Causal Inference

Introduction: Causality and the Fundamental Problem of Causal Inference

Ken Stiller

University of Sarajevo

October 2024

Part 1

Dr Ken Stiller

I'm a postdoctoral Max Weber Fellow at the EUI's School of Transnational Governance. I received a PhD in International Relations from the Department of Politics of International Relations at the University of Oxford, where I was affiliated with Nuffield College.

Causal Inference

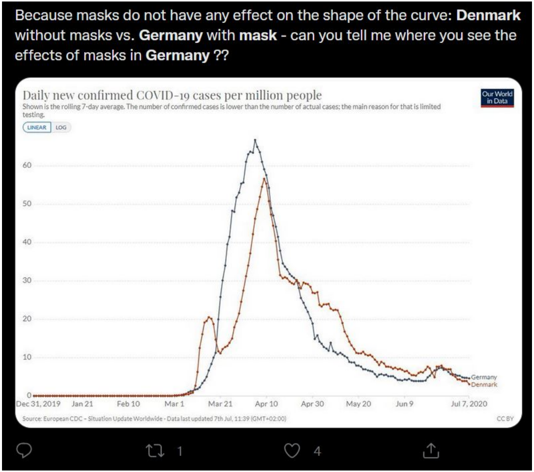
"Felix, qui potuit rerum cognoscere causas"

- Virgil, 29BC.

Why Causal Inference?

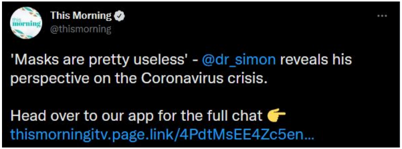
- Our world is full of causal claims!
- This is inevitable!
- Can you think of anything that relates to knowledge that does not rely on a chain of cause and effect?
- Can we answer questions about the (social) world, without making assumptions about whether X and Y are causally related?
- It will make you a better voter, informed citizen, allow you to make better decisions, and differentiate fake news.

Do Masks Protect You Against Covid?



One (among too many!) weird, weird thoughts on Twitter.

Do Masks Protect You Against Covid? II



Backed up by a real doctor - seems legit.

Do Masks Protect You Against Covid? III

So they do help? Probably, but how can we know?

Do Masks Protect You Against Covid? IV

RESEARCH ARTICLE | APPLIED PHYSICAL SCIENCES | OPEN ACCESS

An upper bound on one-to-one exposure to infectious human respiratory particles

Gholamhossein Bagheri, Birte Thiede, Bardia Hejazi, +1, and Eberhard Bodenschatz Authors Info & Affiliations

December 2, 2021 | 118 (49) e2110117118 | https://doi.org/10.1073/pnas.2110117118

We're confident at this point - but it is pretty hard to actually know something!

Let's Look at a Hard Case

Why did Russia decide to invade Ukraine (in early 2022)?

Ukraine - Territorial Gains?

Opinion: What we can expect after Putin's conquest of Ukraine

By Robert Kagan

Contributing columnist

February 21, 2022 at 5:39 p.m. EST

Listen to article 6 min

Ukraine - Domestic Reasons?


Opinion Russian politics

Putin's threats disguise a weakening position

A regional sphere of influence made up of corrupt autocracies is vulnerable to uprisings and instability

GIDEON RACHMAN

Ukraine - Identity?

President of Russia | Events | Structure | Videos and Photos | Documents | Contacts | Search

Events

Article by Vladimir Putin ”On the Historical Unity of Russians and Ukrainians“

July 12, 2021 17:00

Ukraine -Security Concerns?

Why the Ukraine Crisis Is the West's Fault

The Liberal Delusions That Provoked Putin

John J. Mearsheimer

Ukraine - Ukraine's Agency?

NATO recognises Ukraine as Enhanced Opportunities Partner

12 Jun. 2020 - | Last updated: 12 Jun. 2020 15:47

English | French | Russian | Ukrainian

On Friday (12 June 2020), the North Atlantic Council recognised Ukraine as an Enhanced Opportunities Partner. This status is part of NATO's Partnership Interoperability Initiative, which aims to maintain and deepen cooperation between Allies and partners that have made significant contributions to NATO-led operations and missions.

Why did Russia decide to invade Ukraine (in Feb 22)?

For all these reasons?

- Possibly to some extent
- But this is not really a satisfactory or helpful answer

- *We seek to make accurate, precise and 'helpful' (probabilistic) claims*

Why Do We Do Science?

- To *explain* things?
  - Why did something happen?
  - $? \to Y$
- To *predict* things?
  - What will happen?
  - $X \to ?$
- Are these two mutually exclusive?
  - It's always about the relationship between two events (i.e. causation).
  - $X \to Y$
- Is it much better to know a small part about everything than everything about a small part?
  - This - again - is a false dichotomy!
  - Looking at causality might (help) overcome it

Why Do We Do Science?

- To *explain* things?
  - Why did something happen?
  - $? \rightarrow Y$
- To *predict* things?
  - What will happen?
  - $X \rightarrow ?$
- Are these two mutually exclusive?
  - It's always about the relationship between two events (i.e. causation).
  - $X \rightarrow Y$
- Is it much better to know a small part about everything than everything about a small part?
  - This - again - is a false dichotomy!
  - Looking at causality might (help) overcome it

Why Do We Do Science?

- To *explain* things?
  - Why did something happen?
  - $? \rightarrow Y$
- To *predict* things?
  - What will happen?
  - $X \rightarrow ?$
- Are these two mutually exclusive?
  - It's always about the relationship between two events (i.e. causation).
  - $X \rightarrow Y$
- Is it much better to know a small part about everything than everything about a small part?
  - This - again - is a false dichotomy!
  - Looking at causality might (help) overcome it

Why Do We Do Science?

- To *explain* things?
  - Why did something happen?
  - $? \to Y$
- To *predict* things?
  - What will happen?
  - $X \to ?$
- Are these two mutually exclusive?
  - It's always about the relationship between two events (i.e. causation).
  - $X \to Y$
- Is it much better to know a small part about everything than everything about a small part?
  - This - again - is a false dichotomy!
  - Looking at causality might (help) overcome it

Why Do We Do Science?

- To *explain* things?
  - Why did something happen?
  - $? \rightarrow Y$
- To *predict* things?
  - What will happen?
  - $X \rightarrow ?$
- Are these two mutually exclusive?
  - It's always about the relationship between two events (i.e. causation).
  - $X \rightarrow Y$
- Is it much better to know a small part about everything than everything about a small part?
  - This - again - is a false dichotomy!
  - Looking at causality might (help) overcome it

Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form "what is the effect of X on Y"
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents' education on children's income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but often careless interpretations.
- Being reasonably confident about a causal effect is usually not that easy

Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form "what is the effect of X on Y"
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents' education on children's income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but often careless interpretations.
- Being reasonably confident about a causal effect is usually not that easy

Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form "what is the effect of X on Y"
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents' education on children's income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but often careless interpretations.
- Being reasonably confident about a causal effect is usually not that easy

Welcome to Causal Inference

Setup of the short course:

- 4 Days
- We'll discuss theory & background
- You'll receive scripts that implement methods in R
- What do you care about most?

- Make sure to sign up online and on the attendance sheet if you want to receive a **course certificate** from the EUI.

Syllabus: Causal Inference

1. Introduction
2. Potential Outcomes Framework and Selection Bias
3. Selection on Observables, Multiple Regression and Matching
4. Panel Data and Fixed Effects Models
5. Differences-In-Differences
6. Instrumental Variables and 2SLS Estimation
7. Regression Discontinuity Designs

Where to find our course materials

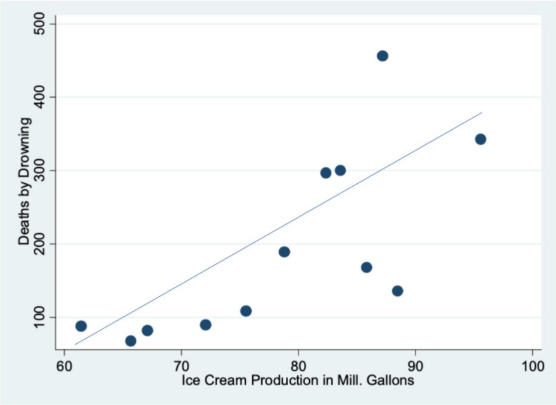
- You'll find all resources (slides, lab code) on our course page: `https://widening-sarajevo.github.io/CI/`

Extra Assignment

- The world is full of incorrect causal claims (unfortunately).
- Look out for them!
- If you find a really bad one, send them to me by email.
- Best one (i.e. most ridiculous incorrect claim) gets an honorary mention at the end of the week for this term!

Table of Contents

What is the Effect of Ice Cream Production on Drowning?

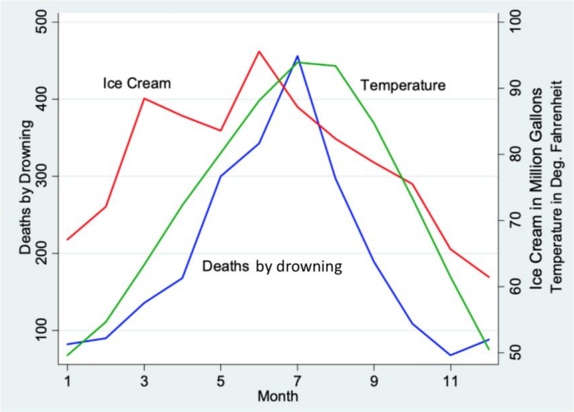


Monthly data for the USA in 2004 – a strong relationship between ice cream sales and drownings

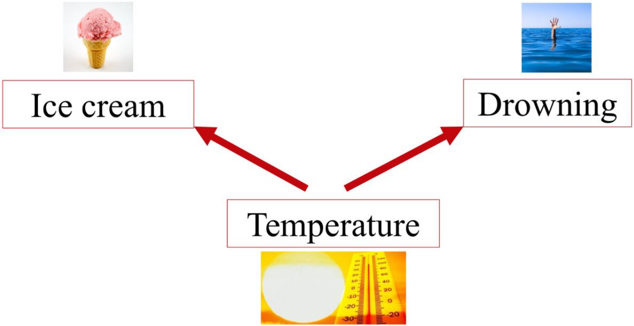
What Causes Drownings?

1. Ice cream
2. Lots of swimming in High temperatures
3. Irresponsible parents
4. Something else
5. I don't know

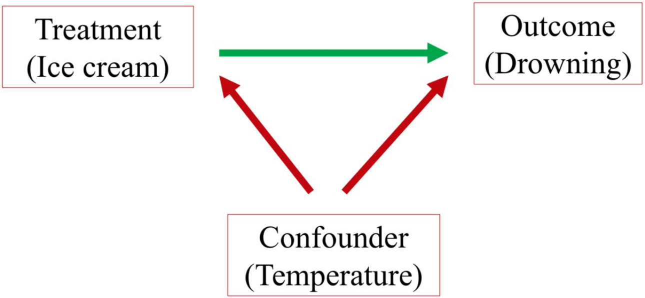
Drowning, Ice Cream, and the Heat



Correlation is Not Causation: A Third Variable can Cause Both



Correlation is Not Causation: A Third Variable can Cause Both



Definitions

Causality

Refers to the relationship between events where one set of events (the effects) is a direct consequence of another set of events (the causes). (Hidalgo & Sekhon 2012)

Causal Inference

The process by which one can use data to make claims about causal relationships. (Hidalgo & Sekhon 2012)

Inferring causal relationships is a central task of science.

Examples

- What is the effect of peace-keeping missions on peace?
- What is the effect of church attendance on social capital?
- What is the effect of minimum wage on employment?

A Counterfactual Logic

Counterfactual Logic

If X had/had not been the case, Y would/would not have happened

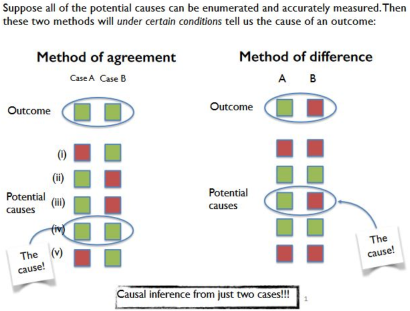
Example:*Does college education increase earnings?*

- If high school grads had instead obtained a college degree, how much would their income change?
- If college grads had only obtained a high school diploma, how much would their income change?

John Stuart Mill: A System of Logic (1843)

A series of "canons" for inductive inference. Two well-known methods of comparison:

- Method of Difference: Units similar in all respects but one manipulable treatment. Units differ in their outcomes.
- (Method of Indifference: Units dissimilar in all respects but one. Outcome same in both units.)



Mill & the Counterfactual Logic of Causality

Causal Inference as a counterfactual Problem

Rather than defining causality purely in reference to observable events, counterfactual models define causation in terms of a comparison of observable and unobservable events.

- We need to construct counterfactuals of the observed world as comparison units (Method of Difference).

Mill & the Counterfactual Logic of Causality

Causal Inference as a counterfactual Problem

Rather than defining causality purely in reference to observable events, counterfactual models define causation in terms of a comparison of observable and unobservable events.

- We need to construct counterfactuals of the observed world as comparison units (Method of Difference).

A hypothetical example

Imagine two students who are interested in getting a very high score on their thesis. They are considering the courses they should take and they are undecided between *Causal Inference* or some other class.

Y_i : Thesis score is the outcome variable of interest for unit i.

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment (taking Causal inference)} \\ 0 & \text{otherwise.} \end{cases}$$

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential thesis score for student } i \text{ with Causal Inference} \\ Y_{0i} & \text{Potential thesis score for student } i \text{ without Causal Inference} \end{cases}$$

Q: What is the effect of taking Causal Inference on your thesis score?

Defining the Potential Outcomes

Definition: Treatment

D_i : Indicator of treatment status for unit i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition: Observed Outcome

Y_i : Observed outcome variable of interest for unit i. (Realized after the treatment has been assigned)

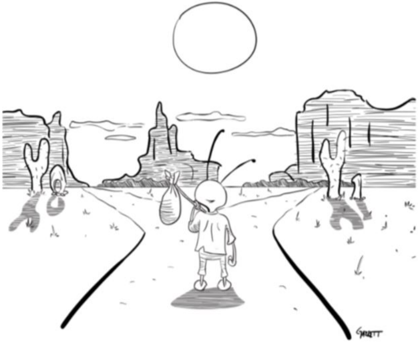
Defining the Potential Outcomes

Definition: Potential Outcomes

Y_{0i} and Y_{1i}: Potential Outcomes for unit i

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

The Road Not Taken (by Robert Frost)



It shows (1) the actual road that you chose, and (2) the counterfactual road that you could have chosen but did not. **How can we know it made the difference? We don't know what would have happened on the other path.**

Causality with Potential Outcomes

Let D_i denote a binary treatment for unit i, where $D_i \in 0, 1$. Let Y_i represent the observed outcome for unit i. The potential outcomes are thus: Y_{1i}, Y_{0i}

The causal effect of D on Y for i is $\tau_i = Y_{1i} - Y_{0i}$

Definition: Causal Effect

Causal Effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

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

The Fundamental Problem of Causal Inference

The Fundamental Problem of Causal Inference

It is impossible to observe for the same unit i the values $D_i = 1$ and $D_i = 0$ as well as the values Y_{1i} and Y_{0i} and, therefore, it is impossible to observe the effect of D on Y for unit i.

This is why we call this a missing data problem. We cannot observe both potential outcomes, hence we cannot estimate:

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

		Y_{i1}	Y_{i0}
Person 1	Treatment Group ($D = 1$)	Observable as Y	Counterfactual
Person 2	Control Group ($D = 0$)	Counterfactual	Observable as Y

But: We Aim to Make Causal Inference!

In the coming days, we will learn ways to address the FPCI.

- Even though it cannot be fully resolved, we can achieve confidence about our findings
- It is absolutely crucial to always be aware what we can claim - as well as about the limitations of our methods

- Now: The limits of linear regressions in the context of the FPCI and why experiments help

Quantities of Interest

Definition ATE

Average Treatment Effect:

$$\tau_{ATE} = E[Y_1 - Y_0]$$

Definition ATT

Average Treatment Effect of the Treated:

$$\tau_{ATT} = E[Y_1 - Y_0|D = 1]$$

Definition ATC

Average Treatment Effect of the Controls:

$$\tau_{ATC} = E[Y_1 - Y_0|D = 0]$$

An Example: ATE

Imagine a population of 4 units:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	?	?	?
2	1	1	?	?	?
3	0	0	?	?	?
4	1	0	?	?	?

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

An Example: ATE

Imagine a population of 4 units:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	1	0	?	1	?

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

Since we cannot observe two worlds at the same time, we cannot calculate the ATE.

An Example: ATE

Imagine a population of 4 units with the counterfactual values being made up! (as are all other values in this example!)

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	1	0	1	1	0

What is the ATE?

An Example: ATE

Imagine a population of 4 units with the counterfactual values are made up! (As all other values in this example!)

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	1	0	1	1	0

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}] = 4/4 = 1$$

An Example: ATE (Continued)

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

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}] = 1/4 \cdot (3 + 0 + 1 + 0) = 1$$

An Example: Incorrect ATE

In reality you only get the following:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	1	0	?	1	?

Wrong $\tau_{ATE} = E[Y_{1i} - Y_{0i}] = 2 - 0.5 = 1.5$

What is the identification problem?

$$\tau_{ATE} = E[Y_1 - Y_0]$$

$$= \pi(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1])$$
$$+ (1\text{-}\pi)(E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0])$$

where π is the share of the treated units in our sample.

What can we observe from the above equation?

1. ?
2. ?
3. ?

What can't we observe from the above equation?

1. ?
2. ?

What is the identification problem?

$$\tau_{ATE} = E[Y_1 - Y_0]$$

$$= \pi(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1])$$
$$+ (1\text{-}\pi)(E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0])$$

where π is the share of the treated units in our sample.

What can we observe from the above equation?

1. π
2. $E[Y_{1i}|D_i = 1]$
3. $E[Y_{0i}|D_i = 0]$

What can't we observe from the above equation?

1. $E[Y_{0i}|D_i = 1]$
2. $E[Y_{1i}|D_i = 0]$

Counterfactual outcomes!

What is the identification problem?

The observed difference in the outcome for the treatment and control group are:

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

$$\mathrm{E}[\mathrm{Y}_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] =$$

$$E[Y_{i1} - Y_{i0}|D_i = 1] + E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]$$

What is the identification problem?

The observed difference in the outcome for the treatment and control group are:

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

$$\mathrm{E}[\mathrm{Y}_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] =$$

$$\underbrace{E[Y_{i1} - Y_{i0}|D_i = 1]}_{ATT} + \underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{SelectionBias}$$

- ATT: Average treatment effect on the treated
- Selection Bias: Differences in the treated and control groups when assigned to the control group.

Both are unobserved and we need to make assumptions!