ESTIMATION OF TREATMENT EFFECTS

AAEC 5126
INSTRUCTOR: KLAUS MOELTNER

Textbooks: Wooldridge (2010), Ch.21; Greene (2012), Ch.19;
Angrist and Pischke (2010), Ch. 3

Preliminaries

The estimation of treatment effects has experienced a resurgence in recent years, primarily due to the arrival of innovative identification methods. The term "treatment" is usually associated with some government policy such as a job training program, drivers' education sessions, or new rules governing standardized tests in public schools. However, it can be used more generally as some form of change in an institutional framework or even the physical environment that is expected to have an effect on individuals' economic decisions.

For example, treatment estimation models are increasingly used in property valuation, where the "treatment" can be an intervention such as a change in zoning laws, cleaning up superfund sites, or improving the quality of well water. In this context, treatments can also be "dished out" by nature - for example via extreme weather events or insect infestations (affecting trees and home values).

There are two fundamentally different treatment scenarios, depending on if the treatment is exogenous to the decision maker or not. In the first case, the treatment will be *ignorable* or *unconfounded* with respect to the outcome of interest (hourly wages, home values, health status, etc.), once we control for observables (our usual $\mathbf{x}$). This case is generally called *selection on observables.* The key point is not if people choose to undergo the treatment or if they are randomly assigned to it, but rather if whatever drives the assignment into a specific treatment status (usually binary - treated or not) is fully observed and can thus be included in the econometric model. Naturally, if the assignment is truly random and thus "forced upon" subjects, unconfoundedness is trivially satisfied.

In the second setting people's assignment to a specific treatment status is driven by observables and unobservables. The latter, in turn, are often also related to the outcome variable (for instance, innate "ability" or "spunk" driving both training participation and wage compensation). This case is generally referred to as *selection on unobservables*, or *treatment endogeneity.*

In this set of introductory lectures we will focus exclusively on the first scenario, selection on observables. In addition, we will limit our discussion to binary (0/1) treatments, which also constitute the bulk of recent applications. However, keep in mind that the treatment could also be continuous or integer counts (such as monetary incentives, amount of medication administered, number of outreach efforts made, etc.).

In terms of estimation methods, we will focus on the three most common methods: (i) regression analysis, (ii) non-parametric matching, and (iii) propensity score analysis.

Adopting Wooldridge's (2010) notation, let w be the treatment indicator. Thus, $w_i = 1$ if individual i receives the treatment *in reality, i.e. in an actually collected sample*, and $w_i = 0$ otherwise. Furthermore, let the hypothetical / theoretical outcome under treatment be y_1, and in absence of treatment y_0. Again, these are purely hypothetical constructs and refer to the underlying population.

In fact, the fundamental dilemma or challenge in treatment estimation arises because for each person i we generally only observe one or the other, i.e. y_{i1} or y_{i0}, but not both. The missing part needs to be inferred using varying econometric methods. It is important to realize from the notation that follows if an outcome is actually observed or only exists in theory.

For example, $y_1|w = 1$ is the actually observed outcome for the actually treated, while $y_0|w = 1$ is the counterfactual outcome for the actually treated - what y would have been for the treated in absence of treatment. The analogous holds for $y_0|w = 0$ versus $y_1|w = 0$. Naturally, we hope that we can learn about $y_0|w = 1$ from $y_1|w = 1$, and about $y_1|w = 0$ from $y_0|w = 0$. Several important conditions need to hold for this to be the case.

There are two population parameters of primary interest: The *average treatment effect*, ATE, and the *average treatment effect on the treated*, ATT. They are given as follows:

$$\begin{aligned} ATE &= \tau_{ate} = E\left(y_1 - y_0\right) \\ ATT &= \tau_{att} = E\left(y_1 - y_0 | w = 1\right) \end{aligned} \tag{1}$$

From a statistical perspective, all three of y_0, y_1 and w are (potentially) random variables with some underlying distribution in the population of interest. Thus, the expectation operator in (1) refers to that population distribution. Note that we can always write:

$$y = (1 - w)\, y_0 + w y_1 = y_0 + w\left(y_1 - y_0\right) \tag{2}$$

ATE and ATT will be equal if w is *statistically independent of both outcomes*, even unconditional on any observables. This is the "ideal" case of treatment analysis - rarely found in reality with observational data (but often assured *by construction* with experimental data). If it holds, we have

$$\tau_{att} = E\left(y_1 - y_0 | w = 1\right) = E\left(y_1 - y_0\right) = \tau_{ate} \tag{3}$$

This makes estimation easy, since we can now directly relate the expectation of the y_1 based on an *observed* sample of y_1's to the general expectation of y_1 (i.e. including both observed and counterfactual outcomes). The analogous holds for y_0. Specifically, we have (using (2)):

$$\begin{aligned} &E\left(y|w=1\right)=E\left(y_0+w\left(y_1-y_0\right)|w=1\right)=E\left(y_1|w=1\right)=E\left(y_1\right) \quad \text{and} \\ &E\left(y|w=0\right)=E\left(y_0+w\left(y_1-y_0\right)|w=0\right)=E\left(y_0|w=0\right)=E\left(y_0\right) \end{aligned} \tag{4}$$

In essence, independence assumes that the expectation of the counterfactual outcomes ($y_0|w=1, y_1|w=0$) is identical to that of the observed outcomes for the opposite treatment category. This immediately implies that we can estimate τ_{ate} (and thus τ_{att}) simply as $E\left(y_1\right)$ - $E\left(y_0\right)$. In practice, this can be accomplish simply by comparing the sample means of the two treatment groups, as discussed below.

Actually, for these results based on expectations to hold we only need the weaker assumption of *mean independence*, i.e.

$$\begin{aligned} E\left(y_0|w\right)&=E\left(y_0\right) \\ E\left(y_1|w\right)&=E\left(y_1\right) \end{aligned} \tag{5}$$

Another useful insight into the relationship between τ_{ate} and τ_{att} can be gained by decomposing outcomes as follows:

$$\begin{aligned} &y_0=E\left(y_0\right)+\left(y_0-E\left(y_0\right)\right)=\mu_0+\nu_0, \quad \text{and} \\ &y_1=E\left(y_1\right)+\left(y_1-E\left(y_1\right)\right)=\mu_1+\nu_1 \end{aligned} \tag{6}$$

Then:

$$\begin{aligned} &y_1-y_0=\mu_1-\mu_0+\left(\nu_1-\nu_0\right)=\tau_{ate}+\left(\nu_1-\nu_0\right), \quad \text{and} \\ &\tau_{att}=E\left(y_1-y_0|w=1\right)=\tau_{ate}+E\left(\nu_0-\nu_1|w=1\right) \end{aligned} \tag{7}$$

So the two effects will be equal if the expected deviation in outcome from the population mean is the same under both treatments for those who actually participated, i.e. $E\left(\nu_0-\nu_1|w=1\right)=0$. The full effect is captured by the difference in population means.

Fortunately, full unconditional independence or even unconditional mean independence between outcomes and treatment is not required to estimate τ_{ate} and τ_{att}. What is important, however, is *independence conditional on observables*. So let's introduce $\mathbf{x}$ to our framework.

Conditional Ignorability and Overlap. Now let's add some observable variables $\mathbf{x}$ that are also random in the population and - potentially - jointly distributed with y_0, y_1 and w. We can then define the *conditional* ATE and ATT as:

$$\begin{aligned} ATE|\mathbf{x}&=\tau_{ate}\left(\mathbf{x}\right)=E\left(y_1-y_0|\mathbf{x}\right) \\ ATT|\mathbf{x}&=\tau_{att}\left(\mathbf{x}\right)=E\left(y_1-y_0|\mathbf{x},w=1\right) \end{aligned} \tag{8}$$

We can then define the KEY assumption of *conditional ignorability* as follows:

Assumption 1. *Conditional on* $\mathbf{x}$*, the treatment* w *is independent of the (hypothetical) outcomes* y_0, y_1*. In other words: The* hypothetical *outcomes under each treatment are not related to* actual *treatment status.*

This assumption is also referred to as *conditional independence* or *unconfoundedness*. It is often sufficient to invoke the milder assumption of *conditional mean independence*, or *conditional ignorability in mean*:

Assumption 2. *Conditional on* $\mathbf{x}$*, expected outcomes are independent of treatment, or:*

$$\begin{aligned} E\left(y_0|\mathbf{x}, w\right) &= E\left(y_0|\mathbf{x}\right) \\ E\left(y_1|\mathbf{x}, w\right) &= E\left(y_1|\mathbf{x}\right) \end{aligned} \tag{9}$$

This says that given $\mathbf{x}$, the expected hypothetical outcome under either treatment is the same regardless of actual treatment status. **As before, this simply implies that we can use observed outcomes from the control group to infer the counterfactual for the treated, and vice versa.**

Some comments on this assumption of ignorability:
It automatically holds if w is a deterministic function of $\mathbf{x}$ (for example, "all men over 40 who have been unemployed for three months or longer receive the treatment"). The more $\mathbf{x}$ can explain about the treatment choice, the more likely will ignorability hold. If w depends on unobservable factors in addition to $\mathbf{x}$, these unobservables must be independent of y_0, y_1 and $\mathbf{x}$ for ignorability to still hold.

As a general rule, $\mathbf{x}$ should not include variables that themselves are affected by w (e.g. education decisions after treatment, but before measurement of outcome). This would violate the ignorability assumption.

Note that under Assumption [2], we have $\tau_{att}\left(\mathbf{x}\right) = \tau_{ate}\left(\mathbf{x}\right)$, since

$$\tau_{att}\left(\mathbf{x}\right) = E\left(y_1|\mathbf{x}, w=1\right) - E\left(y_0|\mathbf{x}, w=1\right) = E\left(y_1|\mathbf{x}\right) - E\left(y_0|\mathbf{x}\right) = \tau_{ate}\left(\mathbf{x}\right) \tag{10}$$

One additional assumption - primarily driven by practical concerns - must hold to allow for the identification of these treatment effects. It is called *overlap* and refers to the distribution of $\mathbf{x}$ over the treated and untreated sub-populations. Specifically:

Assumption 3. *Overlap:* $$\forall \mathbf{x} \in \mathcal{X}: \quad 0 < p\left(w=1|\mathbf{x}\right) < 1, \tag{11}$$

where $\mathcal{X}$ is the support of the covariates. Basically, this assumption states that for any setting of $\mathbf{x}$ in the *population* there is a non-zero probability of observing individuals with such an $\mathbf{x}$ in both the treated and control group. If this weren't the case, we couldn't use observed outcomes given some $\mathbf{x}$ to infer the counterfactual for the reverse outcome, given the same $\mathbf{x}$ (since individuals with such an $\mathbf{x}$ would have never undergone treatment or never *not* undergone treatment).

As an aside, note that the probability $p\left(w=1|\mathbf{x}\right)$ is know as the *propensity score.* It plays a central role in the estimation of treatment effects, as we will see below.

If we're only interested in ATT, a milder overlap assumption suffices for identification (in addition to Assumption [2]):

Assumption 4.
$$\forall\mathbf{x}\in\mathcal{X}:p\left(w=1|\mathbf{x}\right)<1, \tag{12}$$

So in this case, the probability of receiving treatment for some $\mathbf{x}$ can be zero - such $\mathbf{x}$ would then obviously not be observed amongst treated individuals, so finding the counterfactual for it is a moot point. Importantly, however, we need to (at least in theory) be able to observe a control group outcome for each treated outcome, for any setting of $\mathbf{x}$, so not all individuals with this setting can be in the treated group.

Identification

Using (2), Assumption [2], and the notation from (6), we have

$$\begin{aligned}
&E\left(y|\mathbf{x},w\right)=E\left(y_0|\mathbf{x},w\right)+w\left(E\left(y_1|\mathbf{x},w\right)-E\left(y_0|\mathbf{x},w\right)\right)=\\
&E\left(y_0|\mathbf{x}\right)+w\left(E\left(y_1|\mathbf{x}\right)-E\left(y_0|\mathbf{x}\right)\right)=\\
&\mu_0\left(\mathbf{x}\right)+w\left(\mu_1\left(\mathbf{x}\right)-\mu_0\left(\mathbf{x}\right)\right)
\end{aligned} \tag{13}$$

It follows immediately that

$$\begin{aligned}
E\left(y|\mathbf{x},w=0\right)=E\left(y_0|\mathbf{x},w=0\right)=\mu_0\left(\mathbf{x}\right)\\
E\left(y|\mathbf{x},w=1\right)=E\left(y_1|\mathbf{x},w=1\right)=\mu_1\left(\mathbf{x}\right)
\end{aligned} \tag{14}$$

Given a sample of $(y,\mathbf{x},w)$ we can estimate these quantities via the sample means or an *estimable regression function* $m_0\left(\mathbf{x}\right)=E\left(y|\mathbf{x},w=0\right)$ and $m_1\left(\mathbf{x}\right)=E\left(y|\mathbf{x},w=1\right)$. Specifically, if Assumptions [2] and [3] hold, we have:

$$\begin{aligned}
&\tau_{ate}\left(\mathbf{x}\right)=m_1\left(\mathbf{x}\right)-m_0\left(\mathbf{x}\right), \quad \text{and}\\
&\tau_{ate}=E_x\left(m_1\left(\mathbf{x}\right)-m_0\left(\mathbf{x}\right)\right)=E_x\left(\tau_{ate}\left(\mathbf{x}\right)\right)
\end{aligned} \tag{15}$$

The first equation states that we can estimate the ATE for a sub-population with a specific $\mathbf{x}$ via the difference in sample means (or predictions from a regression model) between outcomes for the treated and the control group, where the sample means (or regression models) are based on all individuals that have this exact setting of $\mathbf{x}$. In reality, $\mathbf{x}$ this may be represented by just a few samples points. At the minimum, we need as many sample points per treatment as the dimension of $\mathbf{x}$ to estimate $\tau_{ate}\left(\mathbf{x}\right)$.

The second equation states that we can then obtain the *unconditional* ATE by averaging over the conditional ATE's based on the distribution of $\mathbf{x}$ in the sample.

Using the milder Assumption [4] in addition to Assumption [2], we can derive an estimate for ATT via

$$\begin{aligned} &\tau_{att}\left(\mathbf{x}\right) = m_1\left(\mathbf{x}\right) - m_0\left(\mathbf{x}\right) = \tau_{ate}\left(\mathbf{x}\right), \quad \text{and} \\ &\tau_{att} = E_x\left(m_1\left(\mathbf{x}\right) - m_0\left(\mathbf{x}\right)|w=1\right) \end{aligned} \tag{16}$$

Identification using the propensity score. As shown in Wooldridge (2010), section 21.3.1, we can also identify the conditional and unconditional ATE and ATT using the propensity score $p\left(\mathbf{x}\right) = p\left(w=1|\mathbf{x}\right)$. The results are as follows (maintaining Assumptions [2] and [3] for the ATE, and Assumptions [2] and [4] for the ATT):

$$\begin{aligned} \tau_{ate}\left(\mathbf{x}\right) &= E_{w,y}\left(\frac{\left(w-p\left(\mathbf{x}\right)\right)y}{p\left(\mathbf{x}\right)\left(1-p\left(\mathbf{x}\right)\right)}|\mathbf{x}\right) \\ \tau_{ate} &= E_{w,y,\mathbf{x}}\left(\frac{\left(w-p\left(\mathbf{x}\right)\right)y}{p\left(\mathbf{x}\right)\left(1-p\left(\mathbf{x}\right)\right)}\right) \\ \tau_{att}\left(\mathbf{x}\right) &= E_{w,y}\left(\frac{\left(w-p\left(\mathbf{x}\right)\right)y}{\rho\left(1-p\left(\mathbf{x}\right)\right)}|\mathbf{x}\right) \\ \tau_{att} &= E_{w,y,\mathbf{x}}\left(\frac{\left(w-p\left(\mathbf{x}\right)\right)y}{\rho\left(1-p\left(\mathbf{x}\right)\right)}\right) \end{aligned} \tag{17}$$

where $\rho = p\left(w=1\right)$, the unconditional probability of ending up in the treatment group. Identification via the propensity score can be convenient in practice when the dimension of $\mathbf{x}$ is large, which could make a regression-based approach difficult to implement.

Estimation. Under the assumption of ignorability and overlap, there exist *three general approaches* to estimating treatment effects:

(1) regression-based methods
(2) propensity score methods
(3) matching methods

There also exist various combinations of these three general strategies. For example, regression or matching approaches often utilize the propensity score. We will discuss each approach in turn in the next set of lecture notes.

References

Angrist, J. D. and Pischke, J. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics, *Journal of Economic Perspectives* **24**: 3–30.

Greene, W. (2012). *Econometric Analysis*, 7th edn, Pearson / Prentice Hall.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*, MIT Press.