Session 1: Causal Inference

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

What is econometrics?

- Statistics applied to economics and management
- Tools to **test** (social) scientific theories
- Measuring **causal effects** of a variable on another
- (Forecasting?)

Causal effects

- Causal effects are common in research and business
- Examples?
- *Does a phone alert increase engagement with my app?*
- *What is the effect of an ad on sales?*
- *Does renewable energy increase electricity tariffs?*
- *Does minimum wages reduce employment?*
- *Does carbon pricing reduce CO2?*

Notation and definitions

- **Unit**: i (individual, firm, country, etc.). Complement: $-i$.
- **Treatment**: $D \in \{0, 1\}$ (later: $D \in \mathfrak{R}$)
  - AKA: independent variable, cause, intervention
- **Outcome**: $Y \in \mathfrak{R}$
  - AKA: dependent variable, effect, outcome
- **Potential outcomes**: values of Y under $D = 0$ or $D = 1$
  - Write Y^0 and Y^1, or $(Y|D = 0)$, $(Y|D = 1)$
- **Causal effect**: *comparison* of PO (Rubin 1974):

$$Y(D_i = 1, \mathbf{d}_{-i}) \text{ versus } Y(D_i = 0, \mathbf{d}_{-i})$$

- $\mathbf{d}$: vector of treatment assignment (eg $\mathbf{d}' = [0, 0, 1..., 0]$)

Estimand

- Many ways to do a comparison...
- Most common: **unit treatment effect** (UTE):

$$\text{UTE} \equiv Y(D_i = 1, \mathbf{d}_{-i}) - Y(D_i = 0, \mathbf{d}_{-i})$$

- We could also tak a (risk) ratio: $\frac{Y(D_i=1,\mathbf{d}_{-i})}{Y(D_i=0,\mathbf{d}_{-i})}$
- Excess risk ratio: $\frac{E[Y^1]-E[Y^0]}{E[Y^0]}$
- etc.
- Terminology: quantities of interest like UTE are **estimands**

- Can we simplify? Yes! (Rubin 1980; Imbens and Rubin 2015)
- Stable Unit Treatment Value Assumption (**SUTVA**): *The potential outcomes for any unit do not vary with the treatment assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes*
- Short version: no spillovers, treatment is the same for all units
- **Unit treatment effect** with SUTVA:

$$\begin{aligned} \text{UTE} &\equiv Y(D_i = 1) - Y(D_i = 0) \\ &\equiv \delta_i \\ &\equiv Y_i^1 - Y_i^0 \end{aligned}$$

- Great! With SUTVA: difference between Y_i^1 and Y_i^0

Example 1

For a given **sample** (w/ heterogeneous UTE)...

i	Y^0	Y^1	δ_i	D
1	4	8	4	0
2	2	8	6	0
3	5	6	1	0
4	5	7	2	1
5	4	9	5	1

Example 2

For a given **sample** (w/ homogeneous UTE)...

i	Y^0	Y^1	δ_i	D
1	4	8	4	0
2	4	8	4	0
3	5	9	4	0
4	5	9	4	1
5	4	8	4	1

- Thanks for a great semester!
- Wait...

Fundamental problem of causal inference

Problem: we can't observe $Y_i^1 - Y_i^0$

i	Y^0	Y^1	Y^{observed}	δ_i
1	4		4	?
2	2		2	?
3	5		5	?
4		7	7	?
5		9	9	?

Concretely:

- We don't observe sales with/without an ad
- We don't observe elec tariffs with/without renewables
- We don't observe CO_2 emissions with/without EU ETS
- We don't observe unemployment with/without minimum wage
- **Fundamental problem of causal inference**

Other estimand

- Is there a way around this?
- Shift goal post: instead of i, we focus on averages
- **Average treatment effect** (ATE):

$$\begin{aligned} E[\delta] &= E[Y^1 - Y^0] \\ &= E[Y^1] - E[Y^0] \end{aligned}$$

- Expected value: $E[\cdot] \equiv \sum_{i=1}^{n} x_i p(x_i)$ or $E[\cdot] \equiv \int_{-\infty}^{\infty} x f(x) dx$
- Averages are useful! Eg forecast of effect of new policy

i	Y^0	Y^1	δ_i	D
1	4	8	4	0
2	2	8	6	0
3	5	6	1	0
4	5	7	2	1
5	4	9	5	1
ATE	$= 3.6$			

$$\frac{1}{n}\sum_{i}^{n}\delta_i = \frac{18}{5} = 3.6$$

Nota bene

1. Yet: we can't observe ATE either!
2. $E[\delta]$ is a **population** parameter that needs to be **estimated**.
3. The task of the analyst is to build a **counterfactual**. Many ways to do so!
4. $E[\delta]$ is not the only estimand of interest: sd, quintiles, etc.
5. $E[\cdot]$ isn't always defined (eg if distribution is Cauchy)

Estimating ATE

- As we just saw: ATE can't be observed
- But we have values for Y^0 $(i = 1, 2, 3)$ and Y^1 $(i = 4, 5)$!
- Why not just take the mean of the two groups?

$$\tilde{\delta} = (\bar{Y}|D = 1) - (\bar{Y}|D = 0)$$

- Notation: $\tilde{\delta}$ is an **estimator** of an **estimand**
- To see why not: need to introduce **conditional average treatment effects** (CATE): ATE for subgroups
- **ATE on the treated** (ATT) and **ATE on the control** (ATC):

$$E[Y^1|D = 1] - E[Y^0|D = 1] \equiv \text{ATT}$$
$$E[Y^1|D = 0] - E[Y^0|D = 0] \equiv \text{ATC}$$

- ATT and ATC don't have to be the same!

In our example:

i	Y^0	Y^1	δ_i	D
1	4	8	4	0
2	2	8	6	0
3	5	6	1	0
4	5	7	2	1
5	4	9	5	1
ATE	$=$ 3.6	$(18/5)$		
ATT	$=$ 3.5	$(7/2)$		
ATC	$=$ 3.7	$(11/3)$		

So why is $\widetilde{\delta}$ problematic?

i	Y^0	Y^1	δ_i	D
1	4		?	0
2	2		?	0
3	5		?	0
4		7	?	1
5		9	?	1
ATE	$=$ 3.6	(18/5)		
$\tilde{\delta}$	$=$ 4.3	(8-3.67)		

A bit different. But is it **sampling noise** or **bias** (to be defined)?

Realized Y	$D = 0$	$D = 1$	**Overall**
Y^0	a	b	$E[Y^0] = a(1 - \pi) + \pi b$
Y^1	c	d	$E[Y^1] = c(1 - \pi) + \pi d$

- π: population share of treated units
- a: untreated outcome for control units (observable)
- d: treated outcome for units that are treated (observable)
- b and c: counterfactual cases (unobservable)
- Recall: ATE $= E[Y^1] - E[Y^0]$
- Recall: $\tilde{\delta} = d - a$. Same as ATE?

$$
\begin{aligned}
\text{ATE} \equiv E[\delta] &= E[Y^1] - E[Y^0] \\
&= \underbrace{\left[\pi E[Y^1|D=1] + (1-\pi)\overbrace{E[Y^1|D=0]}^{\text{unobservable}}\right]}_{\text{weighted average of } Y^1} \\
&- \underbrace{\left[\pi \overbrace{E[Y^0|D=1]}^{\text{unobservable}} + (1-\pi)E[Y^0|D=0]\right]}_{\text{weighted average of } Y^0} \\
&\equiv [\pi d + c(1-\pi)] - [a(1-\pi) + \pi b]
\end{aligned}
$$

This gives us:

$$
\begin{aligned}
\text{ATE} &= E[Y^1] - E[Y^0] \\
&= [\pi d + c(1-\pi)] - [a(1-\pi) + \pi b] \\
&\textbf{versus} \\
\tilde{\delta} &= E[Y^1|D=1] - E[Y^0|D=0] \\
&= d - a
\end{aligned}
$$

We can intuitively guess that $\tilde{\delta} = \text{ATE}$ **iff** $d = c$ and $a = b$.

Let's derive this by hand...

$$\underbrace{E[Y^1|D=1] - E[Y^0|D=0]}_{\text{naive difference } \tilde{\delta}} = \underbrace{E[\delta]}_{\text{ATE}}$$
$$+ \underbrace{[E[Y^0|D=1] - E[Y^0|D=0]]}_{\text{baseline bias}}$$
$$+ (1-\pi)\underbrace{[E[\delta]D=1] - E[\delta|D=0]]}_{\text{differential average treatment effect bias}}$$

1. **Baseline bias**: the two groups are different at baseline
2. **Differential treatment effect bias**: average treatment effect isn't the same for T and C

Deep breath

- We want to study causal effects
- Problem: UTE and ATE cannot be observed (fundamental problem of causal inference)
- Naive estimator $\tilde{\delta}$ is only okay if two types of biases are removed (baseline and differential treatment effect).
- Big, big problem: these biases are very likely when we work with **observational** data (vs. **experimental** data)
- Need more complex modeling and estimation methods

Example (1)

- Does a college degree (D) increase a person's income (Y) ($i =$individuals)?
- **Baseline bias**: college students might have been richer even withouth college ($E[Y^0|D=1] > E[Y^0|D=0]$)
- **Differential TE bias**: college students go to college because they know they would get more out of it ($E[\delta]D=1] > E[\delta|D=0] > 0$)
- The two could in theory cancel each other, but who knows?
- **Identification strategy**: design that eliminates these biases.

Example (2)

- Does a carbon tax (D) reduce CO2 emissions (Y) ($i =$states)?
- **Baseline bias**: states who plan to tax CO2 emit more in the first place ($E[Y^0|D=1] > E[Y^0|D=0]$)
- **Differential TE bias**: states who tax CO2 do so if they know the policy will work $(E[\delta]D=1] < E[\delta|D=0] < 0)$

Desperate situation?

Two assumptions that can help us:

1. Same potential treated Y: $E[Y^1|D=1] = E[Y^1|D=0]$
2. Same potential control Y: $E[Y^0|D=1] = E[Y^0|D=0]$

We then have three cases.

- If both assumptions hold: ATE!

- If only the first assumption:

$$\begin{aligned}\tilde{\delta} &= E[Y^1|D=1] - E[Y^0|D=0] \\ &= E[Y^1|D=0] - E[Y^0|D=0] \qquad \text{by assumption} \\ &\equiv \text{ATC}\end{aligned}$$

- If only the second assumption holds, then we can recover:

$$\begin{aligned}\tilde{\delta} &= E[Y^1|D=1] - E[Y^0|D=0] \\ &= E[Y^1|D=1] - E[Y^0|D=1] \qquad \text{by assumption} \\ &\equiv \text{ATT}\end{aligned}$$

Plan for the semester

This course give you the tools to...

- (a) understand inferential statistics,
- (b) understand threats to causal inference, and
- (c) design studies that minimize assumptions
- (d) with an eye on applicability to business, innovation, policy

Session	Date	Topics	Reading		
			[SW]	[V]	Other
1	Feb.19	Causal inference	1-3	A-B	[MW 1, 2]
	Feb.23	*Exercise*			
2	Feb.26	Regression: basics	4	2.1-2.3	
	Mar.01	*Exercise*			
3	Mar.04	Regression: advanced	6	2.8-2.9, 3	[AP 3]
	Mar.08	*Exercise*			
4	Mar.11	Model specification	8	3	
	Mar.15	*Exercise*			
5	Mar.18	Model evaluation and inference (1)	5, 7	2.4-2.6, 4	
	Mar.22	*Exercise*			
6	Mar.25	Model evaluation and inference (2)			
	Mar.29				
	Apr.01 Apr.05	Easter			
7	Apr.08	Threats to inference	9	3.2, 5.1-5.2	[AP 1, 4, 6], [MW 3, 4]
	Apr.12	*Exercise*			
8	Apr.15	Randomized controlled trials	13.1, 13.2		[AP 2]
	Apr.19	*Exercise*			
9	Apr.22	Quasi experiments (1): panel data	10	10	[AP 5], [MW 11]
	Apr.26	*Exercise*			
10	Apr.29	Quasi experiments (2): IV	12	5.3-5.5	[AP 4], [MW 9]
	May.03	*Exercise*			
11	May.06	Quasi experiments (3): others	13		[AP 4], [MW 11.3]
	May.10	*Exercise*			
12	May.13	Discrete dependent variables and maximum likelihood	11	6-7	
	May.17	*Exercise*			
[13]	May.20	Whit Monday			
	May.24	Special topic/buffer class			
14	May.27	Q&A session about the final exam			
	May.31				

Textbooks

- Stock and Watson (2011), Verbeek (2018), on econometrics
- Angrist and Pischke (2008), Morgan and Winship (2014) on causal inference
- Lots of textbooks, some more basic (Wooldridge 2012), some less so (Imbens and Rubin 2015), others useful as references (Greene 2008). Shop around.
- Extra readings will be posted on Moodle

Assignments

- Individual problem sets: 40%. Distributed ~2 weeks before they are due.
- Written exam during the exam session: 60%

Questions?

References

Angrist, Joshua, and Steffan J. Pischke. 2008. *Mostly Harmless Econometics*. Princeton: Princeton University Press.

Greene, William H. 2008. *Econometric Analysis*. 6th ed. New York: Pearson.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Morgan, Steven L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd Edition. Cambridge: Cambridge University Press.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688.

———. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment." *Journal of the American Statistical Association* 75 (371): 591–93.

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.

Wooldridge, Jeffrey M. 2012. *Introductory Econometrics*. 5th ed. South-Western Cengage Learning.