ECON42720 Causal Inference and Policy Evaluation

3 Potential Outcomes and Randomized Experiments

Ben Elsner (UCD)

About this Lecture

Lingo and Notation

The **lingo of causal inference** is borrowed from **medical trials**

- **treatment** is the intervention/variable whose effect we are interested in
- **treatment group** is the group of units that receives the treatment
- **control group** is the group of units that does not receive the treatment or receives a placebo
- **outcome** is the variable that is potentially affected by the treatment

We are after the causal effect: **treatment** (D) $\rightarrow$ **outcome** (Y)

Resources

This lecture is based on

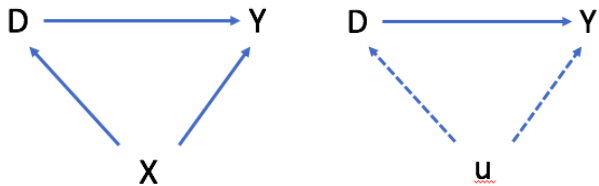
- Cunningham (2020), Chapter 4
- Angrist & Pischke (2009), Chapter 2

Find more about the course on the **course page**

Notation

Causality

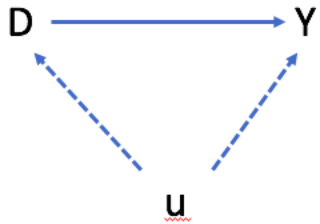
A common challenge in applied econometrics is to **separate a causal effect** from the **influence of third factors**



We often have a **good (theoretical) idea** why $D \to Y$

- but X **is a confounding factor**
- often the problem: we also have an idea what X could be...
- but **cannot observe it** (notation: u for "unobservable")

Causality



Common challenge: **selection into treatment**

In microeconomics we learn

- **people** make **rational choices**...
- ...as do **firms**
- ...as do **governments**

Problem: we do **not observe all the determinants** of these choices (i.e. u)

Causality

Examples for **selection into treatment**:

Going to the gym makes you healthier

- good reason to believe so
- but people who go to the gym are different from those who don't
- observed correlation $\neq$ causation

Exporting boosts firm profitability

- good reason to believe so
- but exporters are different in many ways from non-exporters
- observed correlation $\neq$ causation

Causality

Problem: it is **not easy to account for all confounding factors**

- because we know what they are but can't observe them
- or because we don't know what they are (e.g. "common shocks' ')

The **burden of proof is on the researcher**

- If you make a causal statement, you need to make a convincing case
- but causality cannot be proven without assumptions...
- ...and these assumptions need to be believed

The Potential Outcomes Framework

We will now focus on **binary treatments**

Some **notation from experimental studies**

- i is an index for the units in the population under study.

- D_i is the **treatment status**:
- $D_i = 1$ if unit i has been exposed to treatment,
- $D_i = 0$ if unit i has not been exposed to treatment.

- $Y_i(D_i)$ indicates the **potential outcome according to treatment**:
- $Y_i(1) \equiv Y_{i1}$ is the outcome in case of treatment,
- $Y_i(0) \equiv Y_{i0}$ is the outcome in case of no treatment.

The Potential Outcomes Framework (Rubin, 1974)

The **observed outcome** is then

$$Y_i = \begin{cases} Y_{i1} & \text{if } D_i = 1 \\ Y_{i0} & \text{if } D_i = 0 \end{cases}$$

$$= D_i Y_{i1} + (1 - D_i) Y_{i0}$$

$$= Y_{i0} + \underbrace{(Y_{i1} - Y_{i0})}_{\text{treatment effect}} D_i$$

We are interested in the (individual) **treatment effect** $\Delta_i = Y_{i1} - Y_{i0}$.

The Fundamental Problem of Causal Inference

We cannot identify the individual treatment effect $\Delta_i = Y_{i1} - Y_{i0}$

This is **logically impossible**

- we either observe that a unit was treated or not
- but never both treatment statuses at the same time

The Fundamental Problem of Causal Inference

A **critical ingredient** to establish causality: the **counterfactual**

What would have happened to a unit if

- the treatment status $D \in \{0, 1\}$ was different?
- the treatment intensity was different?

The **counterfactual** is entirely **hypothetical**, but we cannot make causal claims without it!

Statistical solution: **Average Treatment Effect** (**ATE**) for a random unit

$$E(\Delta_i) = E(Y_{i1} - Y_{i0}) = E(Y_{i1}) - E(Y_{i0})$$

If a **random unit was treated**, the **expected difference in their outcome** is the ATE

The Fundamental Problem of Causal Inference

To see how the **ATE can(not) be estimated from observational data**, it is useful to consider a **hypothetical effect**

The **Average Treatment Effect on the Treated** (**ATT**)

$$\begin{aligned} ATT = E(\Delta_i | D_i = 1) &= E(Y_{i1} - Y_{i0} | D_i = 1) \\ &= E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1). \end{aligned}$$

Interpretation: Average difference in potential outcomes for those who were treated

The Average Treatment Effect on the Untreated (ATU)

Similarly, we can define the **Average Treatment Effect on the Untreated** (**ATU**) as:

$$\begin{aligned} ATU = E(\Delta_i | D_i = 0) &= E(Y_{i1} - Y_{i0} | D_i = 0) \\ &= E(Y_{i1} | D_i = 0) - E(Y_{i0} | D_i = 0). \end{aligned}$$

Interpretation: the **average difference in potential outcomes** for those who were **not treated**

The Fundamental Problem of Causal Inference

Example for ATE, ATT, and ATU

Person	D	Treated	Untreated	Causal_Effect
1	1	80	60	20
2	1	75	70	5
3	0	70	60	10
4	1	85	80	5
5	0	75	70	5
6	0	80	80	0
7	0	90	100	-10
8	1	85	80	5

$ATE = 5$: Average of the last column
$ATT = 8.75$: Average of the green cells
$ATU = 1.25$: Average of the white cells

The Average Treatment Effect

Person	D	PO Treated Y^1	PO Treated Y^0	Causal Effect
1	1	80	60	20
2	1	75	70	5
3	0	70	60	10
4	1	85	80	5
5	0	75	70	5
6	0	80	80	0
7	0	90	100	-10
8	1	85	80	5

The Average Treatment Effect

Person	D	PO Treated Y^1	PO Treated Y^0	Causal Effect
1	1	80	60	20
2	1	75	70	5
3	0	70	60	10
4	1	85	80	5
5	0	75	70	5
6	0	80	80	0
7	0	90	100	-10
8	1	85	80	5

The Average Treatment Effect

ATE/ATT/ATU: Which Parameter is most Relevant?

There is **no clear answer to this question**

ATE is the most general parameter

- what if we give the average person/firm/unit a treatment
- This is interesting for medical trials and for many policy questions

ATT is often interesting for policy evaluation

- take those who took up the policy
- what would have happened to them if they had not taken up the policy

ATU is sometimes interesting for policy evaluation

- We may be concerned about the people who did not take up the policy
- How would they be affected if they took up the policy

But: We **need all three to interpret the estimation results**

Comparison by Treatment Status

In many applications, we want to **estimate the ATT**

But we **only observe**

- Whether a unit was treated or not
- The actual outcome of the unit

Solution (?): **comparison of means/simple difference in outcomes**, which can be estimated from two samples of data (treatment and control group)

$$\begin{aligned} SDO &= E(Y_i|D_i = 1)) - E(Y_i|D_i = 0)) \\ &= \frac{1}{N_T}\sum_i (y_i \mid d_i = 1) - \frac{1}{N_C}\sum_i (y_i \mid d_i = 0) \end{aligned}$$

In **out example**, $SDO =$

Comparison by Treatment Status

Problem with SDO: units may **select into treatment**

$$
\begin{aligned}
E(Y_i|D_i=1) &- E(Y_i|D_i=0) \\
&= E(Y_{i1}|D_i=1) - E(Y_{i0}|D_i=0) \\
&= \underbrace{E(Y_{i1}|D_i=1) - E(Y_{i0}|D_i=1)}_{ATT} + \underbrace{E(Y_{i0}|D_i=1) - E(Y_{i0}|D_i=0)}_{\text{Selection bias}}
\end{aligned}
$$

Selection bias: the potential outcomes would differ even if both groups were untreated

Note how we add and subtract $E(Y_{i0}|D_i=1)$ here

Now Suppose We Want to Estimate the ATE

Note that we can write the ATE as: $ATE = \pi ATT + (1 - \pi)ATU$

The **simple difference in outcomes yields**:

$$\begin{aligned} E(Y_i|D_i = 1) - E(Y_i|D_i = 0) &= \\ &= ATE \\ &+ \underbrace{E(Y_{i0}|D_i = 1) - E(Y_{i0}|D_i = 0)}_{\text{Selection bias}} \\ &+ \underbrace{(1 - \pi)\left[ATT - ATU\right]}_{\text{HTE Bias}} \end{aligned}$$

Our estimate of the ATE is contaminates by two biases:

- selection into treatment
- heterogeneous treatment effects

Comparison by Treatment Status

But in most cases, **we run regressions** rather than comparing means

$$\begin{array}{ccccccc} Y_i = & \alpha & + & \beta & D_i & + & \varepsilon_i \\ & = & & = & & & = \\ & E(Y_{i0}) & & (Y_{i1} - Y_{i0}) & & & Y_{i0} - E(Y_{i0}) \end{array}$$

Take expectations conditional on D_i:

$$\begin{aligned} E(Y_i|D_i = 1) &= \alpha + \beta + E(\varepsilon_i|D_i = 1) \\ E(Y_i|D_i = 0) &= \alpha + E(\varepsilon_i|D_i = 0) \end{aligned}$$

Comparison by Treatment Status

The OLS estimate of β is

$$\begin{aligned} E(Y_i|D_i=1) \quad &- \quad E(Y_i|D_i=0) \\ &= \underbrace{\beta}_{\text{treatment effect}} + \underbrace{E(\varepsilon_i|D_i=1) - E(\varepsilon_i|D_i=0)}_{\text{selection bias}}. \end{aligned}$$

Comparison by Treatment Status

Selection bias in theory

$$E(Y_{i0}|D_i = 1) - E(Y_{i0}|D_i = 0) = E(\varepsilon_i|D_i = 1) - E(\varepsilon_i|D_i = 0)$$

- Treated and untreated units don't have the same potential outcomes

In practice, put simply: **treated and untreated units are different**

Selection bias is pervasive because **people choose what's best** for them

- There is simply no reason to believe the bias is zero
- Causal inference provides techniques to eliminate selection bias under assumptions

Randomized Experiments

Experiments provide a **clean way to eliminate selection bias**

Even if in most cases it is **not possible to run experiments**, a **clean experiment** serves as the **benchmark** for all the methods in this course

Idea: the closer we get to the **ideal experimental setting**, the closer we get to estimating a **causal effect**

Randomized Experiments

Basic idea: **randomly assign treatment** across units

Two conditions have to be fulfilled:

- **assignment** to treatment and control group is **random**
- there is **full compliance** with the treatment (plus: no attrition)

In that case, the **treatment and control group** are **statistically identical**

Randomized Experiments

Formally (T: treatment group, C: control group)

$$E\{Y_{i0}|i \in C\} = E\{Y_{i0}|i \in T\}$$

and

$$E\{Y_{i1}|i \in C\} = E\{Y_{i1}|i \in T\}.$$

- ... both groups have the **same potential outcomes** in expectation

Randomized Experiments

Therefore, we can obtain an **unbiased and consistent estimate of the ATE**

$$E(\Delta_i) = E(Y_{i1} - Y_{i0}) = E\{Y_{i1}|i \in T\} - E\{Y_{i0}|i \in C\}.$$

$\Rightarrow$ **comparison of means** is meaningful (causal effect in expectation)

$\Rightarrow$ Randomization solves the **fundamental problem of causal inference**

Randomized Experiments

But **what if not everyone complies** with the treatment?

Examples

- Not every job seeker who is offered training takes part
- Not every person eligible for a medical card/social benefits/etc applies

$$E\{Y_{i1}|i \in T\} - E\{Y_{i0}|i \in C\} \neq ATE$$

As we will learn later in the course, **all is not lost**

- The resulting **treatment effect can still be meaningful**
- . . . as long as **treatment is random**

Randomized Experiments

An **important assumption** underlying causal inference in experiments

SUTVA: **S**table **U**nit **T**reatment **V**alue **A**ssumption

In **plain English**

- treatment of one unit must not affect the potential outcomes of another
- i.e. no spillovers, general equilibrium effects, etc

Randomized Experiments

SUTVA is an **untestable assumption**

Examples for **violations**:

- information leaks from treatment to control group
- large-scale job training programs $\Rightarrow$ GE effects
- health interventions (capacity constraints in medical facilities)

Implications:

- One reason why small interventions don't scale up
- Important to choose the right control group

Cookbook for Analyzing a Randomized Experiment

1) Explain experimental design in detail, in particular

- How was the randomization carried out
- Discuss compliance (or likely non-compliance)
- Argue why SUTVA holds

2) Show balancing tests based on **pre-treatment characteristics**

- Do treatment and control group differ before the experiment?
- Use pre-treatment outcomes if possible
- Use other pre-treatment characteristics that may predict selection into treatment
- If you use regression analysis in the paper, regress the pre-treatment x on the treatment
- Never use post-treatment outcomes (NEVER EVER!)

$$x_i = \delta_1 + \delta_2 D_i + \eta_i$$

Cookbook for Analyzing a Randomized Experiment

3) Show and discuss results

- **Compare means** across treatment and control groups
- Add **pre-treatment characteristics** X_i as controls

$$y_i = \alpha + \beta D_i + \boldsymbol{X_i}\boldsymbol{\gamma} + u_i$$

Why add controls?

- Additional test if randomization worked (β should not change)
- Estimates become more precise (less noise in the model $\Rightarrow$ lower SE)

Example: Tennessee STAR Experiment

Research Question: **does class size matter** to student learning? **(Krueger, 1999)**

Implemented on cohort of kindergartners (i.e. senior infants) in 1985/86 in Tennessee.

Lasted 4 years, then everyone went back to regular size class.

Three treatments

1. Small class 13-17
2. Regular class 22-25
3. Regular class with Aide

Example: Tennessee STAR Experiment

Schools had to have **at least 3 classes** to participate.

Entering cohort **randomly assigned to class type**. Teachers also randomly assigned.

I.e. randomization was carried out **within schools**

Test scores measured towards **end of each school year** in March.

Example: Tennessee STAR Experiment

Balancing tests: do treatment and control groups have the **same pre-treatment characteristics?**

TABLE 2.2.1
Comparison of treatment and control characteristics in the Tennessee STAR experiment

Variable	Class Size			*P*-value for equality across groups
	Small	Regular	Regular/Aide	
Free lunch	.47	.48	.50	.09
White/Asian	.68	.67	.66	.26
Age in 1985	5.44	5.43	5.42	.32
Attrition rate	.49	.52	.53	.02
Class size in kindergarten	15.10	22.40	22.80	.00
Percentile score in kindergarten	54.70	48.90	50.00	.00

Notes: Adapted from Krueger (1999), table I. The table shows means of variables by treatment status for the sample of students who entered STAR in kindergarten. The *P*-value in the last column is for the *F*-test of equality of variable means across all three groups. The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow-up before completing third grade.

- They are similar in age, race and SES
- Attrition rates are similar
- Treatment and outcomes differ

Example: Tennessee STAR Experiment

Krueger (1999) uses a **regression to analyse the experiment**

The **most comprehensive specification** includes **pre-treatment characteristics** and **school fixed effects**

$$y_{is} = \alpha + \beta D_{is} + \boldsymbol{X}_{is}'\boldsymbol{\gamma} + \delta_s + u_{is}$$

Why include fixed effects?

- **Randomization** was carried out **within schools**
- Children in different school types may **react differently to the treatment**
- The FE estimator compares children within the same school

Example: Tennessee STAR Experiment

TABLE 2.2.2
Experimental estimates of the effect of class size on test scores

Explanatory Variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian	—	—	8.35	8.44
			(1.35)	(1.36)
Girl	—	—	4.48	4.39
			(.63)	(.63)
Free lunch	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Teacher Master's degree	—	—	—	−0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Notes: Adapted from Krueger (1999), table V. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. The sample size is 5,681.

Example: Tennessee STAR Experiment

Result: **smaller classes are more effective**

Experimental design appears valid

- Balancing tables point to clean randomization
- Estimates not affected by controls for pre-treatment characteristics

The **same cookbook approach** can be used for **non-experimental studies**

Discrete vs. continuous treatment

So far, we discussed experiments with a **discrete treatment**

The **same assumptions apply** to experiments with a **continuous treatment**

- The **treatment intensity varies** across units
- treatment intensity is **randomly assigned**

Possibilites for researchers:

- **Discretize** and compare mean outcomes (e.g. above/below median)
- Estimate **marginal effect in a regression**

Randomized Experiments as Templates

Randomized experiments are often difficult or impossible to conduct

But they serve as a **template for estimating causal effects** in non-experimental settings

Any study that **claims to estimate a causal effect** should

- explain what the treatment is
- and under what conditions the assignment of the treatment (intensity) is as good as random
- . . . even if the world is not perfect, there is no harm thinking about the perfect

My advice (after 100+ referee reports, several published papers and many rejections): **ignore the experimental template at your own peril**

References

Angrist, Joshua, & Pischke, Jörn-Steffen. 2009. *Mostly Harmless Econometrics - An Empiricist's Companion*. Princeton University Press.

Cunningham, Scott. 2020. *Causal Inference: The Mixtape*. Yale University Press.

Krueger, Alan B. 1999. Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, **114**(2), 497–532.

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

Appendix

Group Work I

ATE, ATT, ATU

Group Work II

Derivation of SDO Decomposition

$$
\begin{aligned}
&E(Y^1 \mid D=1) - E(Y^0 \mid D=0) \\
&\quad = E(Y^1 \mid D=1) - E(Y^1 \mid D=0) + E(Y^1 \mid D=0) - E(Y^0 \mid D=0) \\
&\quad = \underbrace{E(Y^1 \mid D=1) - E(Y^1 \mid D=0)}_{\text{ATT}} + \underbrace{E(Y^1 \mid D=0) - E(Y^0 \mid D=0)}_{\text{Selection Bias}}
\end{aligned}
$$

Derivation of SDO Decomposition

Now consider $ATE = \pi ATT + (1 - \pi)ATU$

$$\begin{aligned} ATE &= \pi ATT + (1 - \pi)ATU \\ &= \pi ATT + ATT - ATT + (1 - \pi)ATU \end{aligned}$$

Re-arranging yields

$$\begin{aligned} ATT &= ATE + (1 - \pi)ATT - (1 - \pi)ATU \\ &= ATE + (1 - \pi)[ATT - ATU] \end{aligned}$$

Derivation of SDO Decomposition

We can plug this into the SDO decomposition and get

$$
\begin{aligned}
E(Y^1 \mid D=1) - E(Y^0 \mid D=0) & \\
&= ATE \\
&+ \underbrace{E(Y^1 \mid D=0) - E(Y^0 \mid D=0)}_{\text{Selection Bias}} \\
&+ \underbrace{(1-\pi)[ATT - ATU]}_{\text{HTE Bias}}
\end{aligned}
$$

benjamin.elsner@ucd.ie

www.benjaminelsner.com

Sign up for office hours

YouTube Channel

@ben_elsner

LinkedIn

Contact

Prof. Benjamin Elsner
University College Dublin
School of Economics
Newman Building, Office G206
benjamin.elsner@ucd.ie

Office hours: book on Calendly