COMPLIER AVERAGE CAUSAL EFFECT

In this lesson, we examine the case where both the instrument, Z, and the intermediate outcome, M, are binary.

We assume the exclusion restriction holds. i.e.

$$Y_i(0, m) = Y_i(1, m) \text{ for all } i \text{ and } m \in \{0, 1\}.$$

This allows us to drop the first argument of the potential outcomes and write $Y_i(z, M_i(z))$ as simply $Y_i(M_i(z))$.

To introduce moment based estimation, we will also assume

1. treatment assignment is unconfounded unconditionally. i.e.

$$M_i(0), M_i(1), Y_i(0), Y_i(1) \perp\!\!\!\perp Z_i$$

2. the average effect of the instrument Z_i on M_i is non-zero. i.e.

$$E(M_i \mid Z_i = 1) - E(M_i \mid Z_i = 0) \neq 0$$

COMPLIER AVERAGE CAUSAL EFFECT

The unconfoundedness assumption implies

$$\tau^c = E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0) = E(Y_i(M_i(1)) - Y_i(M_i(0))).$$

The exclusion restriction implies

$$Y_i(M_i(1)) - Y_i(M_i(0)) = 0 \text{ if } M_i(0) = M_i(1).$$

Together,

$$\begin{aligned}\tau^{(c)} =& E(Y_i(1) - Y_i(0) \mid M_i(0) = 0, M_i(1) = 1) \times \Pr(M_i(0) = 0, M_i(1) = 1) + \\ & E(Y_i(0) - Y_i(1) \mid M_i(0) = 1, M_i(1) = 0) \times \Pr(M_i(0) = 1, M_i(1) = 0).\end{aligned}$$

COMPLIER AVERAGE CAUSAL EFFECT

Consider an experiment with non-compliance. $Z_i = 0$ (1) for assignment to the control (treatment) group, and $M_i = 1$ ($M_i = 0$) for receiving (not receiving) treatment.

Assume

$$\Pr(M_i(0) = 1, M_i(1) = 0) = 0.$$

This is an assumption of no "defiers". Defiers are subjects who would not take up treatment when assigned to treatment but would take up treatment when not assigned.

The no defiers assumption implies
$E(M_i(1) - M_i(0)) = \Pr(M_i(0) = 0, M_i(1) = 1)$. Since the instrument affects M by assumption, the IV estimand

$$\frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(M_i \mid Z_i = 1) - E(M_i \mid Z_i = 0)} = E(Y_i(1) - Y_i(0) \mid M_i(0) = 0, M_i(1) = 1).$$

COMPLIER AVERAGE CAUSAL EFFECT

The assumption Z affects M is testable. The estimand is the average effect in the subpopulation $M_i(0) = 0, M_i(1) = 1$. This is the subpopulation of "compliers", and the estimand is the complier average causal effect (CACE) (or local average treatment effect (LATE)).

Assume the observations $i = 1, ..., n$ are independent and identically distributed Y, M, Z. The IV estimand can be estimated using sample means: $\frac{\bar{Y}_1 - \bar{Y}_0}{\bar{M}_1 - \bar{M}_0}$.

More generally, the CACE can be estimated within levels of covariates $\mathbf{X}$. If $\mathbf{X}$ is discrete, the sample means (within covariate classes) can be used. If $\mathbf{X}$ is continuous or has many levels, it will be necessary to introduce additional modeling assumptions.

COMPLIER AVERAGE CAUSAL EFFECT

It has been argued that for a given intervention Z, the policy relevant question is the overall effect $\tau^{(c)}$, called the intent to treat (ITT) estimand. The CACE only pertains to the subset of units who will comply with whatever assignment they are given.

e.g. Let Z denote a medical treatment that is to be taken in an unsupervised environment. The practical effectiveness of the treatment is the effect in the context in which the treatment is delivered.

n.b. the CACE is not the average effect for all those who would take up treatment if offered. The experiment is not informative about units who take up treatment regardless of their assignment. i.e. $M_i(1) = M_i(0) = 1$).

Even if one wanted to target the treatment toward units who might benefit from being offered treatment, the compliers are a subpopulation that cannot be observed. We cannot observe both potential outcomes $M_i(0)$ and $M_i(1)$.

COMPLIER AVERAGE CAUSAL EFFECT

When there are no defiers, the ATE is a mixture of the average treatment effects among three subgroups:

1. compliers
2. always takers i.e. units with $M_i(0) = M_i(1) = 1$
3. never takers i.e. units with $M_i(0) = M_i(1) = 0$.

If the ATE is the same for compliers and always takers, this gives the effect of treatment on the treated (ATT). If this is also true for the never takers, the complier average causal effect is also the ATE.

COMPLIER AVERAGE CAUSAL EFFECT

It is common to see empirical work where the ATT or ATE seems to be the parameter of substantive interest but where the CACE is estimated and extrapolated without further argument or consideration to the parameter of interest.

When the compliers constitute a large majority of the population, this may be less problematic.

In the Angrist, Imbens, Rubin (1996) paper, the question of interest is the effect of military service on excess civilian mortality. Compliers are approximately 16% of the population. It may be reasonable that never takers are unhealthier than compliers or always takers and that the mortality rate would have been higher had it been possible to take into account the mortality of this sub-population.

COMPLIER AVERAGE CAUSAL EFFECT

We now take confounders, $\mathbf{X}$, into consideration. It will not be possible to proceed non-parametrically as above, unless the $\mathbf{X}$ are discrete and with few levels.

Suppose we observe a sample from random vector $(Y, M, Z, \mathbf{X})$, and the unconfoundedness assumption holds conditional on $\mathbf{X}$.

Let $f_{M(z)=m}(y \mid \mathbf{x})$ denote the observable conditional distribution of the outcome Y among subjects with response $m \in \{0, 1\}$ to $z \in \{0, 1\}$.

COMPLIER AVERAGE CAUSAL EFFECT

The monotonicity condition $M(1) \geq M(0)$ implies that subjects with $M(0) = 1$ are always-takers (A) i.e.

$$f_{M(0)=1}(y \mid \mathbf{x}) = f_{A0}(y \mid \mathbf{x}),$$

where $f_{A0}(y \mid \mathbf{x})$ is the distribution of always takers when $z = 0$.

Similarly, subjects with $M(1) = 0$ are never takers (N) i.e.

$$f_{M(1)=0}(y \mid \mathbf{x}) = f_{N1}(y \mid \mathbf{x}),$$

where $f_{N1}(y \mid \mathbf{x})$ is the distribution of never takers when $z = 1$.

COMPLIER AVERAGE CAUSAL EFFECT

The remaining distributions are mixtures:

$$f_{M(0)=0}(y \mid \mathbf{x}) = \pi_C(\mathbf{x}) f_{C0}(y \mid \mathbf{x}) + \pi_N(\mathbf{x}) f_{N0}(y \mid \mathbf{x}),$$
$$f_{M(1)=1}(y \mid \mathbf{x}) = \pi_C(\mathbf{x}) f_{C1}(y \mid \mathbf{x}) + \pi_A(\mathbf{x}) f_{A1}(y \mid \mathbf{x}),$$

where

$\pi_C(\mathbf{x})$, $\pi_N(\mathbf{x})$, $\pi_A(\mathbf{x})$ are the conditional complier, never taker and always taker probabilities, respectively

$f_{C0}(y \mid \mathbf{x})$ and $f_{C1}(y \mid \mathbf{x})$ are the complier distributions when $z = 0$ and $z = 1$, respectively

$f_{N0}(y \mid \mathbf{x})$, $f_{N1}(y \mid \mathbf{x})$, $f_{A0}(y \mid \mathbf{x})$, and $f_{A1}(y \mid \mathbf{x})$ are defined similarly.

These can be identified (in some instances) under additional assumptions.

COMPLIER AVERAGE CAUSAL EFFECT

e.g. suppose the outcome Y is continuous and the distributions

$$f_{M(0)=0}(y \mid \mathbf{x}) = \pi_C(\mathbf{x}) f_{C0}(y \mid \mathbf{x}) + \pi_N(\mathbf{x}) f_{N0}(y \mid \mathbf{x}),$$
$$f_{M(1)=1}(y \mid \mathbf{x}) = \pi_C(\mathbf{x}) f_{C1}(y \mid \mathbf{x}) + \pi_A(\mathbf{x}) f_{A1}(y \mid \mathbf{x}),$$

are assumed to be mixtures of normal distributions with means $\mu_{C0}(\mathbf{X})$, $\mu_{C1}(\mathbf{X})$, $\mu_{N0}(\mathbf{X})$, $\mu_{N1}(\mathbf{X})$, $\mu_{A0}(\mathbf{X})$, $\mu_{A1}(\mathbf{X})$ and common variance σ^2, and the mixing probabilities follow a multinomial logit model

$$\pi_C(\mathbf{x}) = \frac{\exp(\beta_C' \mathbf{x})}{1 + \exp(\beta_C' \mathbf{x}) + \exp(\beta_A' \mathbf{x})},$$
$$\pi_A(\mathbf{x}) = \frac{\exp(\beta_A' \mathbf{x})}{1 + \exp(\beta_C' \mathbf{x}) + \exp(\beta_A' \mathbf{x})}$$

COMPLIER AVERAGE CAUSAL EFFECT

The assumption of a common variance can be relaxed, but this can create difficulties.

An assumption that is commonly made is that the distributions of never takers and always takers are the same for $z = 0$ and $z = 1$. This is consistent with the exclusion restriction we made in conjunction with the discussion of instrumental variables. n.b. this assumption does not require potential outcomes $Y(z, m)$ to be well defined.

In the common situation where the treatment cannot be accessed outside of the control group, there are no always takers and the analysis above reduces accordingly. See Little and Yau (1998).