treated in a counterfactual world where they are not treated

2.1.2   Fruitless and Fruitful comparisons

The critical thing to be mindful of is keeping things equal i.e. *ceteris paribus*. Comparisons of people with and without health insurance are not apples to apples; such contrasts are apples to oranges, or worse. Among other differences, those with health insurance are better educated, have higher income, and are more likely to be working than the uninsured. Many of the differences in characteristics between the two classes are large and most are statistically precise enough to rule out the hypothesis that these discrepancies are merely chance findings. Thus it won't be surprising that most variables listed are high correlated with health as well as with health insurance status. Some other highly associated characteristics

- Education

- Family income

- Age

- Employed vs Unemployed (less powerful indicator)

We use the Robert Frost metaphor for understanding the potential outcomes.

- For everybody there are two potential outcomes **but only one is actually observed**:

$$Counterfactual = \begin{cases} Y_{1i} & if \ D_i = 0 \\ Y_{0i} & if \ D_i = 1 \end{cases}$$

- The treatment effect is the difference between the actual outcome and the counterfactual given the person has insurance. Thus the effect of having insurance for individual i is $Y_{1i} - Y_{0i}$

To cement this down even further, lets consider the story of two students (K and M) going to MIT who have the option to subscribe to the MIT health insurance plan. Upon reflection, K decides that the MIT insurance is worth paying for since he fears he might get sick. Let's say that for K $Y_{0i} = 3$ and $Y_{1i} = 4$ for $i = K$ For K, the causal effect of insurance is one step up on the NHIS scale:

$$Y_{1,K} - Y_{0,K} = 1$$

M is also coming to MIT but does not fear sickness whatsoever. Not concerned about the winters in Boston, M does not think she will fall sick easily so opts not to buy the MIT insurance. Because $Y_{0,M} = Y_{1,M} = 1$ we have:

$$Y_{1,M} - Y_{0,M} = 0$$

We can summarise this data into a table below:

	K	M
Potential outcome without insurance: Y_{0i}	3	5
Potential outcome with insurance: Y_{1i}	4	5
Treatment (insurance status chosen): D_i	1	0
Actual health outcome: Y_i	4	5
Treatment effect: $Y_{1i} - Y_{0i}$	1	0

There is a selection problem for K and M. This is because:

- We have:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

  Which we can call the observed difference in average health.

- This is equivalent to:

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \text{ (Average treatment effect on the treated)} +$$
$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \text{ (Selection bias)}$$

  Which we can call the average treatment effect on the treated plus the selection bias

If we look at K and M again and compare the two outcomes they have given themselves, based on their decision to buy or not to buy insurance, we can see that K has a NHIS rating of $Y_K = Y_{1,K} = 4$ while M has a NHIS rating of $Y_M = Y_{0,M} = 5$. Taking the difference between them we obtain:

$$Y_K - Y_M = -1$$

From this, we might believe that K's decision to purchase insurance is counter-productive since M, even without purchasing insurance, has a higher NHIS rating than K. In fact, the comparison between frail K and hearty M tells us little about the causal effects of their choices. This can be seen by linking observed and potential outcomes as follows:

$$Y_K - Y_M = Y_{1,K} - Y_{0,M} = [(Y_{1,K} - Y_{0,K}) + (Y_{0,K} - Y_{0,M})]$$

The second part of the equation (in square brackets) is derived byu adding and subtracting $Y_{0,K}$, thereby generating two hidden comparisons that determine the one we see. The first comparison is the causal effect of health insurance on K, which is equal to 1. The second is the difference between the two student's health status were both to decide insurance which is equal to -2. This -2 reflects K's relative frailty. Therefore, in the context of our effort to uncover causal effects we find the lack of comparability captured by the second term, as mentioned above, which is the **selection bias**.

3   Introduction to Experiments

How do we solve the selection bias? One method is through random assignment of the treatment factor. That is to say that if D_i is randomly assigned, it is (statistically) independent of potential outcomes such that there is no difference between $E[Y_{0i}|D_i = 1]$ and $E[Y_{0i}|D_i = 0]$. That means we eliminate out selection bias factor $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = 0$.

3.1   Developing the experiment

Remember K and M from the MIT observation. If we were to randomly assign treatment in this case, the sample size would be too small. Further, characteristics between K and M would be too different and we would not be able to check for balance.

Thus, we can use the law of large numbers (LLN) to get a larger sample to test on. The theorem states:

The larger the sample, the closer the sample average will become to the population mean. By making the sample large enough, sample statistics and population can be brought as close to what we want.

The mathematical expectation is such that:

The mathematical expectation of a variable, Y_i, is written $E[Y_i]$, is the population average of this variable. If Y_i is a variable generated by a random process, such as throwing a die, $E[Y_i]$ is the average in infinitely many repetitions of this process. If Y_i is a variable that comes from a sample survey, $E[Y_i]$ is the average obtained if everyone in the population from which the sample is drawn were to be enumerated.

By having a large sample, we can wash out the random characteristics within our sample size making our experiment more indicative. Our aim is to remove the selection bias when inferring a causal effect, thus, with enough subjects, we can assure balance across the groups. Balance means no *systematic differences* in characteristics e.g. we hold a similar fraction of men and women in each group. We can balance for observed characteristics but we can only assume randomisation would work for unobserved characteristics, such as motivations, incentives and inner behaviours.

3.2   Randomized Trials

Experimental random assignment eliminates selection bias. The logistics of a randomized experiment, sometimes called a randomized trial, can be complex, but the logic is simple. Taking note of our previous issue of characteristics between K and M, random assignment, with the LLN, makes this comparison *ceteris paribus*.

In randomized trials, experimental samples are created by sampling from a population we'd like to study rather than by repeating a game, but the LLN works just the same. When sampled subjects are randomly divided (as if by a coin toss) into treatment and control groups, they come from the same underlying population. the LLN therefore promises that those in randomly assigned treatment and control samples will be similar if the samples are large enough. Randomly assigned groups should be similar in every way, including in ways that we cannot easily measure or observe. This is the approach to eliminating selection bias.

We can describe the power of random assignment using this conditional expectation:

The conditional expectation of a variable, Y_i, is given a dummy variable, $D_i = 1$, is written $E[Y_i|D_i = 1]$. This is the average of Y_i in the population that has D_i equal to 1. Likewise, the conditional expectation of a variable, Y_i, given $D_i = 0$, written $E[Y_i|D_i = 0]$, is the average of Y_i in the population that has D_i equal to 0. If Y_i and D_i are variables generated by a random process, such as throwing a die under different circumstances, $E[Y_i|D_i = d]$ is the average of infinitely many repetitions of this process while holding the circumstances indicated by D_i at D. If Y_i and D_i come from a sample survey, $E[Y_i|D_i = d]$ is the average computed when everyone in the population who has $D_i = d$ is sampled.

Because randomly assigned treatment and control groups come from the same underlying population, they are the same in every way, including their expected Y_{0i}. In other words, the conditional expectations $E[Y_{0i}|D_i = 1]$ and $E[Y_{0i}|D_i = 0]$ are the same. This means that when D_i is randomly assigned, $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ and the difference in expectations by treatment status captures the causal effect of treatment:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$
$$= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$
$$= E[Y_{0i} + \kappa|D_i = 1] - E[Y_{0i}|D_i = 0]$$
$$= \kappa + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$
$$= \kappa; \text{ which is the average causal effect of insurance on health in this case.}$$

3.3   Results and Replication

The replication of the results provided a much richer interpretation of a health insurance policy that potentially affects millions of people. Below are some reasons for differences in replication between the RAND and Oregon Trail experiments:

- Economics is not physics: there are no causal relationships fixed across time and space

- There were different set ups: RAND - with its varying generosity, Oregon trial - insured to uninsured

- There were different populations

- Medicine changed between 1974 and 2008

Similar results from multiple experiments for related treatments and related population of interests boosts our confidence about the external validity i.e. the transferability of the theory, of our result. The same goes for non-experimental (observational) methods.

Sometimes experiments simply cannot be done. Here are some reasons why:

- Ethical issues

- Budgets

- Inability to experiment with things that have already happened e.g. Fiscal Policy

But experiments are constantly occurring within business and government. For example, Google Adwords will randomise adverts to see clicks or compare search results or search trends with marketing targets.

Even if we can't do an experiment, it is useful to think about one i.e. experiments which can be a benchmark for observational methods. Here, we can mimic the ideal experiment through a comparison. However, the experiment is not a silver bullet, there are questions that cannot be answered. For example, events that have happened once and cannot be redone again such as Brexit or the 2007 financial crisis. Although there are similar scenarios, we cannot say that it is the same situation.

4 Experimental Approaches

4.1 Sampling

When estimating the treatment effect or checking for balance, how can we tell whether a difference in the means did not arise by chance? There are serveral sources for variability in our estimates:

- Sample selection:

- Statistical modelling: our choice of estimators differs

- Imprecision in measurement: the information is not precisely measured

- Sampling: variability caused by different samples from the population

Here, we focus on how sampling can affect the means in this lecture.

4.1.1 Populations and Samples

We try to analyse a representative sample rather than a whole population. Below are a few terms to take note of:

- Population mean: the expectation of the population distribution of the Data Y_i or μ. This is a parameter (a fixed number). $\mu = E[Y_i]$

- We would like to also *estimate* this parameter. The sample average:

$$\overline{Y}_n = Avg_n(Y_i) = \frac{1}{n}\sum_{i=1}^{n}Y_i$$

  is an estimator of the population mean. An estimator is an example of a sample statistic, a function of sample data used to *infer* the parameter.

- An *estimate* is the value the estimator takes on for a particular sample

- Sampling variability is about the size of the differences between: $\overline{Y}_n - \mu$

4.1.2 Unbiasedness

If the estimator is unbiased, we must have $E[\overline{Y}_n] = E[Y_i] = \mu$. This is a very desirable property. Unbiasedness means that in many repeated random samples, the average of the sample averages will equate the population mean. The sample average does not deviate from the mean in any direction, on average. Unbiasedness must hold for samples of any size.

We want to assess the size of $\overline{Y}_n - \mu$. Unbiasedness means that:

$$E[\overline{Y}_n - \mu] = 0$$

We are also interested in the absolute values of these differences, so we typically look at the squared error terms otherwise we will bear no good result since the expected value, not squared, will average out to 0. Hence we have the equation: $E[\overline{Y}_n - \mu]^2 = 0$

For any random variable Y_i, we define variance as:

$$Var(Y_i) = E[\overline{Y}_n - \mu]^2$$

Conceptually, $E[\overline{Y}_n - E[\overline{Y}_n]]^2$ is the variance of the sample average we would obtain if we drew (infinitely) many samples. If the estimator is unbiased we will have:

$$E[\overline{Y}_n - E[\overline{Y}_n]]^2 = E[\overline{Y}_n - \mu]^2$$

Variance of the sample average:

$$Var(\overline{Y}_n) = Var(\frac{1}{n}\sum_{i=n}^{n}Y_i)$$

The variance of the sample average depends on two things:

1. The variance of the underlying observations Y_i

2. The sample size n

The standard deviation of the sample average
Note that variances don't have the same scale to the original data (due to the square involved) - it makes sense to work with its square root. The sampling standard deviation of the sample average is:

$$SE(\overline{Y}_n) = \sqrt{Var(\overline{Y}_n)} = \frac{\sigma}{\sqrt{n}}$$

Comparing averages. In analysing experiments, we want to compare the two averages from the treatment and control group. We want to estimate:

$$Var(\overline{Y}_n^1 - \overline{Y}_m^2)$$

Large and small standard errors Hypothesis testing The student's t-statistic Could the estimate of \$x be consistent with the hypothesis that the two groups averages are the same? Let:

$$\overline{Y}_n^1 - \overline{Y}_n^2 = \mu$$

And we want to see if $\mu = \mu_0 = 0$