Causal Inference

Rajeev Dehejia

What is a causal effect?

- Difficult to define. Perhaps by example?
- Billiard ball (A) rolls, hits billiard ball (B). If we see a "constant conjunction" of these, then (A) causes (B) to move (Hume).
- I had a headache before lecture. I took a pill. The headache went away. Did the pill cause the pain to go away?
  - What else do we have to consider here? E.g. Might it be the Gatorade I drank before coming to class?

What is a causal effect?

- We can imagine a "structural" model: model the molecules of the pill and how they interact with pain receptors (the theory makes the relationship causal).
- More basically what do mean when we say the pill caused the pain to go away?
- That my pain would not have gone away if I did not take the pill.
  - Comparing my pain level having taking the pill with what *it would have been* if I had not taken the pill, the pain is lower.

What is a causal effect: The pill example

- Y= pain levels; **Potential outcomes**:
  - $Y_{1=}$ my pain if I take the pill
    - Outcome under treatment regime 1: takes pill
  - Y_0 my pain if I don't take the pill
    - Outcome under treatment regime 0: no pill; status quo; control
- So the causal claim is $Y_1 < Y_0$.
  - "Pain if I take the pill will be less than had I not taken the pill"; "The pill causes a reduction in pain compared to not having taken it"
  - Many ways to scale this (Y_1/Y_0 <1), **but most common is $\tau = Y_1 - Y_0 < 0$ (call this the treatment effect).**
- Think of the "Y's" as data, things that could have been observed, except that only one of these was observed.

Causality with potential outcomes

Treatment Effect

The treatment effect or causal effect of the treatment on the outcome for unit *i* is the difference between its two potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

Observed Outcomes

Observed outcomes are realized as

$$Y_i = Y_{1i}T_i + Y_{0i}(1 - T_i) \text{ or } Y_i = \begin{cases} Y_{1i} \text{ if } T_i = 1 \\ Y_{0i} \text{ if } T_i = 0 \end{cases}$$

The fundamental problem of causal inference

- Only one of Y_1 and Y_0 can ever be observed for a given unit.
  - For a given unit, either you receive the pill or you didn’t. Both are not possible. So you can only observe one these.
  - **So the individual-level treatment effect can *never* be observed**.
  - Of course, with some assumptions you might make progress… and that’s what we’ll be working on all semester!

Identification problem for causal inference

Problem

Causal inference is difficult because it involves missing data. How can we find $Y_{1i} - Y_{0i}$*?*

- A large amount of homogeneity would solve this problem:
  - (Y_{1i}, Y_{0i}) constant across time: Easiest one is temporal stability. Try the same experiment on the same person at different points in time.
  - (Y_{1i}, Y_{0i}) constant across individuals: Another easy one is to assume that two units both starting with a headache are otherwise identical. Give one the pill, and not to the other.
- Unfortunately, often there is a large degree of heterogeneity in the individual responses to participation in public programs

Stable unit treatment value assumption (SUTVA)

Assumption

Observed outcomes are realized as

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

Two key components of this assumption:

1. **No interference**: (units do not interfere with each other)- treatment applied to one unit does not effect the outcome for another unit

2. There is only a single version of each treatment level (**potential outcomes must be well defined**)

Does not solve identification (fundamental problem) but makes things easier…

What are some threats to validity you remember from stat 2 or Program Evaluation that might violate these assumptions?

No causation without manipulation

- One of the implications of this definition of causality is that a cause ("X"; "independent variable") is a **treatment that we could plausibly *manipulate***.
- If it can't be manipulated, it's not a treatment.
- Example:
  - She did well on the exam because she is a woman. ✖
  - She did well on the exam because she studied for it. ✓
  - She did well on the exam because she was coached by her teacher. ✓ ✓

The problem with multiple units

Unit	Outcome under treatment	Outcome under control	Treatment effect
1	Y_{11}	~~Y_{01}~~	τ_1
2	~~Y_{12}~~	Y_{02}	τ_2
3	Y_{13}	~~Y_{03}~~	τ_3
4	Y_{14}	~~Y_{04}~~	τ_4
N	~~Y_{1N}~~	Y_{0N}	τ_N
Sum	$\mu_T = \Sigma Y_{1n}/N$	$\mu_C = \Sigma Y_{0n}/N$	$\tau = \Sigma \tau_n/N = \mu_T - \mu_C$

The fundamental problem (again)

$$\tau = E(Y_{1i}) - E(Y_{0i})$$
$$= \Pr(T_i = 1)E(Y_{1i} | T_i = 1) + \Pr(T_i = 0)E(Y_{1i} | T_i = 0)$$
$$- \left[\Pr(T_i = 1)E(Y_{0i} | T_i = 1) + \Pr(T_i = 0)E(Y_{0i} | T_i = 0)\right]$$

- Or selection bias (example)

The fundamental problem, multiple units

Unit	Outcome under treatment	Outcome under control	Treatment effect
1	Y_{11}	~~Y_{01}~~	τ_1
2	~~Y_{12}~~	Y_{02}	τ_2
3	Y_{13}	~~Y_{03}~~	τ_3
4	Y_{14}	~~Y_{04}~~	τ_4
N	~~Y_{1N}~~	Y_{0N}	τ_N
Sum	$\mu_T = \boldsymbol{\Sigma} Y_{1n}/N$	$\mu_C = \boldsymbol{\Sigma} Y_{0n}/N$	$\tau = \boldsymbol{\Sigma}\tau_n/N = \mu_T - \mu_C$
	$\bar{x}_T = \sum_{i \in \{T\}} y_{1i}/N_T$	$\bar{x}_C = \sum_{i \in \{C\}} y_{01i}/N_C$	
	$E(\bar{x}_T) \neq \mu_T$	$E(\bar{x}_C) \neq \mu_C$	

Quantities of interest (estimands)

ATE

Average treatment effect is: $\alpha_{ATE} = E\left[Y_1 - Y_0\right]$

Difference in mean potential outcomes for the overall population of interest

ATET

Average treatment effect on the treated is:

$$\alpha_{ATET} = E\left[Y_1 - Y_0 \mid T = 1\right]$$

The ATE on the treated (ATET) is like the ATE, but it uses only the subjects who were observed in the treatment group.

Average treatment effect (ATE)

Imagine a population with 4 units:

i	Y_{1i}	Y_{0i}	Y_i	T_i	Y_{1i} - Y_{0i}
1	3	?	3	1	?
2	1	?	1	1	?
3	?	0	0	0	?
4	?	1	1	0	?

What is $\alpha_{ATE} = E\left[Y_1 - Y_0\right]$?

Average treatment effect (ATE)

Imagine a population with 4 units:

i	Y_{1i}	Y_{0i}	Y_i	T_i	Y_{1i} - Y_{0i}
1	3	0	3	1	3
2	1	1	1	1	0
3	1	0	0	0	1
4	1	1	1	0	0

What is $\alpha_{ATE} = E\left[Y_1 - Y_0\right]$?

Average treatment effect (ATE)

Imagine a population with 4 units:

i	Y_{1i}	Y_{0i}	Y_i	T_i	$Y_1 - Y_{0i}$
1	3	0	3	1	3
2	1	1	1	1	0
3	1	0	0	0	1
4	1	1	1	0	0
$E[Y_1]$	1.5				
$E[Y_0]$		0.5			
$E[Y_1 - Y_0]$					1

$$\alpha_{ATE} = E\left[Y_1 - Y_0\right] = 3\cdot\left(1/4\right) + 0\cdot\left(1/4\right) + 1\cdot\left(1/4\right) + 0\cdot\left(1/4\right) = 1$$

Average treatment effect on the treated (ATET)

Imagine a population with 4 units:

i	Y_{1i}	Y_{0i}	Y_i	T_i	Y_1 - Y_{0i}
1	3	0	3	1	3
2	1	1	1	1	0
3	1	0	0	0	1
4	1	1	1	0	0

What is $\alpha_{ATET} = E\left[Y_1 - Y_0 \middle| T = 1\right]$?

Average treatment effect on the treated (ATET)

Imagine a population with 4 units:

i	Y_{1i}	Y_{0i}	Y_i	T_i	$Y_1 - Y_{0i}$
1	3	0	3	1	3
2	1	1	1	1	0
3	1	0	0	0	1
4	1	1	1	0	0
$E[Y_1 \mid T = 1]$	2				
$E[Y_0 \mid T = 1]$		0.5			
$E[Y_1 - Y_0 \mid T = 1]$					1.5

$$\alpha_{ATET} = E\left[Y_1 - Y_0 \mid T = 1\right] = 3 \cdot \left(1/2\right) + 0 \cdot \left(1/2\right) = 1.5$$

Selection bias

Problem

Comparisons of earnings for the treated and the untreated do not usually give the right answer:

$$E\left[Y \middle| T=1\right]-E\left[Y \middle| T=0\right]=E\left[Y_1 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]$$

$$=\underbrace{E\left[Y_1-Y_0 \middle| T=1\right]}_{ATET}+\underbrace{\left\{E\left[Y_0 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]\right\}}_{BIAS}$$

- Bias term is not likely to be zero for most public policy applications
- Selection into treatment is often associated with the potential outcomes

Selection bias

Problem

Comparisons of earnings for the treated and the untreated do not usually give the right answer:

$$E\left[Y \middle| T=1\right]-E\left[Y \middle| T=0\right]=E\left[Y_1 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]$$

$$=\underbrace{E\left[Y_1-Y_0 \middle| T=1\right]}_{ATET}+\underbrace{\left\{E\left[Y_0 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]\right\}}_{BIAS}$$

Example: Job training program for disadvantaged

- Participants are self-selected from a subpopulation of individuals in difficult labor situations
- Post-training period earnings for participants would be lower than those for nonparticipants in the absence of the program

$$\left(E\left[Y_0 \middle| T=1\right]-E\left[Y_0 \middle| T=0\right]<0\right)$$

Training program for the disadvantaged in the U.S.

Data from the National Supported Work Demonstration (NSW)

TABLE 1.—MEAN EARNINGS PRIOR, DURING, AND SUBSEQUENT TO TRAINING FOR 1964 MDTA CLASSROOM TRAINEES AND A COMPARISON GROUP

	White Males		Black Males		White Females		Black Females	
	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group
1959	$1,443	$2,588	$ 904	$1,438	$ 635	$ 987	$ 384	$ 616
1960	1,533	2,699	976	1,521	687	1,076	440	693
1961	1,572	2,782	1,017	1,573	719	1,163	471	737
1962	1,843	2,963	1,211	1,742	813	1,308	566	843
1963	1,810	3,108	1,182	1,896	748	1,433	531	937
1964	1,551	3,275	1,273	2,121	838	1,580	688	1,060
1965	2,923	3,458	2,327	2,338	1,747	1,698	1,441	1,198
1966	3,750	4,351	2,983	2,919	2,024	1,990	1,794	1,461
1967	3,964	4,430	3,048	3,097	2,244	2,144	1,977	1,678
1968	4,401	4,955	3,409	3,487	2,398	2,339	2,160	1,920
1969	$4,717	$5,033	$3,714	$3,681	$2,646	$2,444	$2,457	$2,133
Number of Observations	7,326	40,921	2,133	6,472	2,730	28,142	1,356	5,192

Assignment mechanism

Assignment Mechanism

Assignment mechanism is the procedure that determines which units are selected for treatment intake. Examples include:

- Random assignment
- Selection on observables
- Selection on unobservables

Most models of causal inference attain identification of treatment effects by restricting the assignment mechanism in some way.

Key ideas

- Causality is define by potential (not realized or observed) outcomes.
- Observed association is neither necessary nor sufficient for causation.
- Estimation of treatment effects usually starts with studying the assignment mechanism.

How does this relate to regression?

$$
\begin{aligned}
Y_i &= T_i Y_{1i} + (1 - T_i) Y_{0i} \\
&= Y_{0i} + T_i (Y_{1i} - Y_{0i}) \\
&= \alpha + \beta_i T_i + \varepsilon_i, \quad \text{where } \varepsilon_i = Y_{0i} - \alpha
\end{aligned}
$$

```
In Stata:
reg outcome_var treatment_var
```

In potential outcomes framework

$$
\begin{aligned}
Y_i &= T_i Y_{1i} + (1 - T_i) Y_{0i} \\
&= Y_{0i} + T_i (Y_{1i} - Y_{0i}) \\
&= \alpha + \beta_i T_i + \varepsilon_i, \quad \text{where } \varepsilon_i = Y_{0i} - \alpha
\end{aligned}
$$

- Think of Y_{0i} as the omitted variable.
- What is the correlation between T_i and Y_{0i}?
- If positive then OLS will overestimate the treatment effect (those with high Y_0 assigned to treatment; we observe low Y_0's of those in control; overstate treatment effect).
- Finally a good explanation of error term and why it would ever be uncorrelated with “x”

A regression approach to causality

- Want effect of *X* on distribution of *y*, other relevant things being held constant.
- Most common to be interested in effect on mean of y, i.e.:

$$\frac{\partial E\left(y \middle| X, ?\right)}{\partial X}$$

Bonus slides

- The next slides dive deeper into the idea of how the Rubin Causal model or potential outcomes framework relates to the regression framework.
- These slides are optional – feel free to peruse or bypass.

Estimation of linear regression offers promising approach

- Can interpret regression function (*Xβ*) as an approximation of E(*y*|*X*).
- If error term uncorrelated with X can use OLS
- If conditional expectation linear in *X* then exact.
- If conditional expectation non-linear then *Xβ* linear approximation to true function.
- This is same as:

$$p\lim \hat{\beta}^{OLS} = \frac{Cov(X, y)}{Var(X)}$$

Problems with inferring causal effects from regressions

- Absent all those assumptions, regressions tell us about correlations but 'correlation is not causation'
- Example: Regression of hours of child labor work on weight of the child:

hours	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
weight	.5479186	.022497	24.36	0.000	.5038152	.592022
_cons	2.204502	.7003772	3.15	0.002	.8314705	3.577533

- Do you work more because you are healthier? A causal effect?
- Interpretation: Each additional kg in weight is associated with a 0.54 hour increase in labor hours.

General problems in estimating causal effects

- Omitted Variables
- Reverse Causality
- Measurement Error
- Sample selection

Omitted variables (should be familiar)

- Suppose we want to estimate E(y|X,W) assumed to be linear in (X,W), so that E(y|X,W) =Xβ+Wγ or:

$$y = X\beta + W\gamma + \varepsilon$$

- But you estimate

$$y = X\beta + u$$

- i.e. E(y|X). Where there is only one variable:

$$p\lim \hat{\beta} = \beta + \gamma \frac{Cov(W, X)}{Var(X)}$$

- Extent of omitted variables bias related to:
  - size of correlation between X and W
  - strength of relationship between y and W

In child labor example…

- Age an obvious omitted variable:

hours	Coef.	Std. Err.	t
weight	.180696	.0356908	5.06
age	1.581047	.1205299	13.12
_cons	-4.381884	.8527244	-5.14

- The coefficient on weight goes down from 0.54 to 0.18
- The omitted variable (age) was partially absorbed by weight. In what direction was the bias? Why does this intuitively make sense? How would you interpret these outputs?

Reverse causality/ endogeneity

- Idea is that correlation between *y* and *X* may be because it is *y* that causes *X* not the other way round
- Interested in causal model:

$$y = X\beta + \varepsilon$$

- But also causal relationship in other direction:

$$X = \alpha y + u$$

How does this relate to RCM?

- Recall

$$\begin{aligned} Y_i &= T_i Y_{1i} + (1 - T_i) Y_{0i} \\ &= Y_{0i} + T_i (Y_{1i} - Y_{0i}) \\ &= \alpha + \beta_i T_i + \varepsilon_i, \quad \text{where } \varepsilon_i = Y_{0i} - \alpha \end{aligned}$$

- In the standard causal model, we maintain a clear timeline, with all variables on the RHS being pre-treatment variables. So we “design” this problem away.
- So while $T \neq f(Y_i)$, it could dependent on expectations of post-treatment $Y_{i.}$

Measurement error

- Most (all?) of our data are measured with error.
- Suppose causal model is:

$$y = X^{*}\beta + \varepsilon$$

- But only observe X which is X* plus some error:

$$X = X^{*} + u$$

- Classical measurement error:

$$E(u|X^{*})=0$$

Measurement error

- Can write causal relationship as:

$$Y = X\beta - u\beta + \varepsilon$$

- Note that *X* correlated with composite error.
- Should know this leads to bias/ inconsistency in OLS estimator
- Can make some useful predictions about nature of bias – later on in course.
- Want $E(y|X^*)$ but can only estimate $E(y|X)$
- No specific RCM reinterpretation. Nothing specific to the evaluation framework.

Selection effects

- The following regression seems to show that children who work spend less hours in school:

schlhrs	Coef.	Std. Err.	t
econhrs	-.2401388	.022659	-10.60
_cons	17.16715	.2712731	63.28

- Is this sensible? Possibly. But not necessarily causal…

In stata:

```
reg schlhrs econhrs
```

Selection effects

- One explanation is sample selection
  - Sample of children who work are also those less likely to go to school for reasons other than the work (poverty, good physical skills, parental motivation).
  - This will be a major theme in most quasi-experimental or “natural” experiments, and we will discuss ways to mitigate this error using statistical modelling

Why is this? A brief exposition

- Standard example, causal model for everyone:

$$y = X\beta + \varepsilon$$

- But only observe if work, W=1, so estimate E(y|X,W=1) not E(y|X)
- Sample selection bias if W correlated with ε.
- In an evaluation/ causal framework, we observe Y_1 (Y_0) for a selected sample of individuals, T=1 (0). Recall that ε contains Y_{0i}, so the issue is that assignment to treatment is can be correlated with Y_0.

Common features of problems

- ✓ All problems have an expression in everyday language – omitted variables, reverse causality etc
- ✓ All have an econometric form – the same one
- ✓ A correlation of *X* with the 'error'
- ✓ All relate to the relationship between the assignment rule and potential outcomes.

How to surmount the problems?

- More sophisticated econometric methods than OLS e.g. IV.
- Better data – Griliches:
  - "since it is the 'badness' of the data that provides us with our living, perhaps it is not at all surprising that we have shown little interest in improving it"

But recent trends

- Much more emphasis on good quality data and research design than 'statistical fixes' – the 'credibility revolution'
- Probably started in labour economics but now arriving in most fields
- Will illustrate this in course through wide-ranging examples

Choosing your data...

- Suppose interested in causal effect of *X* on *y*.
- Can choose the way in which *X* is determined in your sample
  - may seem fanciful but field experiments becoming more common in economics
- Good reason to choose to do randomized controlled experiment