Causality and the Potential Outcomes Model

Giselle Montamat

Harvard University

Spring 2020

A road map

Two components of econometrics:

1. Identification
2. Estimation, inference

Model

Identifying assumptions $\Downarrow \Uparrow$ *(1) Identification*

Population distribution of observable variables

Sampling $\Downarrow \Uparrow$ *(2) Estimation, Inference*

Observations

- Model: underlying structure that details relationships between variables (these could be causal relationships, based on some definition of causality).
- Identifying assumptions: further assumptions about the joint distribution of variables.

A road map

1. Identification:
   - Learning about underlying structures (e.g. a causal effect) from a population distribution (e.g. an expectation)
   - What could one learn from "ideal" data? (aka, if we have an infinitely large sample/the population data/if we know the distribution)
   - To "identify" (1): Take an object from the underlying structure (e.g. a causal effect) $\rightarrow$ can one write it as a function of the moments (e.g. expectation, variance) of the distribution of the data, that is, of the distribution of variables that one can get a sample from?
   - What these moments can identify depends on model's assumptions and other identifying assumptions
   - To "identify" (2): how do we back out parameters of a structural object (aka, a model parameter) given knowledge of the population joint distribution of observable variables?
2. Estimation, inference:
   - Learning about a population distribution from a finite number of observations.

More formally (as seen in Lecture 1):

$$\text{Model: } \{F(\theta) : \theta \in \Theta\}$$

$$\textit{Identifying assumptions} \Downarrow \Uparrow \textit{(1) Identification of } \theta$$

$$\text{Population distribution of observable variables: } D \sim F$$

θ is point-identified (based on observing D) if the mapping $\theta \to F(\theta)$ is one-to-one. In other words, if for every possible distribution F for D, $\theta \in \Theta : F(\theta) = F\}$ contains at most one element.

Simple example: Potential Outcomes Model (with binary treatment)

Model: how a certain amount of T affects outcome Y for individual i.

$$Y_i = Y_i(T_i) = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

Can think as: $Y_i = Y_i(T_i) \equiv h(T_i, U_i)$, where $U_i = (Y_i(0), Y_i(1))$ captures all other determinants of Y_i.

Implicitly imposes assumptions "SUTVA":

- Potential outcomes for any unit do not vary with the treatments assigned to other units
- No hidden versions of the treatment (i.e., no hidden quality differences in treatment that is missed by the treatment measure T)

Simple example: Potential Outcomes Model

Define **"causal effect" or "treatment effect" (TE)** for individual i as:

$$TE_i \equiv Y_i(1) - Y_i(0)$$

Note 1: "causality" defined in terms of "potential outcomes"
Note 2: TE may be heterogeneous!
Note 3: Based on the model, observed outcome is:

$$Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$$

$$Y_i = \underbrace{Y_i(0)}_{\text{Baseline}} + \underbrace{[Y_i(1) - Y_i(0)]}_{\text{Causal effect}} T_i$$

Define **"average treatment effect" (ATE)** as:

$$ATE \equiv E[Y_i(1) - Y_i(0)]$$

$\Rightarrow$ TE and ATE are examples of structural objects that we may wish to identify.

Simple example: Potential Outcomes Model

Identifying assumptions: how is treatment assigned?
(In other words, how do $Y_i(1)$ and $Y_i(0)$ relate to T_i?)

Examples:

1) T is randomly assigned: $\{Y_i(1), Y_i(0)\} \perp T_i$
2) T is *not* randomly assigned: $\{Y_i(1), Y_i(0)\} \not\perp T_i$
3) T is not randomly assigned but there is random assignment of an instrument Z: $\{Y_i(1), Y_i(0), T_i(1), T_i(0)\} \perp Z_i$
4) T is randomly assigned conditional on a set of observable characteristics X: $\{Y_i(1), Y_i(0)\} \perp T_i \mid X_i$

Model with identifying assumptions lead to data: $D_i = (Y_i, T_i)$ (or $D_i = (Y_i, T_i, Z_i)$, or $D_i = (Y_i, T_i, X_i)$)

Key idea: we *don't* observe both $Y_i(0)$ *and* $Y_i(1)$. The outcome that we observe (Y_i) is one or the other depending on whether treatment was assigned to i ($T_i = 1$) or not ($T_i = 0$). Hence, we don't observe TE nor ATE.

Simple example: Potential Outcomes Model

Identification. Example: how can we write ATE (which we do not observe) as a function of moments of the distribution of D_i (which we do observe)? In other words, can we (point) identify ATE?

1) If random assignment of T:

$$\begin{aligned} ATE &\equiv E[Y_i(1) - Y_i(0)] \\ &= E[Y_i(1)] - E[Y_i(0)] \\ &=^{*} E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] \\ &= E[Y_i|T_i = 1] - E[Y_i|T_i = 0] \end{aligned}$$

Note 1: * uses identifying assumption of independence $\{Y_i(1), Y_i(0)\} \perp T_i$. The trick is to know what value of T_i to condition on.

Note 2: $E[Y_i|T_i]$ can be estimated from data since both Y_i and T_i are observed.

Simple example: Potential Outcomes Model

A brief disgression back to ECON 2120...

Conditional expectation function (CEF):

$$E[Y_i|T_i] = \underset{m(T_i)}{argmin}\ E[(Y_i - m(T_i))^2]$$

$$E[m(T_i)(Y_i - E[Y_i|T_i])] = 0\ \forall\ m(.)$$

(i.e, minimize over all possible functions $m(T_i)$; CEF is orthogonal projection over space of all functions $m(.)$)

Best linear predictor (BLP):

$$E^*[Y_i|T_i] \equiv T_i'\beta = T_i' \times \underset{b}{argmin}\ E[(Y_i - T_i'b)^2]$$

$$E[l(T_i)(Y_i - T_i'\beta)] = 0\ \forall\ l(.) \text{ linear}$$

(i.e, minimize over linear functions of T_i; BLP is orthogonal projection over space of linear functions $l(.)$)

Simple example: Potential Outcomes Model

Note: In 2120 you called the CEF the "regression function". If you have read Mostly Harmless Econometrics, the BLP is called the "Population Regression Function"...yes, I know, very confusing.

A useful relationship between CEF and BLP to remember:

- If CEF is a linear function, then it coincides with the BLP
- Examples of cases when CEF is linear:
  - If (Y, T) has a multivariate normal distribution, then $Y|T$ has a normal distribution with $E[Y|T]$ linear in T

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right)$$

$$Z_1|Z_2 \sim N(\underbrace{\mu_1 - \sigma_{12}\sigma_{22}^{-1}(Z_2 - \mu_2)}_{\equiv E[Z_1|Z_2]}, \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21})$$

  - If T is discrete (CEF is linear in T if T is dummy, or it is linear in appropriately defined transformations of T if T not dummy).

Simple example: Potential Outcomes Model

Example: T_i binary.

$$\begin{aligned}
\text{CEF: } E[Y_i|T_i] &= E[Y_i|T_i=1]\mathbb{I}_{(T_i=1)} + E[Y_i|T_i=0]\mathbb{I}_{(T_i=0)} \\
&= E[Y_i|T_i=1]T_i + E[Y_i|T_i=0](1-T_i) \\
&= \underbrace{E[Y_i|T_i=0]}_{\equiv\delta_0} + \underbrace{(E[Y_i|T_i=1] - E[Y_i|T_i=0])}_{\equiv\delta_1} T_i
\end{aligned}$$

$$\text{BLP: } E^*[Y_i|1, T_i] = \beta_0 + \beta_1 T_i$$

$$\beta_0 = E[Y_i] - \beta_1 E[T_i]$$

$$\beta_1 = \frac{Cov(Y_i, T_i)}{Var(T_i)}$$

Simple example: Potential Outcomes Model

Exercise: show that $\delta_1 = \beta_1$ and $\delta_0 = \beta_0$.

Simple example: Potential Outcomes Model

Example: $T \in \{t_1, ..., t_K\}$. Define $X_k = \mathbb{I}_{(T=t_k)}$ (one dummy per value that T can take).

$$\begin{aligned} \text{CEF: } E[Y|T] &= E[Y|T=t_1]\mathbb{I}_{(T=t_1)} + ... + E[Y|T=t_K]\mathbb{I}_{(T=t_K)} \\ &= E[Y|T=t_1]X_1 + ... + E[Y|T=t_K]X_K \end{aligned}$$

So we conclude the CEF is linear in $X_1, ..., X_K$. Then:

$$E[Y|T] = E^*[Y|X_1, ..., X_K]$$

Note 1: T is a particular case of this one in which $T \in \{0, 1\}$ so we have $X_1 = \mathbb{I}_{(T=0)} = (1 - T)$ and $X_2 = \mathbb{I}_{(T=1)} = T$.

Note 2: if more than one regressor (say $\{T_1, ..., T_K\}$), define a dummy for every possible combination of values that the set of regressors can take.

...now let's go back to ECON 2140.

Simple example: Potential Outcomes Model

Exercise: show that, under random assignment of T, one can also point identify (in addition to ATE):

- the average treatment effect on the treated (ATET)

$$\begin{aligned} ATET &\equiv E[Y_i(1) - Y_i(0)|T_i = 1] \\ &= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 1] \\ &= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] \\ &= E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = ATE \end{aligned}$$

- the average treatment effect on the untreated (ATEU)

$$\begin{aligned} ATEU &\equiv E[Y_i(1) - Y_i(0)|T_i = 0] \\ &= E[Y_i(1)|T_i = 0] - E[Y_i(0)|T_i = 0] \\ &= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] \\ &= E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = ATE \end{aligned}$$

- the conditional average treatment effect (CATE)

Simple example: Potential Outcomes Model

Exercise: show that, under random assignment of T, one can also point identify (in addition to ATE):

- the marginal distribution of potential outcomes $Y(0)$ and of $Y(1)$

$$F_{Y(0)}(y) = \underbrace{P(Y_i(0) \leq y)}_{\text{This is the dist of sth you } \textbf{don't} \text{ observe}} = P(Y_i(0) \leq y | T_i = 0) = \underbrace{P(Y_i \leq y | T_i = 0)}_{\text{This is the dist of sth you } \textbf{do} \text{ observe}}$$

$$F_{Y(1)}(y) = P(Y_i(1) \leq y) = P(Y_i(1) \leq y | T_i = 1) = P(Y_i \leq y | T_i = 1)$$

Example: compare fraction of poor people when treatment is assigned vs fraction of poor people when treatment is not assigned.

- the quantile treatment effect (QTE)

$$F^{-1}_{Y(1)}(\tau) = inf\{y : P(Y_i(1) \leq y) \geq \tau\} = inf\{y : P(Y_i \leq y | T_i = 1) \geq \tau\}$$

$$F^{-1}_{Y(0)}(\tau) = inf\{y : P(Y_i(0) \leq y) \geq \tau\} = inf\{y : P(Y_i \leq y | T_i = 0) \geq \tau\}$$

Example: compare the median person that is treated with the median person that is not treated.

Simple example: Potential Outcomes Model

Isaiah's remark: identifying the *treatment effect over the distribution of outcomes* $(F_{Y(1)}(y) - F_{Y(0)}(y);\ F^{-1}_{Y(1)}(\tau) - F^{-1}_{Y(0)}(\tau))$ is different than identifying the *distribution of the treatment effect* $(F_{Y(1)-Y(0)}(y);\ F^{-1}_{Y(1)-Y(0)}(\tau))$.

So far, we have only shown point identification of the mean of $F_{Y(1)-Y(0)}(y) = P(Y_i(1) - Y_i(0) \leq y)$ (aka, the ATE). Can we do more?

Exercise: Problem Set 1, Problem 2 - Find upper and lower bounds on $F_{Y(1)-Y(0)}(y)$

Note: next step is to find estimates for the expectations and probabilities that identify the causal objects of interest. Useful trick for probabilities: remember $P(Y_i \leq y) = E[\mathbb{I}_{(Y_i \leq y)}]$.

Simple example: Potential Outcomes Model

Identification. (Cont.)

2) If T **not** randomly assigned:

$E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$ no longer identifies ATE

$$
\begin{aligned}
&= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] \\
&= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] + E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 1] \\
&= \underbrace{E[Y_i(1) - Y_i(0)|T_i = 1]}_{ATET} + \underbrace{E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 0]}_{\text{Selection Bias}}
\end{aligned}
$$

Example: suppose treatment is "to be hospitalized" and outcome is "health". Selection bias is the difference in average "baseline" health ($Y(0)$) between those who are and those who aren't hospitalized. If the sick are more likely than the healthy to get treatment, then those who are hospitalized have worse baseline values of health (i.e., of $Y(0)$), making selection bias negative so that $E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$ understates the causal effect of treatment on treated.

In other words, $E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$ is not causal unless we impose certain identifying assumptions.

Simple example: Potential Outcomes Model

Exercise. Provide the smallest possible bounds on ATE when potential outcomes can take on only two values, 0 and 1. (Seen in ECON 2120).

We know that $Y_i(1),\ Y_i(0) \in \{0,1\}$. So $-1 \leq Y_i(1) - Y_i(0) \leq 1 \Rightarrow$ length of this interval is 2. Can we do better (aka, tighter) for the mean of $Y_i(1) - Y_i(0)$ (aka, ATE)? Yes!

Show:

$$\underbrace{E[Y_i(1)|T_i=1]P(T_i=1)}_{\text{when } P(Y_i(1)=1|T_i=0)=0} \leq E[Y_i(1)] \leq \underbrace{E[Y_i(1)|T_i=1]P(T_i=1)+P(T_i=0)}_{\text{when } P(Y_i(1)=1|T_i=0)=1}$$

$$E[Y_i(1)] = E[E[Y_i(1)|T_i]] =$$

$$= \underbrace{E[Y_i(1)|T_i=1]}_{\text{indentified}}\underbrace{P(T_i=1)}_{\text{identified}} + \underbrace{E[Y_i(1)|T_i=0]}_{\text{not identified}}\underbrace{P(T_i=0)}_{\textit{identified}}$$

$$= \underbrace{E[Y_i(1)|T_i=1]}_{\text{indentified}}\underbrace{P(T_i=1)}_{\text{identified}} + \underbrace{P(Y_i(1)=1|T_i=0)}_{\in[0,1]}\underbrace{P(T_i=0)}_{\textit{identified}}$$

Simple example: Potential Outcomes Model

Similarly:

$$\underbrace{E[Y_i(0)|T_i=0]P(T_i=0)}_{\text{when } P(Y_i(0)=1|T_i=1)=0} \leq E[Y_i(0)] \leq \underbrace{E[Y_i(0)|T_i=0]P(T_i=0)+P(T_i=1)}_{\text{when } P(Y_i(0)=1|T_i=1)=1}$$

Bounds on ATE are then:

$$E[Y_i(1)-Y_i(0)] \leq E^H[Y_i(1)] - E^L[Y_i(0)] =$$

$$= E[Y_i(1)|T_i=1]P(T_i=1) + P(T_i=0) - E[Y_i(0)|T_i=0]P(T_i=0)$$

$$E[Y_i(1)-Y_i(0)] \geq E^L[Y_i(1)] - E^H[Y_i(0)] =$$

$$= E[Y_i(1)|T_i=1]P(T_i=1) - E[Y_i(0)|T_i=0]P(T_i=0) - P(T_i=1)$$

Length of this interval is 1!

Called "worst-case" bounds because we are under assumption that T is not randomly assigned and we don't know anything more.

Simple example: Potential Outcomes Model

Identification. (Cont.)

3) If T **not** randomly assigned but there's a randomly assigned instrument that affects treatment.

   1. Could simply define Z_i to be the treatment (e.g. "assignment to treatment" is the treatment).
      $\Rightarrow$ Goes back to case 1, where Z_i is taken to be T_i and is randomly assigned.
      $\Rightarrow$ TE called an "intent" to treat effect.

   2. Z_i is an instrument (e.g. "assignment to treatment") and we stick to identifying causal objects of the treatment T_i.

Simple example: Potential Outcomes Model

Additions to model: random instrument $Z_i \in \{0, 1\}$

$$T_i = T_i(Z_i) = \begin{cases} T_i(0) & \text{if } Z_i = 0 \\ T_i(1) & \text{if } Z_i = 1 \end{cases}$$

Implicitly imposes:

- SUTVA: treatment of i is not affected by the instrument values of other units.
- Exclusion restriction: Z_i does not affect Y_i directly. That is, we still have that observed outcome is $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$.

Note: treatment can be written:

$$T_i = T_i(1)Z_i + T_i(0)(1 - Z_i)$$

Simple example: Potential Outcomes Model

Example: Angrist (1990).

- T: military service $\rightarrow$ not random assignment
- Z: draft eligibility $\rightarrow$ random assignment

Identifying assumptions:

- Random assignment/independence: $\{Y_i(1), Y_i(0), T_i(1), T_i(0)\} \perp Z_i$
- Monotonicity/no defiers: $T_i(1) \geq T_i(0)$
- Relevance/first stage: $P(T_i(1) \neq T_i(0)) > 0$ (aka, instruments are "strong", meaning the instrument affects treatment)

Note: 4 possible combinations of $(T_i(0), T_i(1))$:

1. Compliers: $(T_i(0) = 0, T_i(1) = 1) \Rightarrow T_i(1) > T_i(0)$
2. Always takers: $(T_i(0) = 1, T_i(1) = 1) \Rightarrow T_i(1) = T_i(0)$
3. Never takers: $(T_i(0) = 0, T_i(1) = 0) \Rightarrow T_i(1) = T_i(0)$
4. Defiers: $(T_i(0) = 1, T_i(1) = 0) \Rightarrow T_i(1) < T_i(0)$

Simple example: Potential Outcomes Model

Exercise: show that the relevance assumption can be checked with the data (aka, identify $P(T_i(1) \neq T_i(0)) > 0$).

Monotonicity implies: $P(T_i(1) \neq T_i(0)) > 0 \Leftrightarrow P(T_i(1) > T_i(0)) > 0 \Leftrightarrow P(T_i(1) - T_i(0) = 1) > 0 \Leftrightarrow E[T_i(1) - T_i(0)] > 0 \Leftrightarrow E[T_i(1)] > E[T_i(0)]$

Independence implies: $E[T_i(1)] > E[T_i(0)] \Leftrightarrow E[T_i(1)|Z_i = 1] > E[T_i(0)|Z_i = 0] \Leftrightarrow E[T_i|Z_i = 1] - E[T_i|Z_i = 0] > 0$

Exercise: show that $Cov(T_i, Z_i) > 0 \Leftrightarrow E[T_i|Z_i = 1] - E[T_i|Z_i = 0] > 0$

$$
\begin{aligned}
Cov(T_i, Z_i) &> 0 \\
E[T_i Z_i] - E[T_i]E[Z_i] &> 0 \\
E[E[T_i Z_i|Z_i]] - E[E[T_i|Z_i]]E[Z_i] &> 0 \\
E[T_i|Z_i = 1]P(Z_i = 1) - (E[T_i|Z_i = 1]P(Z_i = 1) + E[T_i|Z_i = 0]P(Z_i = 0))\, E[Z_i] &> 0 \\
E[T_i|Z_i = 1]E[Z_i] - (E[T_i|Z_i = 1]E[Z_i] + E[T_i|Z_i = 0](1 - E[Z_i]))\, E[Z_i] &> 0 \\
E[T_i|Z_i = 1] - (E[T_i|Z_i = 1]E[Z_i] + E[T_i|Z_i = 0](1 - E[Z_i])) &> 0 \\
E[T_i|Z_i = 1](1 - E[Z_i]) - E[T_i|Z_i = 0](1 - E[Z_i]) &> 0 \\
E[T_i|Z_i = 1] - E[T_i|Z_i = 0] &> 0
\end{aligned}
$$

Simple example: Potential Outcomes Model

Define **"local average treatment effect" (LATE)** as:

$$LATE \equiv E[Y_i(1) - Y_i(0) | T_i(1) > T_i(0)]$$

Note: this is an average treatment effect for a "local" group, namely, the compliers.

$\Rightarrow$ LATE is a structural object that we can identify.

Simple example: Potential Outcomes Model

Exercise: show that LATE is identified by $\frac{Cov(Y_i,Z_i)}{Cov(Y_i,T_i)}$

$$
\begin{aligned}
\frac{Cov(Y_i, Z_i)}{Cov(Y_i, T_i)} &= \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[T_i|Z_i = 1] - E[T_i|Z_i = 0]} \\
&=^{*} \frac{E[Y_i(1)T_i + Y_i(0)(1 - T_i)|Z_i = 1] - E[Y_i(1)T_i + Y_i(0)(1 - T_i)|Z_i = 0]}{E[T_i|Z_i = 1] - E[T_i|Z_i = 0]} \\
&= \frac{E[Y_i(1)T_i(1) + Y_i(0)(1 - T_i(1))|Z_i = 1] - E[Y_i(1)T_i(0) + Y_i(0)(1 - T_i(0))|Z_i = 0]}{E[T_i(1)|Z_i = 1] - E[T_i(0)|Z_i = 0]} \\
&=^{**} \frac{E[Y_i(1)T_i(1) + Y_i(0)(1 - T_i(1))] - E[Y_i(1)T_i(0) + Y_i(0)(1 - T_i(0))]}{E[T_i(1)] - E[T_i(0)]} \\
&= \frac{E[(Y_i(1) - Y_i(0))(T_i(1) - T_i(0))]}{E[T_i(1) - T_i(0)]} \\
&= \frac{E[E[(Y_i(1) - Y_i(0))(T_i(1) - T_i(0))|T_i(1) - T_i(0)]]}{E[T_i(1) - T_i(0)]} \\
&=^{***} \frac{E[Y_i(1) - Y_i(0)|T_i(1) - T_i(0) = 1]P(T_i(1) - T_i(0) = 1)}{P(T_i(1) - T_i(0) = 1)} \\
&= E[Y_i(1) - Y_i(0)|T_i(1) > T_i(0)]
\end{aligned}
$$

Note: * uses *exclusion restriction*, ** uses *random assignment of instrument*, *** uses *no defiers*; the denominator $E[T_i(1) - T_i(0)]$ is different from 0 because of *relevance*.

Simple example: Potential Outcomes Model

Note: we ruled out defiers by assumption (monotonicity) and we were able to identify a causal effect for compliers. We can't identify who is or not a complier, however, we can identify features of the distribution of covariates for compliers, as in the following exercise.

Exercise (practice at home): show that $E[g(X_i)|T_i(1) > T_i(0)]$ is identified by:

$$E[g(X_i)|T_i(1) > T_i(0)] = \frac{E[g(X_i)T_i|Z_i = 1] - E[g(X_i)T_i|Z_i = 0]}{E[T_i|Z_i = 1] - E[T_i|Z_i = 0]}$$

Exercise (practice at home): show that, under random assignment of instrument Z, we can identify:

- the local average treatment effect conditional on covariates (CLATE)
- the marginal distribution of potential outcomes for compliers

Simple example: Potential Outcomes Model

Summing up, in this example:

- A causal relationship describes what would happen to a given i in a hypothetical comparison of two different scenarios (one is a counterfactual).
- Identification: identify an average causal effect (a structural object) by a difference in expectations / a ratio of covariances (of observed variables) if we have a randomized experiment (of the treatment / of an instrument).
- Estimation: estimate expectations using sample means.

Other examples on identification

- From Lecture 1:
  - Linear regression model

    Model: $Y_i = X_i'\beta + \epsilon_i \;\; ; \;\; E[\epsilon_i|X_i] = 0$

    Data: (X_i, Y_i)

    Structural object to identify: β

    Identifying assumption: $E[X_i X_i']$ has full rank (no multicollinearity)

    Result: $\beta = E[X_i X_i']^{-1} E[X_i Y_i']$

  - Binary choice model - Manski (1975)

    Model: $Y_i = 1\{X_i'\beta + \epsilon_i > 0\} \;\; ; \;\; Med(\epsilon_i|X_i) = 0$

    Data: (X_i, Y_i) (observed values of X_i don't include x^*)

    Structural objects to identify: $\frac{\beta}{||\beta||} \;\; ; \;\; E[Y_i|X_i = x^*]$

    Identifying assumptions:

    -$E[X_i X_i']$ has full rank (no multicollinearity);

    -$P(0 < E[Y_i|X_i] < 1) = 1$;

    -at least one element $X_{i,j}$ with support equal to $\mathbb{R}$

    Result: can identify $\frac{\beta}{||\beta||}$ but not $E[Y_i|X_i = x^*]$ (can only say if $\lesseqgtr \frac{1}{2}$)

Other examples on identification

- From Problem Set 1:
  - Problem 1: Measurement Error

    Model: $Y_i = \beta_1 + \beta_2 W_i + \epsilon_i \;\; ; \;\; E[X_i \epsilon_i] = 0 \;\; ; \;\; X_i = [1, W_i]'$

    Data: (X_i^*, Y_i) with $X_i^* = X_i + \eta_i$:

    Structural object to identify: β

    Identifying assumptions:

    -$\eta_i = [0, \eta_{i,2}]'$ (no measurement error in constant);

    -$E[X_i \eta_i'] = 0$;

    -$E[\eta_i \epsilon_i] = 0$

    Results:

    -β_2 not identified: show an example in which a given distribution of observables can be associated to two different values of β_2 (pick a normal because all information is summarized in mean and variance of (X_i^*, Y_i))

    -*sign*(β_2) is identified

    -can identify a lower bound for $|\beta_2|$

Other examples on identification

- From Problem Set 1:
  - Problem 1: Measurement Error

$$\beta_2^* = \frac{Cov(W_i^*, Y_i)}{Var(W_i^*)} = \frac{Var(W_i)}{Var(W_i) + Var(\eta_{i,2})}\beta_2$$

Conclude:

- ★ $\frac{Var(W_i)}{Var(W_i)+Var(\eta_{i,2})} \geq 0 \Rightarrow sign(\beta_2) = sign(\beta_2^*)$
- ★ $0 \leq \frac{Var(W_i)}{Var(W_i)+Var(\eta_{i,2})} \leq 1 \Rightarrow |\beta_2| \geq |\beta_2^*|$

Recap

Model: $Y_i = h(T_i, U_i)$
$\rightarrow$ Structural object, e.g.: $ATE \equiv E[h(t_2, U_i) - h(t_1, U_i)]$

Identification problem:
$\rightarrow$ How do we back out ATE (or other causal objects) given knowledge of the population distribution of observable variables D_i?
$\rightarrow$ What assumptions are needed? (identifying assumptions)

Two types of exercises:

1. I have this object from the pop distribution $\Rightarrow$ what does it identify under such and such assumptions? *Example: Problem Set 2, exercise 1.*
2. I have this object that I wish to identify $\Rightarrow$ what object from the pop distribution can identify it given such and such assumptions?

Estimation/inference problem:
$\rightarrow$ How to come up with estimates/tests of these objects using sampled data set?

Recap: Identification problem - two main cases

1. T_i is randomly assigned, so U_i can be seen as "pre-treatment" personal characteristics for $i \Rightarrow U_i \perp T_i$ (usually with experimental data)

$$E[h(t, U_i)] = E[h(t, U_i)|T_i = t] = E[Y_i|T_i = t]$$

$$ATE = E[Y_i|T_i = t_2] - E[Y_i|T_i = t_1]$$

(See Section 1, 1))

2. T_i is **not** randomly assigned, aka, there is "selection": people with $T_i = t_1$ have systematically different U_i than people with $T_i = t_2 \Rightarrow U_i \not\perp T_i$ (usually with observational data)
(Section 1, case 2))

What to do in order to identify a causal object?

1. Exploit *natural* experiments: treatment or instrument assignment is as good as random. (Section 1, cases 1) and 3))
2. Control for additional observed variables (components of U_i) $\rightarrow$ "selection on observables" or "conditional independence assumption": T_i is as good as randomly assigned conditional on X_i. (Section 1, case 4) - today's section)

Simple example: Potential Outcomes Model

Identification. (Cont.)

4) T randomly assigned if we condition on a set of observables X.
Idea: individuals select into treatment based on observable characteristics; within a group $X = x$ treatment is randomly assigned.

   Called "selection on observables".

   Identifying assumptions:

   - Selection on observables/conditional independence/conditional unconfoundedness given X: $\{Y_i(1), Y_i(0)\} \perp T_i \mid X_i$
   - In every group there are some people treated and some not treated (overlap condition): $P(T_i = 1|X_i = x) = E[T_i|X_i = x] \in (0, 1)$

Note: if $T_i \perp X_i$, then we're back to random assignment of treatment: $\{Y_i(1), Y_i(0)\} \perp T_i$

Simple example: Potential Outcomes Model

Exercise: show that, under random assignment of T conditional on observables X, one can point identify:

- the conditional average treatment effect (CATE)

$$\begin{aligned} CATE &\equiv E[Y_i(1) - Y_i(0)|X_i] \\ &= E[Y_i(1)|X_i] - E[Y_i(0)|X_i] \\ &= E[Y_i(1)|X_i, T_i = 1] - E[Y_i(0)|X_i, T_i = 0] \\ &= E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0] \end{aligned}$$

- the average treatment effect (ATE)

$$\begin{aligned} ATE &\equiv E[Y_i(1) - Y_i(0)] \\ &= E_X[E[Y_i(1) - Y_i(0)|X_i]] \\ &= E_X[E[Y_i(1)|X_i]] - E_X[E[Y_i(0)|X_i]] \\ &= E_X[E[Y_i(1)|X_i, T_i = 1]] - E_X[E[Y_i(0)|X_i, T_i = 0]] \\ &= E_X[E[Y_i|X_i, T_i = 1]] - E_X[E[Y_i|X_i, T_i = 0]] \\ &= E_X[\underbrace{E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0]}_{\text{CATE}}] = E_X[CATE] \end{aligned}$$

Simple example: Potential Outcomes Model

- the average treatment effect on treated (ATET)

$$\begin{aligned} ATET &\equiv E[Y_i(1) - Y_i(0)|\, T_i = 1] \\ &= E[E[Y_i(1) - Y_i(0)|X_i, T_i = 1]|\, T_i = 1] \\ &= E[\underbrace{E[Y_i(1) - Y_i(0)|X_i]}_{\text{CATE}}|\, T_i = 1] = E_{X|\,T=1}[CATE] \end{aligned}$$

- the average treatment effect on untreated (ATEU)

$$\begin{aligned} ATEU &\equiv E[Y_i(1) - Y_i(0)|\, T_i = 0] \\ &= E[E[Y_i(1) - Y_i(0)|X_i, T_i = 0]|\, T_i = 0] \\ &= E[\underbrace{E[Y_i(1) - Y_i(0)|X_i]}_{\text{CATE}}|\, T_i = 0] = E_{X|\,T=0}[CATE] \end{aligned}$$

- the marginal distribution of potential outcomes $Y(0)$ and of $Y(1)$, conditional on X, aka $F_{Y(1)|X}(y)$ and $F_{Y(0)|X}(y)$

Simple example: Potential Outcomes Model

- the marginal distribution of potential outcomes $Y(0)$, conditional on being treated, aka, $F_{Y(0)|T=1}(y)$

$$
\begin{aligned}
F_{Y(0)|T=1}(y) &= P(Y(0) \leq y | T = 1) \\
&= E[\mathbb{I}_{(Y(0)\leq y)} | T = 1] \\
&=^{*} E_{X|T=1}[E[\mathbb{I}_{(Y(0)\leq y)} | T = 1, X]] \\
&= E_{X|T=1}[E[\mathbb{I}_{(Y(0)\leq y)} | T = 0, X]] \\
&=^{**} E_{X|T=1}\left[E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)} | X\right]\right]
\end{aligned}
$$

Note 1: * uses the fact that $Y(0) \perp T|X$.
Note 2: ** uses $E[\mathbb{I}_{(Y\leq y)}(1-T)|X] = E[\mathbb{I}_{(Y\leq y)}|T=0,X](1-p(X))$ (show!)

Simple example: Potential Outcomes Model

We can keep working to find a "nicer" expression with intuitive interpretation:

$$
\begin{aligned}
F_{Y(0)|T=1}(y) &= E_{X|T=1}\left[E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}|X\right]\right] \\
&= \sum_x E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}|X\right] P(X|T=1) \\
&= \sum_x E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}|X\right] \frac{P(T=1|X)P(X)}{P(T=1)} \\
&= \sum_x E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}\frac{p(X)}{P(T=1)}|X\right] P(X) \\
&= E_X\left[E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}\frac{p(X)}{P(T=1)}|X\right]\right] \\
&= E\left[\frac{\mathbb{I}_{(Y\leq y)}(1-T)}{1-p(X)}\frac{p(X)}{P(T=1)}\right] \\
&= E\left[\mathbb{I}_{(Y\leq y)}(1-T)\frac{p(X)}{P(T=1)}\frac{1}{1-p(X)}\right]
\end{aligned}
$$

Note: don't confuse $p(X) \equiv P(T=1|X)$ with $P(X)$!

Simple example: Potential Outcomes Model

Propensity score:

$$p(X_i) = P(T_i = 1|X_i)$$

Note 1: $P(X_i)$ is a random variable that takes on values $\in [0, 1]$; $P(X_i = x)$ is a specific value for this random variable, notably the probability of being treated if characterstic X_i takes on value x.
Note 2: if $T_i \in \{0, 1\}$, then $P(T_i = 1|X_i) = E[T_i|X_i]$

Why useful? For estimation!

- Identification step: instead of conditioning on X_i, condition on $p(X_i)$.
- *Curse of dimensionality*: to justify selection on observables assumption, would want to condition on many covariates $X_1, X_2, ..., X_K$, but then very few observations (or even none!) of both treated *and* untreated people within each group (there are as many groups as combinations of values for the covariates $\{x_1, x_2, ..., x_k\}$).
- Two groups, aka, two combinations $\{x_1, x_2, ..., x_k\}$ and $\{x'_1, x'_2, ..., x'_k\}$, might have the same propensity score:
  $p(X_{i,1} = x_1, ..., X_{i,K} = x_K) = p(X_{i,1} = x'_1, ..., X_{i,K} = x'_K)$, i.e.,
  $P(T_i = 1|X_{i,1} = x_1, ..., X_{i,K} = x_K) = P(T_i = 1|X_{i,1} = x'_1, ..., X_{i,K} = x'_K)$
- So we can pool their observations if instead of conditioning on $X_1, X_2, ..., X_K$, we condition on $p(X_1, X_2, ..., X_K)$

Simple example: Potential Outcomes Model

Exercise: show that one can rewrite the ATE in the following way:

$$ATE \equiv E[Y_i(1)] - E[Y_i(0)] = E\left[Y_i \frac{T_i - p(X_i)}{p(X_i)(1-p(X_i))}\right]$$

We showed before: $ATE \equiv E_X[E[Y_i|X_i, T_i = 1]] - E_X[E[Y_i|X_i, T_i = 0]]$

Show that: $E[Y_i|X_i, T_i = 1] = E\left[\frac{Y_i T_i}{p(X_i)}|X_i\right]$

$$\begin{aligned} E[Y_i T_i|X_i] &= E[E[Y_i T_i|X_i, T_i]|X_i] \\ &= E[T_i E[Y_i|X_i, T_i]|X_i] \\ &= 1 \times E[Y_i|X_i, T_i = 1]P(T_i = 1|X_i) + 0 \times E[Y_i|X_i, T_i = 0]P(T_i = 0|X_i) \\ &= E[Y_i|X_i, T_i = 1]p(X_i) \end{aligned}$$

$$\Rightarrow E[Y_i|X_i, T_i = 1] = \frac{E[Y_i T_i|X_i]}{p(X_i)} = E\left[\frac{Y_i T_i}{p(X_i)}|X_i\right]$$

Show that: $E[Y_i|X_i, T_i = 0] = E\left[\frac{Y_i(1-T_i)}{1-p(X_i)}|X_i\right]$

Simple example: Potential Outcomes Model

Therefore:

$$E[Y(1)] = E_X[E[Y_i|X_i, T_i = 1]] = E\left[\frac{Y_i T_i}{p(X_i)}\right]$$

$$E[Y(0)] = E_X[E[Y_i|X_i, T_i = 0]] = E\left[\frac{Y_i(1 - T_i)}{1 - p(X_i)}\right]$$

Intuition for $E[Y_i(0)] = E\left[\frac{Y_i(1-T_i)}{1-p(X_i)}\right]$:
$Y_i(1 - T_i)$ keeps the observations for those that are not treated, for which we observe the potential outcome under no treatment, $Y_i(0)$. Units that have a small $P(T_i = 0|X_i) = 1 - p(X_i)$ are "under-represented" in these observations, so we upweight them with the inverse of $1 - p(X_i)$.

Last step is algebra:

$$ATE = E\left[\frac{Y_i T_i}{p(X_i)}\right] - E\left[\frac{Y_i(1 - T_i)}{1 - p(X_i)}\right] = E\left[Y_i \frac{T_i - p(X_i)}{p(X_i)(1 - p(X_i))}\right]$$

Note: this exercise was proved slightly differently in lecture using the fact that, by construction, $T_i \perp X_i|p(X_i)$. Check it out!

Simple example: Potential Outcomes Model

There's a fifth case that we talked about in class...

Identification. (Cont.)

5) Z is randomly assigned if we condition on a set of observables X. *Idea:* individuals receive instrument (e.g., get draft letter) based on observable characteristics; within a group $X = x$ instrument is randomly assigned.

   Identifying assumption:

   - Conditional independence of instrument: $\{Y_i(1), Y_i(0), T_i(1), T_i(0)\} \perp Z_i \mid X_i$

Simple example: Potential Outcomes Model

Exercise (practice at home): show that, under random assignment of Z conditional on observables X, one can point identify:

- the conditional local average treatment effect (CLATE)
- the local average treatment effect (LATE)
- features of the distribution of covariates for compliers: $E[g(X_i)|\, T_i(1) > T_i(0)]$

Simple example: Potential Outcomes Model

Notice the analogies between the CATE and ATE under case 4), and the CLATE and LATE under case 5):

$$ATE \equiv E[Y_i(1) - Y_i(0)] = E_X[E[Y_i(1) - Y_i(0)]|X_i] = E_X[CATE]$$

So by identifying CATE, we can also identify ATE.

$$LATE \equiv E[Y_i(1) - Y_i(0)|\, T_i(1) > T_i(0)] =$$

$$= E_X\left[\frac{P(T_i(1) > T_i(0)|X_i)}{E[P(T_i(1) > T_i(0)|X_i)]} E[Y_i(1) - Y_i(0)|\, T_i(1) > T_i(0), X_i]\right] =$$

$$E_X[W(X_i)CLATE]$$

So by identifying CLATE, we can also identify LATE.

Simple example: Potential Outcomes Model

ATE can be re-written with an expression that uses the propensity score:

$$ATE \equiv E[Y_i(1) - Y_i(0)] = E\left[\frac{T_i}{p(X_i)}Y_i\right] - E\left[\frac{(1-T_i)}{1-p(X_i)}Y_i\right]$$

LATE can be re-written:

$$LATE = E\left[\frac{\kappa_i^1}{E[\kappa_i^1]}Y_i\right] - E\left[\frac{\kappa_i^0}{E[\kappa_i^0]}Y_i\right]$$

$$\kappa_i^0 \equiv (1-T_i)\frac{(1-Z_i) - E[1-Z_i|X_i])}{E[1-Z_i|X_i]E[Z_i|X_i]}$$

$$\kappa_i^1 \equiv T_i\frac{Z_i - E[Z_i|X_i])}{E[1-Z_i|X_i]E[Z_i|X_i]}$$

...but this one has less of a clear intuition.

BLP and what it identifies

If the object you're identifying requires you to estimate a CEF, remember:

- We know that CEFs are the same as BLPs (i.e., CEFs are linear) in some special cases (normal distributions; discrete regressors)
- Otherwise, might want to *assume* linearity of CEF
- But if this assumption is wrong, then BLP will be identifying something different than desired.

BLP and what it identifies

Exercise: Problem Set 2, Exercise 1 walked you through an example of conditions under which, in the context of the potential outcomes model with covariates, BLP identifies or not the ATE.

- Define the CEF of Y_i given X_i, for treatment and control groups:

$$g_0(X_i) \equiv E[Y_i | T_i = 0, X_i]$$

$$g_1(X_i) \equiv E[Y_i | T_i = 1, X_i]$$

- Define the BLP of Y_i given X_i, for treatment and control groups:

$$g_0^L(X_i) \equiv E^*[Y_i | T_i = 0, X_i] = X_i' \gamma_0$$

$$g_1^L(X_i) \equiv E^*[Y_i | T_i = 1, X_i] = X_i' \gamma_1$$

Where:

$$\gamma_0 \equiv E[X_i X_i' | T_i = 0]^{-1} E[X_i Y_i | T_i = 0]$$

$$\gamma_1 \equiv E[X_i X_i' | T_i = 1]^{-1} E[X_i Y_i | T_i = 1]$$

BLP and what it identifies

Exercise: Problem Set 2, Exercise 1 walked you through an example of conditions under which, in the context of the potential outcomes model with covariates, BLP identifies or not the ATE.

- Recall that the ATE is the average (wrt to X) of the CATE:

$$ATE = E_X[E[Y_i(1)|X_i] - E[Y_i(0)|X_i]]$$

- If $Y_i(0), Y_i(1) \perp T_i|X_i$, then ATE is identified by:

$$ATE = E_X[\underbrace{E[Y_i|T_i = 1, X_i]}_{\equiv g_1(X_i)} - \underbrace{E[Y_i|T_i = 0, X_i]}_{\equiv g_0(X_i)}]$$

  - If CEF is linear, then it coincides with the BLP: $g_0(X_i) = g_0^L(X_i)$ and $g_1(X_i) = g_1^L(X_i)$, so can use BLP to identify ATE:

$$ATE = E_X[\underbrace{E^*[Y_i|T_i = 1, X_i]}_{\equiv g_1^L(X_i)} - \underbrace{E^*[Y_i|T_i = 0, X_i]}_{\equiv g_0^L(X_i)}]$$

BLP and what it identifies

- If $Y_i(0), Y_i(1) \perp T_i | X_i$, then ATE is identified by:

$$ATE = E_X[\underbrace{E[Y_i|T_i = 1, X_i]}_{\equiv g_1(X_i)} - \underbrace{E[Y_i|T_i = 0, X_i]}_{\equiv g_0(X_i)}]$$

  - If CEF is **not** linear, then can't use BLP to identify ATE:

$$ATE \neq E_X[\underbrace{E^*[Y_i|T_i = 1, X_i]}_{\equiv g_1^L(X_i)} - \underbrace{E^*[Y_i|T_i = 0, X_i]}_{\equiv g_0^L(X_i)}]$$

    Instead, use BLP to identify a (sort of) weighted average of treatment effects:

$$E[w_1(X_i)Y_i(1)|T_i = 1] - E[w_0(X_i)Y_i(0)|T_i = 0] =$$

$$E_X[\underbrace{E^*[Y_i|T_i = 1, X_i]}_{\equiv g_1^L(X_i)} - \underbrace{E^*[Y_i|T_i = 0, X_i]}_{\equiv g_0^L(X_i)}]$$

BLP and what it identifies

- If $Y_i(0), Y_i(1) \perp T_i$, then BLP can be used to identify the ATE (as long as you remember to include a constant)

$$ATE = E_X[\underbrace{E^*[Y_i|T_i = 1, X_i]}_{\equiv g_1^L(X_i)} - \underbrace{E^*[Y_i|T_i = 0, X_i]}_{\equiv g_0^L(X_i)}]$$

Note: throughout this exercise, we're running separate regressions on the treated and control groups, so we have a BLP for each group. The last questions asks you to pool both groups together and consider a BLP of Y_i on X_i **and** T_i.