Lecture 1: Causal Identification

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

January 29, 2020

Plan for the week

Mon:

- Discuss overview of class and syllabus.
- Explain what "causal identification" means.
- Introduce potential outcomes and causal graphs.

Wed:

- Randomized experiments.
- Explain *estimation* concepts (estimand, estimators, bias, consistency, efficiency).
- Explain *statistical inference* concepts (sampling distribution, randomization distribution, CLT, confidence intervals, p-value).

Where this class fits in

Model of quantitative research process:

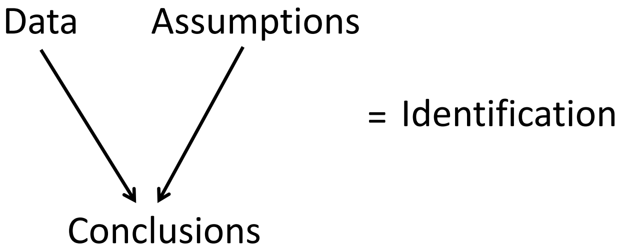
- Theory motivates causal hypothesis or target of inference:
  - *H: manipulating X results in (...) effect on Y.*
- Hypothesis, statistical theory, and substantive theory motivate a research design:
  - Operationalize X and Y.
  - Define ways to get optimal variation in X and Y given constraints.
- Research design and statistical theory motivate analysis plan:
  - Optimal estimation strategy, given constraints.
  - Optimal testing strategy, given constraints.

Where this class fits in

Model of quantitative research process:

- Theory motivates causal hypothesis or target of inference:
  - *H: manipulating X results in (...) effect on Y.*
- Hypothesis, statistical theory, and substantive theory motivate a research design:
  - Operationalize X and Y.
  - Define ways to get optimal variation in X and Y given constraints.
- Research design and statistical theory motivate analysis plan:
  - Optimal estimation strategy, given constraints.
  - Optimal testing strategy, given constraints.

What do we mean by "identification"?



What do we mean by "identification"?

Identification refers generally to sufficiency for drawing a conclusion given *the type of data* that are available.

What do we mean by "identification"?

Identification refers generally to sufficiency for drawing a conclusion given *the type of data* that are available.

Distinguish from sampling uncertainty:

What do we mean by “identification”?

Identification refers generally to sufficiency for drawing a conclusion given *the type of data* that are available.

Distinguish from sampling uncertainty:

Manski (1995, 4) : “Studies of identification seek to characterize the conclusions that could be drawn if one could...obtain an unlimited number of observations. Identification problems cannot be solved by gathering more of the same kind of data.”

What do we mean by "identification"?

Identification refers generally to sufficiency for drawing a conclusion given *the type of data* that are available.

Distinguish from sampling uncertainty:

Manski (1995, 4) : "Studies of identification seek to characterize the conclusions that could be drawn if one could...obtain an unlimited number of observations. Identification problems cannot be solved by gathering more of the same kind of data."

When we do not have identification, "the danger exists that the desired quantity will not be discernible unambiguously from the data – even when infinitely many samples are taken" (Pearl 2009, 77).

What do we mean by "identification"?

Identification refers generally to sufficiency for drawing a conclusion given *the type of data* that are available.

Distinguish from sampling uncertainty:

Manski (1995, 4) : "Studies of identification seek to characterize the conclusions that could be drawn if one could...obtain an unlimited number of observations. Identification problems cannot be solved by gathering more of the same kind of data."

When we do not have identification, "the danger exists that the desired quantity will not be discernible unambiguously from the data – even when infinitely many samples are taken" (Pearl 2009, 77).

In this class we focus on causal identification. This is a specific application of the general idea of "identification," distinct from some other applications:

What do we mean by “identification”?

Alternative application of “identification” (I):

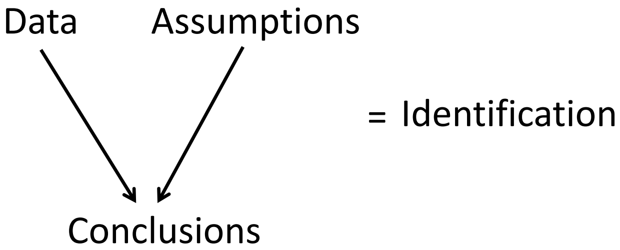
Suppose someone says...

...they prefer Klobuchar over Sanders, and

...they prefer Warren over Klobuchar.

Does this information (data) identify the person’s preference ordering over these three candidates?

What do we mean by “identification”?



What do we mean by "identification"?

Alternative application of "identification" (II):

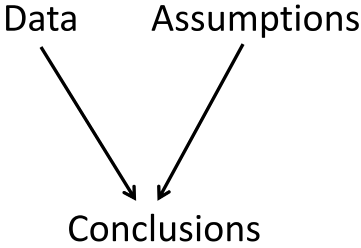
Suppose none of the coefficients below are equal to zero but the error terms (last ones) are iid mean zero draws. Which system of simultaneous equations identifies its coefficients?

$$x_t = \alpha_1^a + \alpha_2^a y_t + \nu_t^a$$
$$y_t = \beta_1^a + \beta_2^a x_t + \varepsilon_t^a$$

$$x_t = \alpha_1^b + \alpha_2^b y_t + \alpha_3^b w_t + \alpha_4^b v_t + \nu_t^b$$
$$y_t = \beta_1^b + \beta_2^b x_t + \beta_3^b w_t + \beta_4^b v_t + \varepsilon_t^b$$

$$x_t = \alpha_1^c + \alpha_2^c y_t + \alpha_3^c w_t + \nu_t^c$$
$$y_t = \beta_1^c + \beta_2^c x_t + \beta_4^c v_t + \varepsilon_t^c$$

What do we mean by “identification”?



= Identification

What do we mean by "causal identification"?

Causal identification refers to sufficiency for drawing a conclusion about a *causal effect* given the type of data at hand.

What do we mean by “causal identification”?

Causal identification refers to sufficiency for drawing a conclusion about a *causal effect* given the type of data at hand.

This class focuses on strategies for non-parametric causal identification and, in turn, estimation of causal effects.

What do we mean by “causal identification”?

Causal identification refers to sufficiency for drawing a conclusion about a *causal effect* given the type of data at hand.

This class focuses on strategies for non-parametric causal identification and, in turn, estimation of causal effects.

Angrist and Krueger (1999):

> *The combination of a clearly labeled source of identifying variation in a causal variable and the use of a particular econometric[/statistical] technique to exploit this information is what we call an identification strategy.*

Potential outcomes

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

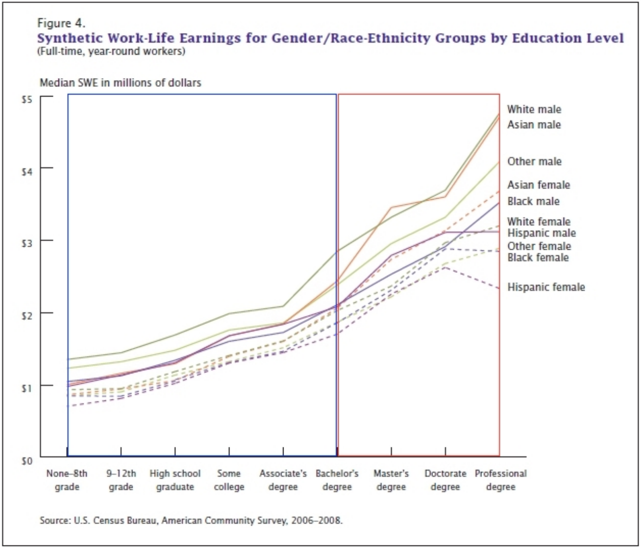
I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

Modern frameworks for causal analysis:

- Potential outcomes (Neyman, 1923; Rubin, 1974, 1978).
- Causal graphs (Pearl, 2009).

Both rely on “counterfactual” logic.

Running Example: Effect of College on Earnings*



*SWE = expected total earnings over 25-64.

Potential outcomes

A causal effect can be defined as
a contrast between "potential outcomes."

Potential outcomes

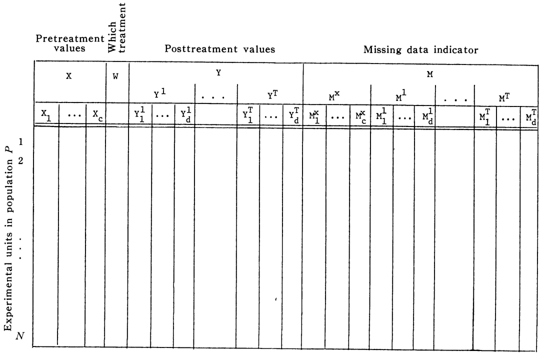


FIG. 1. All values in a study of T treatments.

Potential outcomes

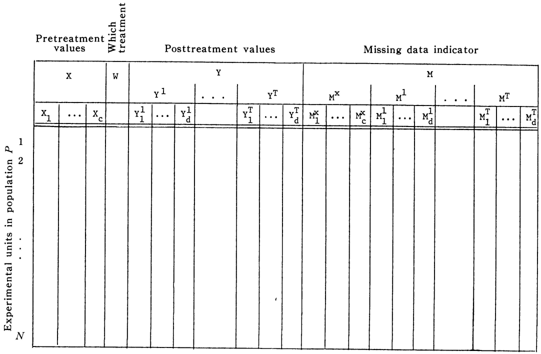


FIG. 1. All values in a study of T treatments.

- Population $\mathscr{P}$ indexed by $i = 1, \ldots, N$.

Potential outcomes

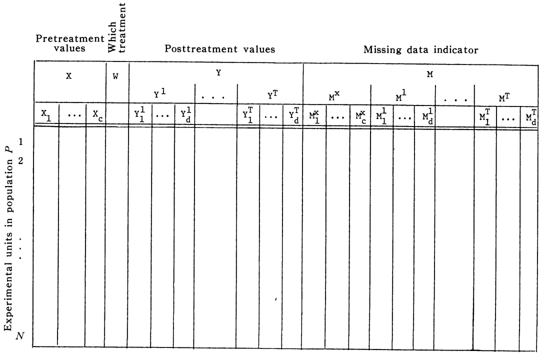


FIG. 1. All values in a study of T treatments.

- Population $\mathscr{P}$ indexed by $i = 1, \ldots, N$.
- Covariates, x_i.

Potential outcomes

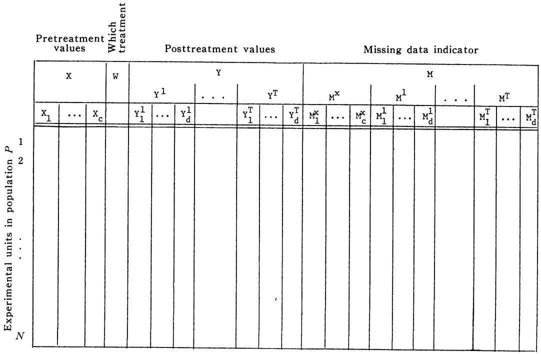


FIG. 1. All values in a study of T treatments.

- Population $\mathscr{P}$ indexed by $i = 1, \ldots, N$.
- Covariates, x_i.
- Treatment variable, W_i with support $\mathscr{W}$.

Potential outcomes

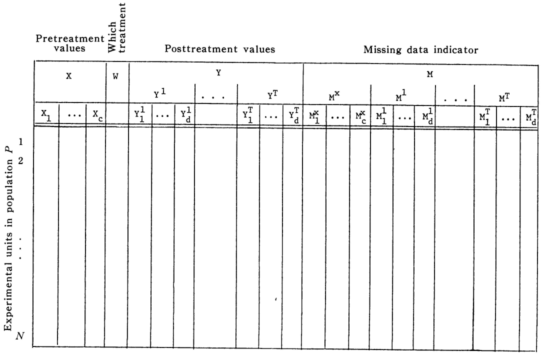


FIG. 1. All values in a study of T treatments.

- Population $\mathscr{P}$ indexed by $i = 1, ..., N$.
- Covariates, x_i.
- Treatment variable, W_i with support $\mathscr{W}$.
- Potential outcomes, y_{wi} given treatment $w \in \mathscr{W}$.

Potential outcomes

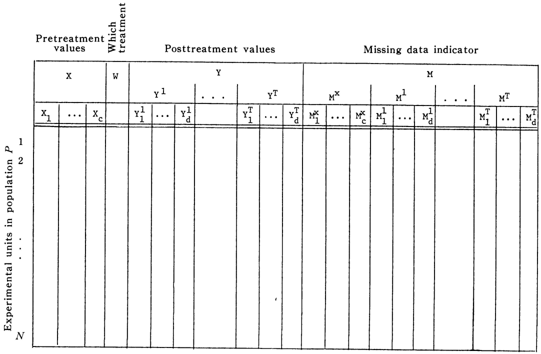


FIG. 1. All values in a study of T treatments.

- Population $\mathscr{P}$ indexed by $i = 1, \ldots, N$.
- Covariates, x_i.
- Treatment variable, W_i with support $\mathscr{W}$.
- Potential outcomes, y_{wi} given treatment $w \in \mathscr{W}$.
- Missing data indicators, $M_{i,j}^{(k)}$.

Potential outcomes

Things to note:

Potential outcomes

Things to note:

- For now, we will assume outcomes for i depend only on treatment values for i and no one else — "stable unit treatment value assumption" (SUTVA), a.k.a. "no interference."

Potential outcomes

Things to note:

- For now, we will assume outcomes for i depend only on treatment values for i and no one else — "stable unit treatment value assumption" (SUTVA), a.k.a. "no interference."
- *Unit level causal effects* for members of $\mathscr{P}$ are fixed a priori, and they compare y_{wi} to $y_{\tilde{w}i}$ for $w \neq \tilde{w}$ and $i \in \mathscr{P}$.

Potential outcomes

Things to note:

- For now, we will assume outcomes for i depend only on treatment values for i and no one else — "stable unit treatment value assumption" (SUTVA), a.k.a. "no interference."
- *Unit level causal effects* for members of $\mathscr{P}$ are fixed a priori, and they compare y_{wi} to $y_{\tilde{w}i}$ for $w \neq \tilde{w}$ and $i \in \mathscr{P}$.
- *Population causal effects* for compare aggregates of unit level causal effects for members of $\mathscr{P}$.
- Effects are defined in an "agnostic" or "non-parametric" way.
- Potential outcomes and covariates are fixed, treatments and response indicators stochastic.
- Effects are defined by letting only treatments vary, holding units fixed.
- Thus, causal effects are clearly defined for units that can conceivably receive different treatment values.
- A test for the above is "manipulation" (Holland, 1986).

Potential outcomes, causal effects, and manipulability

Holland (1986) : "For causal inference, it is critical that each unit be potentially exposable to any one of the causes."

Angrist and Krueger (1999) : "The problem of ambiguous counterfactuals is typically resolved by focusing on hypothetic manipulations in the world as is."

Potential outcomes, causal effects, and manipulability

Holland (1986) : "For causal inference, it is critical that each unit be potentially exposable to any one of the causes."

Angrist and Krueger (1999) : "The problem of ambiguous counterfactuals is typically resolved by focusing on hypothetic manipulations in the world as is."

Issues arise when trying to interpret things like race or gender. See VanderWeele and Robinson (2014) for a formal treatment of ways to interpret "race effects."

Potential outcomes and fundamental problem of causal inference

Recall, a unit level causal effect compares y_{wi} to $y_{\tilde{w}i}$ for $w \neq \tilde{w}$.

"Fundamental problem of causal inference" (Holland, 1986) : For each i potential outcomes for all w exist, but we only observe the potential outcome for the treatment value that i receives.

- "Scientific solution": Use theory to determine when units are interchangeable.
- "Statistical solution": Study features of conditional distributions, such as averages.

Causal identification under the potential outcomes model

Causal identification under the potential outcomes model

Basic statistical setting (adjusting notation to conform to MHE):

Causal identification under the potential outcomes model

Basic statistical setting (adjusting notation to conform to MHE):

- Consider a random draw, i, from $\mathscr{P}$, countable but large.

Causal identification under the potential outcomes model

Basic statistical setting (adjusting notation to conform to MHE):

- Consider a random draw, i, from $\mathscr{P}$, countable but large.
- Each draw is characterized by
  - a covariate vector, X_i,
  - potential outcomes that under SUTVA are characterized as Y_{di} for all $d \in \mathscr{D}$, as well as
  - treatment assignments, $D_i \in \mathscr{D}$.

Causal identification under the potential outcomes model

- Suppose $\mathscr{D} = \{0, 1\}$. Then under SUTVA, potential outcomes for an arbitrary draw from $\mathscr{P}$ are Y_{1i} and Y_{0i}.

Causal identification under the potential outcomes model

- Suppose $\mathscr{D} = \{0, 1\}$. Then under SUTVA, potential outcomes for an arbitrary draw from $\mathscr{P}$ are Y_{1i} and Y_{0i}.
- A unit level treatment effect for an arbitrary draw from $\mathscr{P}$ is,

$$\rho_i = Y_{1i} - Y_{0i},$$

for which,

$$E[\rho_i] = E[Y_{1i} - Y_{0i}] = \rho, \tag{1}$$

the average treatment effect (ATE).

Causal identification under the potential outcomes model

- Suppose $\mathscr{D} = \{0, 1\}$. Then under SUTVA, potential outcomes for an arbitrary draw from $\mathscr{P}$ are Y_{1i} and Y_{0i}.
- A unit level treatment effect for an arbitrary draw from $\mathscr{P}$ is,

$$\rho_i = Y_{1i} - Y_{0i},$$

for which,

$$E[\rho_i] = E[Y_{1i} - Y_{0i}] = \rho, \tag{1}$$

the average treatment effect (ATE).

- For an arbitrary draw from $\mathscr{P}$, we observe,

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{i0}.$$

Causal identification under the potential outcomes model

- Suppose $\mathscr{D} = \{0,1\}$. Then under SUTVA, potential outcomes for an arbitrary draw from $\mathscr{P}$ are Y_{1i} and Y_{0i}.
- A unit level treatment effect for an arbitrary draw from $\mathscr{P}$ is,

  $$\rho_i = Y_{1i} - Y_{0i},$$

  for which,

  $$E[\rho_i] = E[Y_{1i} - Y_{0i}] = \rho, \tag{1}$$

  the average treatment effect (ATE).
- For an arbitrary draw from $\mathscr{P}$, we observe,

  $$Y_i = D_i Y_{1i} + (1 - D_i) Y_{i0}.$$

- For our running example, we have $D_i = 1$ if college, $D_i = 0$ if not. Outcome of interest is income. ρ is the average income benefit of college.

Causal identification under the potential outcomes model

- Consider simple difference in mean college grad incomes vs mean no college incomes:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

- Then what does this difference equal?

Causal identification under the potential outcomes model

- Various decompositions are possible to illustrate sources of bias.

Causal identification under the potential outcomes model

- Various decompositions are possible to illustrate sources of bias.
- E.g, consider a decomposition wrt ATT:

$$
\begin{aligned}
E[Y_i|D_i=1]-E[Y_i|D_i=0] &= E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0] \\
&= \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{\text{Average treatment effect on the treated (ATT)}} \\
&\quad + \underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{Selection bias}}.
\end{aligned}
\tag{2}
$$

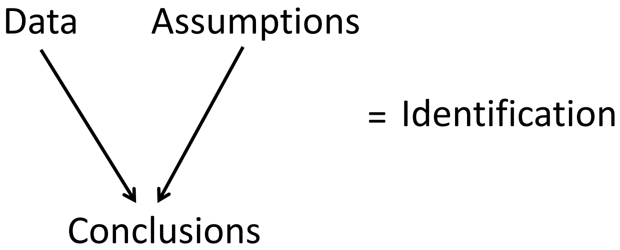
Causal identification under the potential outcomes model

- Various decompositions are possible to illustrate sources of bias.
- E.g, consider a decomposition wrt ATT:

$$\begin{aligned} E[Y_i|D_i=1]-E[Y_i|D_i=0] &= E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0] \\ &= \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{\text{Average treatment effect on the treated (ATT)}} \\ &\quad + \underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{Selection bias}}. \end{aligned} \tag{2}$$

- Could do similar wrt to ATC or effect heterogeneity (cf. CCI).

Causal identification under the potential outcomes model



Causal identification under the potential outcomes model

Now consider some identifying assumptions:

Causal identification under the potential outcomes model

Now consider some identifying assumptions:

Identifying assumption 1 (random assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) \text{ and } 0 < Pr[D_i = 1] < 1 \tag{3}$$

Causal identification under the potential outcomes model

Now consider some identifying assumptions:

Identifying assumption 1 (random assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) \text{ and } 0 < Pr[D_i = 1] < 1 \tag{3}$$

Under (3),

$$\underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection bias}} = 0.$$

(Why? E.g., if $A \perp\!\!\!\perp B$, what does this imply about $E[A|B]$?)

Causal identification under the potential outcomes model

Now consider some identifying assumptions:

Identifying assumption 1 (random assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) \text{ and } 0 < Pr[D_i = 1] < 1 \tag{3}$$

Under (3),

$$\underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection bias}} = 0.$$

(Why? E.g., if $A \perp\!\!\!\perp B$, what does this imply about $E[A|B]$?)

As such,

$$\underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{ATT}} = E[Y_{1i} - Y_{0i}],$$

so the simple difference, (2), equals ρ.

Causal identification under the potential outcomes model

Identifying assumption 2 (conditionally independent/unconfounded/strongly ignorable assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) | X_i \text{ and } 0 < Pr[D_i = 1 | X_i = x] < 1 \text{ for all } x \in \mathscr{X} \quad (4)$$

Causal identification under the potential outcomes model

Identifying assumption 2 (conditionally independent/unconfounded/strongly ignorable assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) | X_i \text{ and } 0 < Pr[D_i = 1 | X_i = x] < 1 \text{ for all } x \in \mathscr{X} \quad (4)$$

The conditional average treatment effect (CATE) is given by,

$$\rho(x) = E[Y_{1i} - Y_{0i} | X_i = x].$$

Causal identification under the potential outcomes model

Identifying assumption 2 (conditionally independent/unconfounded/strongly ignorable assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) | X_i \text{ and } 0 < Pr[D_i = 1 | X_i = x] < 1 \text{ for all } x \in \mathscr{X} \quad (4)$$

The conditional average treatment effect (CATE) is given by,

$$\rho(x) = E[Y_{1i} - Y_{0i} | X_i = x].$$

By the same logic as before, assumption (4) implies,

$$E[Y_i | D_i = 1, X_i = x] - E[Y_i | D_i = 0, X_i = x] = \rho(x).$$

(Show this.)

Causal identification under the potential outcomes model

Identifying assumption 2 (conditionally independent/unconfounded/strongly ignorable assignment):

$$D_i \perp\!\!\!\perp (Y_{1i}, Y_{0i}) | X_i \text{ and } 0 < Pr[D_i = 1 | X_i = x] < 1 \text{ for all } x \in \mathscr{X} \quad (4)$$

The conditional average treatment effect (CATE) is given by,

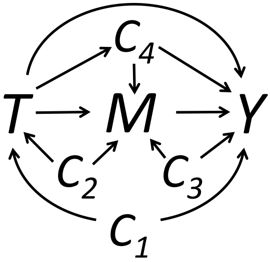
$$\rho(x) = E[Y_{1i} - Y_{0i} | X_i = x].$$

By the same logic as before, assumption (4) implies,

$$E[Y_i | D_i = 1, X_i = x] - E[Y_i | D_i = 0, X_i = x] = \rho(x).$$

(Show this.)
Marginalization over $\mathscr{X}$, the support of X_i, yields,

$$\int_{\mathscr{X}} \rho(x) dF(x) = \rho.$$



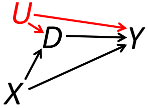
Causal graphs

Causal graphs

- Complementary analytical framework uses causal graphs (Pearl 2009).

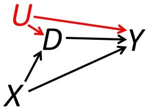
Causal graphs

- Complementary analytical framework uses causal graphs (Pearl 2009).
- Working with our example, we have in the population



Causal graphs

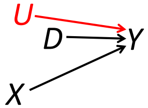
- Complementary analytical framework uses causal graphs (Pearl 2009).
- Working with our example, we have in the population



- Our target of inference is based on an hypothetical intervention:

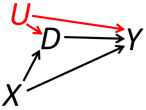
$$\mathrm{E}\left[Y|do(D_i = 1)\right] - \mathrm{E}\left[Y|do(D_i = 0)\right],$$

  equivalent to the post-intervention graph



Causal graphs

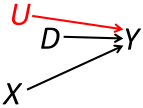
- Complementary analytical framework uses causal graphs (Pearl 2009).
- Working with our example, we have in the population



- Our target of inference is based on an hypothetical intervention:

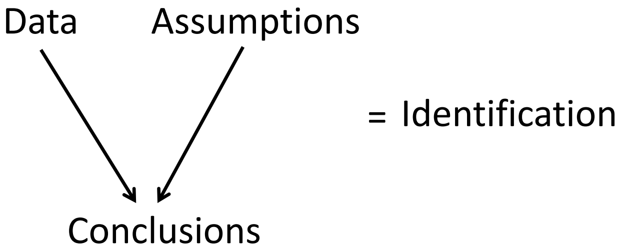
$$\mathrm{E}\left[Y|do(D_i = 1)\right] - \mathrm{E}\left[Y|do(D_i = 0)\right],$$

  equivalent to the post-intervention graph



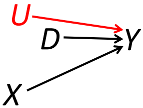
- Difference between $\mathrm{E}\left[Y|do(D_i = 1)\right] - \mathrm{E}\left[Y|do(D_i = 0)\right]$ and $\mathrm{E}\left[Y|D_i = 1\right] - \mathrm{E}\left[Y|D_i = 0\right]$ is "backdoor paths" from D to Y.

Causal graphs



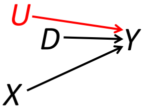
Causal graphs

- Random assignment ⇒ intervention graph *is* population graph:

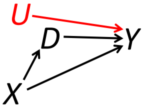


Causal graphs

- Random assignment ⇒ intervention graph *is* population graph:

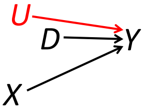


- CIA implies no backdoor path through U:

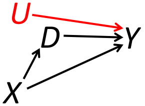


Causal graphs

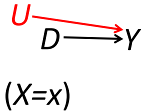
- Random assignment $\Rightarrow$ intervention graph *is* population graph:



- CIA implies no backdoor path through U:



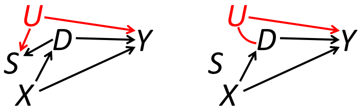
- Conditioning on X removes the other backdoor path:



- Marginalize over x to recover the intervention graph.

Causal graphs

- These are examples of "closing" backdoor paths.
- Other operations, e.g., "opening" backdoor paths by conditioning on "colliders":



Looking forward to the rest of the class

- With respect to our example and cases that resemble it:
  - These identifying assumptions rule out confounding by U_i.
  - If true, sufficient for identifying average treatment effect.
  - Plausibility is suspect, so need either other sources of identification or ways to check sensitivity to violations.

Looking forward to the rest of the class

- With respect to our example and cases that resemble it:
  - These identifying assumptions rule out confounding by U_i.
  - If true, sufficient for identifying average treatment effect.
  - Plausibility is suspect, so need either other sources of identification or ways to check sensitivity to violations.
- More generally, this semester we will look at...
  - ...these and other identifying assumptions and then statistical methods that make use of them.
  - ...other target parameters (e.g., distributional effects), relax SUTVA, etc.
  - ...both potential outcomes and causal graphs representations — each yields insights.

Looking forward to the rest of the class

- With respect to our example and cases that resemble it:
  - These identifying assumptions rule out confounding by U_i.
  - If true, sufficient for identifying average treatment effect.
  - Plausibility is suspect, so need either other sources of identification or ways to check sensitivity to violations.
- More generally, this semester we will look at...
  - ...these and other identifying assumptions and then statistical methods that make use of them.
  - ...other target parameters (e.g., distributional effects), relax SUTVA, etc.
  - ...both potential outcomes and causal graphs representations — each yields insights.
- Time for questions.