Treatment Effects and Potential Outcomes

Introduction

- We are going to look at a slightly different way (linked to, but slightly different to regression) to think about causality.
- This framework was proposed by Donald Rubin in the 1970's and can help us think about what causality really means.
- It will also be helpful as we come to talk about the last few techniques to identify causality.

An Example

- To help keep things concrete, consider the following example: we want to know if having health insurance improves your health.
- It may be difficult to determine the exact causal effect because:
    - The more educated are more likely to buy health insurance but are also likely to take better care of themselves - Confounding variable (OVB).
    - The less healthy you are, the more likely you are to buy health insurance - Reverse causality.

Potential Outcomes

- Health insurance coverage for individual i is described by a random variable, **the treatment**

  $$D_i = \{0, 1\},$$

  that is, $D_i = 1$ if person i has health coverage, and $D_i = 0$ if person i does not.
- The outcome of interest, some measure of health status, is denoted Y_i.
- We can think of **potential outcomes** in the following way: What would have happened to someone who was insured if they had not been insured?
- Hence, for everyone, there are two potential outcomes:

  $$\begin{array}{ll} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{array}$$

Observed Outcomes

- The effect of having insurance for individual i is $Y_{i1} - Y_{i0}$, this is known as **the treatment effect**.

- The **observed outcome**, Y_i , can be written in terms of potential outcomes as follows

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i. \end{aligned}$$

- The problem is, we only observe **either** Y_{i1} **or** Y_{i0} for a single individual, **not both!**

The Selection Problem

- We can write the difference in observed average outcomes for those with and those without health insurance as the sum of two terms:

$$
\begin{aligned}
E\left[Y_i \middle| D_i = 1\right] - E\left[Y_i \middle| D_i = 0\right] &= E\left[Y_{1i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 0\right] \\
&= E\left[Y_{1i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 1\right] + \\
&\quad E\left[Y_{0i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 0\right] \\
&= E\left[Y_{1i} - Y_{0i} \middle| D_i = 1\right] + \\
&\quad E\left[Y_{0i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 0\right].
\end{aligned}
$$

- The blue term is known as the **average treatment effect on the treated (ATET)** . The term in red is the **selection bias**.
- The selection bias is related to the **assignment rule**.

Random Assignment

- Recall the fertiliser and crop yield example. We said that we could obtain the causal effect if the assignment of fertiliser to different fields was random.
- We can now see mathematically, why this works... If D_i is randomly assigned, it is independent of potential outcomes. Therefore:

$$\begin{aligned} E\left[Y_i \middle| D_i = 1\right] - E\left[Y_i \middle| D_i = 0\right] &= E\left[Y_{1i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 0\right] \\ &= E\left[Y_{1i} \middle| D_i = 1\right] - E\left[Y_{0i} \middle| D_i = 1\right]. \end{aligned}$$

- The second equality follows because $E\left[Y_{0i}\right]$ does not change if $D_i = 0$ or if $D_i = 1$.
- In fact, we can write

$$\begin{aligned} E\left[Y_i \middle| D_i = 1\right] - E\left[Y_i \middle| D_i = 0\right] &= E\left[Y_{1i} - Y_{0i} \middle| D_i = 1\right] \\ &= E\left[Y_{1i} - Y_{0i}\right]. \end{aligned}$$

Balance in Characteristics

- If we **only use two fields** and randomly assign one to be fertislised and one not, the differences in the outcomes could be due to other differences between the two fields.
- However, as our sample grows larger, the law of large numbers (LLN) says that the average characteristics of the fertilised fields will be equal to the average characteristics of the un-fertislised fields.
- We have washed out any idiosyncratic differences between the two groups by having a large sample and random assignment. We are now comparing apples with apples!

Balance in Characteristics

- So, we should now have two groups where there are no systematic differences between them, apart from one group has been treated and the other has not.
- We can check for balance among our regressors, to see if the randomisation has worked as it should.
- Most importantly, this randomisation should balance all unobserved characteristics too. For example, motivation, intelligence, incentives, tastes, etc.

Relating this to Regression

- Suppose that the treatment effect is the same for every person: $Y_{1i} - Y_{0i} = \beta$.

- Recall

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i . \end{aligned}$$

- We can write

$$\begin{array}{ccccccc} Y_i = & \alpha & + & \beta & D_i & + & \epsilon_i \\ & \Downarrow & & \Downarrow & & & \Downarrow \\ & E[Y_{0i}] & & (Y_{1i} - Y_{0i}) & & & Y_{0i} - E[Y_{0i}] \end{array}$$

- This is just a regression!

Relating this to Regression

- Now consider the conditional expectation of Y_i for when treatment status is switched on and off:

$$\begin{aligned} E\left[Y_i \middle| D_i = 1\right] &= \alpha + \beta + E\left[\epsilon_i \middle| D_i = 1\right] \\ E\left[Y_i \middle| D_i = 0\right] &= \alpha + E\left[\epsilon_i \middle| D_i = 0\right]. \end{aligned}$$

- From this, we can write

$$E\left[Y_i \middle| D_i = 1\right] - E\left[Y_i \middle| D_i = 0\right] = \beta + \left(E\left[\epsilon_i \middle| D_i = 1\right] - E\left[\epsilon_i \middle| D_i = 0\right]\right)$$

  where β is the treatment effect (actually the ATET), and the second term is the selection bias.

- This is the same formula we had before. In fact, if you plug in $\epsilon_i = Y_{0i} - E\left[Y_{0i}\right]$ in the selection bias term, you get exactly the form that we gave on Slide 6 (in red).

Endogeneity / Selection Bias

- In order for $E\left[Y_i \middle| D_i = 1\right] - E\left[Y_i \middle| D_i = 0\right]$ to give us the causal effect, we need the selection bias to be zero

$$E\left[\epsilon_i \middle| D_i = 1\right] = E\left[\epsilon_i \middle| D_i = 0\right]$$

- But this just says that we want the error term to be mean independent of the treatment... the exact same thing we wanted in our regressions.
- The only difference is the name that we're giving it!
- We can extend this to have some control variables X and then we need that the error term is mean independent of the treatment conditional on the X.

Heterogenous Treatment Effects

- A couple of slides back, we mentioned that we were going to assume that the treatment effect is the same for everyone: $Y_{1i} - Y_{0i} = \beta$.

- But, why should this be the case! We speak about 'the' causal effect, but it could be completely different for different people. Maybe the *average* treatment effect (ATE) isn't so informative.

- It's unrealistic to try to estimate the causal effect for every type of person. Of course, you can split your sample according to X, e.g. by gender, and estimate two separate models, but this can get unwieldy pretty quickly.

- There is a broader way to think about the types of people we have. But first, an example...

Heterogenous Treatment Effects - Example

- Imagine the following (unrealistic) experiment:
    - The government takes a sample of high school students and randomly gives some of them 50 000DKK to help them go to university. The others get nothing.

- We can then use this experiment to determine the causal impact of going to university on some outcome of choice (earnings/marriage/criminal activity).

- We can do this by using whether the student received 50 000DKK as an instrument for going to university. (Think about why this works as a valid and relevant IV).

Four Types of People

- In this example, we have a binary instrument, Z, (receives money) and a binary treatment, D, (goes to university).

- The decision to go to university or not depends on whether you receive money. So we write D as a function of Z, $D_i(Z_i = z)$

- We can imagine four different groups of people based on how they react to the instrument:
    - **Always Takers (A)**: $D_i(Z_i = 1) = 1$ and $D_i(Z_i = 0) = 1$ They always go to university.
    - **Never Takers (N)**: $D_i(Z_i = 1) = 0$ and $D_i(Z_i = 0) = 0$ They never go to university.
    - **Compliers (C)**: $D_i(Z_i = 1) = 1$ and $D_i(Z_i = 0) = 0$ They only go to university if they get the money.
    - **Hipsters (H)**: $D_i(Z_i = 1) = 0$ and $D_i(Z_i = 0) = 1$ They do the opposite of what you expect/want (aka **Defiers**).

Average Outcomes

- Suppose our outcome of interest, y, is future wage. We can write our regression as

$$y_i = \alpha + \beta_i D_i + u_i.$$

- What is the expected wage difference for those who received money and those who did not? The expected wage for those given no money is

$$E\left[y_i | Z_i = 0\right] = \alpha + E\left[\beta_i D_i | Z_i = 0\right]$$

  note that $E\left[u_i | Z_i\right] = E\left[u_i\right] = 0$ because of the random allocation of Z.

- We can decompose the second part of this equation into parts for each type of person...

Average Outcomes - No Money

- When given no money ($Z_i = 0$), never-takers and compliers will choose $D_i = 0$. So these two groups drop out.
- For always-takers and hipsters, $D_i = 1$. So the second term becomes

$$\pi_A \beta_A + \pi_H \beta_H,$$

  where π_A denotes the proportion of always takers in the population, π_H similarly for hipsters.
- Notice that we assume the causal effect is the same for all hipsters, β_H, but this can be different to the causal effect for always-takers, β_A.

Average Outcomes - Money

- When given money ($Z_i = 1$), never-takers and hipsters will choose $D_i = 0$. So these two groups drop out.
- For always-takers and compliers, $D_i = 1$. So the second term becomes
$$\pi_A \beta_A + \pi_C \beta_C.$$

Differences

- This means that the expected difference between those who receive money and those who do not is:

$$\pi_C \beta_C - \pi_H \beta_H.$$

- The never-takers and the always-takers get washed out because they never change their behaviour, so we have no way of looking at the causal effect of university on wages because the instrument doesn't affect their choice.

- Using the same approach, we can determine the difference in the proportion of people attending college between those who received money and those who did not. It is given by:

$$\pi_C - \pi_H.$$

The IV Estimator

- In this binary IV / binary treatment model, it can be shown that the IV estimand is just the ratio of these two differences:

$$\beta_{IV} = \frac{\pi_C \beta_C - \pi_H \beta_H}{\pi_C - \pi_H}.$$

- We are likely to be interested in different components within this object, but there is no way for us to get at them from this IV estimator.

- This formula is also a little worrying because even if both causal effects are positive, if the proportion of hipsters is very high, the IV estimator can come out negative!

Solutions

- Some econometricians like to make a **monotonicity assumption**: $D_i(Z_i = 1) \geq D_i(Z_i = 0)\ \forall i$.

- This rules out the hipsters. Which means our IV estimator becomes

$$\beta_{IV} = \frac{\pi_C \beta_C}{\pi_C} = \beta_C.$$

- So the IV estimator identifies the causal effect for the compliers. This effect is known as the **Local Average Treatment Effect** (**LATE**).

- With the removal of the hipsters, LATE only tells us the causal effect for those for which the instrument has an effect.

- Other solutions involve placing different assumptions on the underlying structural model.

A Final Worry

- The last thought to keep you up at night...
- Imagine a slightly different experiment: instead of being given 50 000DKK, you get 100 000DKK.
- Even assuming we have no hipsters, the set of people who are now compliers is different to the original experiment. These people will generally have different effects of going to college than the former set of compliers.
- So, we get different estimates of “the” causal effect of going to college on wages, even though both instruments are completely valid.

Summary

- We have seen a new way to think about causality centered around counterfactual outcomes.
- We have linked this to what we have already seen in regression and shown that the two are equivalent.
- We looked at the Average Treatment Effect on the Treated, and the Average Treatment Effect.
- We discussed the problems that arise when we allow for heterogenous effects and looked at the Local Average Treatment Effect.