Treatment Effects

Christopher Taber

Department of Economics
University of Wisconsin-Madison

September 6, 2017
Notation

First a word on notation

I like to use $i$ subscripts on random variables to be clear that this is individual specific (or firm specific, or time specific, or state specific etc). 

As much as possible I will use capital letters to denote random variables

I will also try to be clear about conditioning so

$$E (Y_i \mid X_i = x)$$

means the expected value of $Y_i$ conditional on $X_i$ having the value $x$. This thing can be thought of as a function of $x$ and I’ll try to write it that way
I will use the notation

\[ E(Y_i \mid X_i) \]

to essentially mean

\[ E(Y_i \mid X_i = x) \]

for all values of \( x \) (in the support of \( X_i \))

For example if \( G_i \) is gender and takes on the value \( m \) and \( f \) then

\[ E(Y_i \mid G_i) = 0 \]

means that

\[ E(Y_i \mid G_i = m) = 0 \]
\[ E(Y_i \mid G_i = f) = 0 \]
I will also use the notation \( \approx \) to mean asymptotically equivalent so

\[
\frac{1}{N} \sum_{i=1}^{N} X_i \approx E(X_i)
\]

means that as \( N \) gets large, \( \frac{1}{N} \sum_{i=1}^{N} X_i \) converges (in probability or almost surely) to \( E(X_i) \).

I will not distinguish between different types of convergence (feel free to ignore this comment if you don’t know it means)
The key theme in this course will be causal estimation—we want to estimate the effect of some intervention or policy on an individual.

The standard way to think about these issues is in the “Treatment Effects” framework.

We will start with the simplest case where the default is that an individual (or firm or state or whatever) either receives some treatment or they do not.
The classic and easiest way to think about this is a drug.

We have a patient who is sick and either we give them the drug or we do not.

We let $Y_{0i}$ be the outcome that individual $i$ would receive if they did not get the drug.

and $Y_{1i}$ the outcome they would receive if they did receive the drug.
A key concept here is that of a “counterfactual”

Let $T_i$ be an indicator (0 or 1) for whether individual $i$ receives the drug treatment then what we generally get to see in the data is

$$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$$
That is

- for persons who receive the drug we observe $Y_{1i}$
- for persons who don’t receive the drug we observe $Y_{0i}$

So for people who receive the drug, $Y_{0i}$ is counterfactual—it is the outcome that person $i$ would have received if they had not received the drug.

Similarly $Y_{1i}$ is counterfactual for people who did not receive the drug. It is the outcome they would have obtained had they received the drug.
We define the treatment effect for individual $i$ as the change in their outcomes resulting from the treatment:

$$\Delta_i \equiv Y_{1i} - Y_{0i}$$
There are many many different treatments we can think of:

- a drug for Ebola
- a big fence between here and Mexico
- a year of school
- a month of training
- an increase in the federal minimum wage to $15
- a welfare program
- free tuition for public colleges
- inclusion of a public option in health exchanges
- exposure to polluted air
- access to clean water
- race or gender
- a cigarette a day while pregnant
- a glass of wine every night for dinner
- increased financial regulations on banks
- helmets on motorcyclists
- the death penalty
Our first issue: even if we knew the value of $\Delta_i$ for every person in the population that we study,

What do we present in the paper?

We must focus on a feature of the distribution.
The most common:

- **Average Treatment Effect (ATE)**
  \[ E(\Delta_i) \]

- **Treatment on the Treated (TT)**
  \[ E(\Delta_i \mid T_i = 1) \]

- **Treatment on the Untreated (TUT)**
  \[ E(\Delta_i \mid T_i = 0) \]

(Heckman and Vytlacil discuss Policy Relevant Treatment effects, but I need more notation than I currently have to define those)
These each answer very different questions

I will ignore TUT for the rest of these lecture notes because it is symmetric with TT

In terms of identification they are symmetric.
What does Identification mean?

Answer: different things to different people

I will stick to a more statistical definition (though informal)

A parameter $\theta$ is identified if under the maintained assumptions of the model it is uniquely determined by the population version of the data

This can be explained within this treatment effect framework:

In the data we observe $(Y_i, T_i)$, so the population version of the data is the full joint distribution of $Y_i$ and $T_i$

I used the phrase “population version of the data” to mean everything we could ever learn from the data. For example when $T_i = 1$ the data doesn’t inform us about $Y_{0i}$ so this is not part of the population version of the data.
What is identified from this type of data?

We can directly identify $Pr(T_i = 1)$

We can also observe

- the distribution of $Y_{0i}$ conditional on $T_i = 0$
- the distribution of $Y_{1i}$ conditional on $T_i = 1$
Identification of TT

By definition the TT parameter is

\[ TT = E(\Delta_i \mid T_i = 1) = E(Y_{1i} - Y_{0i} \mid T_i = 1) = E(Y_{1i} \mid T_i = 1) - E(Y_{0i} \mid T_i = 1) \]

Is this identified?

Well one part we get immediately: \( E(Y_{1i} \mid T_i = 1) \)

The other part is counterfactual: \( E(Y_{0i} \mid T_i = 1) \)

It is the expected value the treatments would have received if the treatment option were not available to them.

This is not measured in the data
To a large extent what this course and more generally causal estimation is about is trying to say something about $E(Y_{0i} \mid T_i = 1)$

All the different empirical approaches we examine do this in one way or another
Average Treatment Effect
This is slightly more complicated

\[ ATE = E(\Delta_i) = E(\Delta_i \mid T_i = 1)Pr(T_i = 1) + E(\Delta_i \mid T_i = 0)Pr(T_i = 0) \]
\[ = [E(Y_{1i} \mid T_i = 1) - E(Y_{0i} \mid T_i = 1)] Pr(T_i = 1) \]
\[ + [E(Y_{1i} \mid T_i = 0) - E(Y_{0i} \mid T_i = 0)] [1 - Pr(T_i = 1)] \]

All we can directly identify from the data is:

\[ E(Y_{1i} \mid T_i = 1), E(Y_{0i} \mid T_i = 0), Pr(T_i = 1) \]

In this case there are two missing pieces:

\[ E(Y_{1i} \mid T_i = 0), E(Y_{0i} \mid T_i = 1) \]

If we could figure these out we could identify the ATE
Extensions:

We have kept things simple by focusing on binary treatments, but more generally treatments could have multiple values:

- drug 1/drug 2
- classroom training/ job search assistance
- 4 yr college/2 yr college

Could also have continuous treatment

- hours of training
- dose of drug
This does not drastically change things-just makes more counterfactuals

For example with 2 different drugs we have three things:

- $Y_{0i}$: outcome taking neither drug
- $Y_{1i}$: outcome taking drug 1
- $Y_{2i}$: outcome taking drug 2

This gives a bunch of different things we can look at:

$$E(Y_{1i} - Y_{0i})$$
$$E(Y_{2i} - Y_{0i} \mid T_i = 2)$$
$$E(Y_{2i} - Y_{1i} \mid T_i = 1)$$

etc.
Randomized Control Trial

So how do I deal with this general identification problem?

Simplest thing is to randomize, if I randomly force some people to receive the treatment and others not to receive the treatments then $T_i$ is random and independent of underlying variables.

In particular it will be independent of $Y_{1i}$ and $Y_{0i}$ so

$$E (Y_{1i} | T_i = 0) = E (Y_{1i} | T_i = 1)$$
$$E (Y_{0i} | T_i = 1) = E (Y_{0i} | T_i = 0)$$

Then I have solved the problem

Since the things on the right hand side are identified, the things on the left hand side are as well
They work great when you can do them, but often experiments are not feasible:

- prohibitively expensive
  - typically social experiments are very expensive to run, and we sometimes need very large samples

- Unethical
  - effect of growing up with your own parents vs. foster home
  - effect of exposure to carbon monoxide on mortality

- Compliance very difficult
  - compare people who drink a glass of wine every day (for 30 years) to people who don’t drink at all

- Not feasible
  - monetary policy (GE effects)
  - effects of preschool program on adult outcomes, but we need to know the answer for a current policy discussion
Since we can not typically use experimental methods we typically are restricted other econometric methods.

We will spend the rest of this part of the course trying other methods.