Impact Evaluation Methods

Topic 1: Introduction to Causal Inference

Alex Alekseev

April 28, 2025

University of Regensburg, Department of Economics

What Is Causal Inference?

What Is Causal Inference?

- Causal inference is the art and science of estimating causal relationships
- John Snow (1855), Ronald Fisher (1935), Trygve Haavelmo (1943), Donald Rubin (1974)
- Labor economists from the late 1970s to late 1990s

Nobel Prize in Economics 2021

David Card ("for his empirical contributions to labour economics") and to Josh Angrist + Guido Imbens ("for their methodological contributions to the analysis of causal relationships")

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

© Nobel Prize Outreach. Photo: Paul Kennedy

David Card

Prize share: 1/2

© Nobel Prize Outreach. Photo: Risdon Photography

Joshua D. Angrist

Prize share: 1/4

© Nobel Prize Outreach. Photo: Paul Kennedy

Guido W. Imbens

Prize share: 1/4

Causal Inference Applications

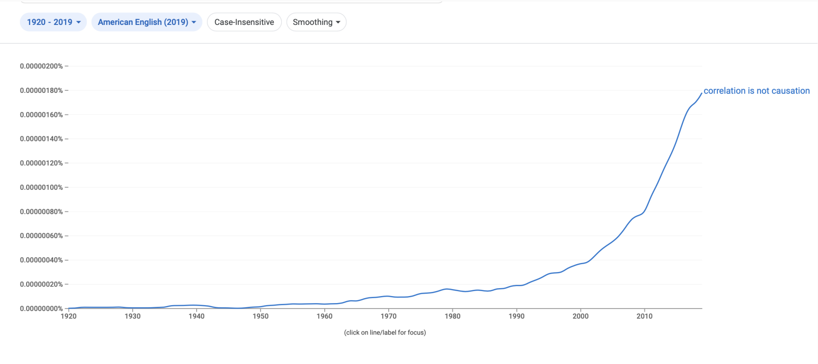
- "Mostly Harmless Econometrics" by Joshua Angrist and Jörn-Steffen Pischke
- Causal inference is not limited to labor economics
- It applies throughout the field of empirical economics: political economy, health economics, environmental economics
- And not just economics: political science, sociology

Mainstream Approach

- Causal inference matured into a distinct field.
- Sometimes reviewed in a chapter on “program evaluation” in econometrics textbooks (Wooldridge 2010)
- Angrist and Pischke (2009) "Mostly Harmless Econometrics," Morgan and Winship (2014) "Counterfactuals and Causal Inference," Imbens and Rubin (2015) "Causal Inference for Statistics, Social, and Biomedical Sciences"
- Specific strategies: Angrist and Krueger (2001) on Instrumental Variables, Imbens and Lemieux (2008) on Regression Discontinuity.

Correlation is not Causation

Correlation is not Causation



What Does It Mean?

One possible definition

Just because two variables are correlated does not necessarily mean that one causes the other

Or...

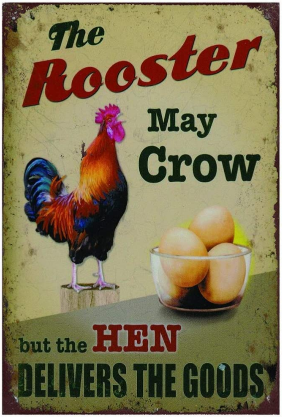
The inability to legitimately deduce a cause-and-effect relationship between two variables on the basis of an observed association between them

Purpose of the Course

To explain why correlations, particularly in observational data, are unlikely to be reflective of causal relationships and how to uncover true causal relationships

Example 1

When the rooster crows, the sun soon after rises, but does it mean that the rooster causes the sun to rise?



Example 2

We can observe that the number of people who wear shorts is much higher on days when people eat ice cream. Does it mean that shorts-wearing causes people to buy ice-cream?

Photo by Choi Mo on Unsplash

Example 3

Sometimes there are causal relationships between variables and yet no observable correlation between them. Consider a central bank that expects a recession.

Example 3, ctd.

- The bank enters into open-market operations
- These operations will show no relationship with actual output
- Banks may engage in aggressive trading, and we would be unable to see any evidence that it was working
- Had the central bank not intervened...
- Would need a potential outcome

Who Cares?

Many research questions are causal

Not interested	**Interested**
if countries with higher minimum wages have less poverty	if raising the minimum wage **reduces** poverty
if people who take a popular common-cold-shortening medicine get better	if the medicine made them get better **more quickly**
if the central bank raising interest rates was shortly followed by a recession	if the interest rate increase **caused** the recession

What is Causality?

A useful way to think about it

We can say that **X** causes **Y** if, were we to intervene and change the value of **X**, then the distribution of **Y** would also change as a result

Using the Definition

- If a fox eats the rooster, would it mean that the sun would no longer rise?
- If we were to swap out someone's pants for shorts, would it make them more likely to eat ice cream?
- If we were to stop the central bank and not let it engage in the open-market operations, would it lead to a change in GDP?

Highway Example

Research question

Does adding an additional highway lane reduce traffic?

- Compare traffic patterns on three-lane highways and on two-way highways
- It turns out that more lanes have **more** traffic
- Why do those highways have more lanes in the first place?
- The busiest routes tend to be the ones that get expanded

Highway Example, ctd.

- We are not interested in how much traffic there is on three-lane highways vs. two-lane highways
- We interested in whether we can make traffic go down by turning a two-lane highway into a three-lane highway
- The numbers we have don't tell us that right away
- We typically don't have a **"what if"** highway
- Find the best possible proxy to that **"what if"** highway

Hospital Example

Research question

Do hospitals make people healthier?

- A poor elderly population that uses hospital emergency rooms for primary care
- Expensive, crowds hospital facilities, not very effective
- Exposure to other sick patients might have a net negative impact on their health

How Can We Answer It?

- Hospitals provide many valuable services
- The ostensible answer to the hospital-effectiveness question: yes (?)
- Let's compare the health status of those who have been to the hospital to the health of those who have not.
- The National Health Interview Survey (NHIS)
- "During the past 12 months, was the respondent a patient in a hospital overnight?"
- "Would you say your health in general is excellent, very good, good, fair, poor?"

Let's Look at The Data

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

The difference in means is 0.71 (t-statistic $= 58.9$)

Should We Believe It?

- It's not impossible this is the right answer
- Hospitals are full of other sick people who might infect us
- However, people who go to the hospital are probably **less healthy** to begin with
- Even after hospitalization people who have sought medical care are not as healthy, on average
- Though they may well be better than they otherwise **would have been**

Training Programs

- Government-subsidized training programs
- Provide a combination of classroom instruction and on-the-job training for groups of disadvantaged workers
- The idea is to increase employment and earnings
- Paradoxically, studies based on non-experimental comparisons of participants and non-participants often show that after training, the trainees **earn less** than plausible comparison groups (see, e.g., Ashenfelter, 1978; Ashenfelter and Card, 1985; Lalonde 1995)

Should We Believe These Results?

- Subsidized training programs are meant to serve men and women with low earnings potential
- Simple comparisons of program participants with non-participants often show lower earnings for the participants
- Evidence from randomized evaluations of training programs generate mostly positive effects (Lalonde, 1986; Orr et al. 1996)

Education Example

- Class size and learning outcomes
- Some studies suggest there is **little or no link** between a class size and student learning
- Perhaps school systems can save money by hiring fewer teachers?
- Should not be taken at face value: weaker students are often **deliberately grouped** into smaller classes.

Human Capital

- The theory says that education has a causal effect on the subsequent labor market earnings
- Educational training provides skills that **increase the productivity** of workers
- Productivity is prized in the labor market, hence firms are willing to **pay educated workers more**

Should We Believe The Result?

- Suppose we find that people with more education **earn higher wages** than people with less education
- Should we take this result at face value?

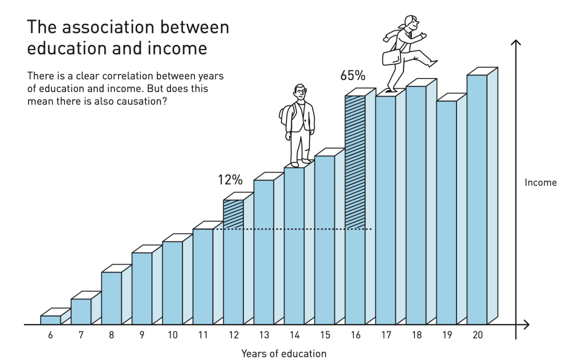


Photo by Johan Jarnestad/The Royal Swedish Academy of Sciences

What Can Be Wrong Here?

- **Ability** enhances productivity as well
- People with relatively high ability are more likely to obtain higher educational degrees
- Highly educated will have higher ability and higher natural rates of productivity
- Some of the uncovered effect of education on earnings may instead reflect innate ability
- “Ability bias” (Card 1999)

Types of Data

Two Types of Data

- Experimental data and non-experimental, or observational, data
- Our focus: observational data.

Cox and Reid (2000)

The word **experiment** is used ... to mean an investigation where the system under study is under the control of the investigator. This means that the individuals or material investigated, the nature of the treatments or manipulations under study and the measurement procedures used are all selected, in their important features at least, by the investigator. By contrast in an **observational study** some of these features, and in particular the allocation of individuals to treatment groups, are outside the investigator's control.

Experimental Data

- Ronald A. Fisher, 1935 book "The Design of Experiments"
- Experimental data come in a variety of different flavors (Harrison and List 2004):
  - a conventional lab experiment
  - an artefactual field experiment
  - a framed field experiment
  - a natural field experiment

Perry Preschool Project

- A 1962 randomized experiment
- Assess the effects of an early-intervention program involving 123 Black preschoolers in Ypsilanti (Michigan)
- The treatment group was randomly assigned to an intensive intervention that included preschool education and home visits
- Follow-up data through 1993 on the participants at age 27
- Dozens of academic studies cite or use the Perry findings (Barnett 1992)
- A basis for the Head Start preschool program

Tennessee STAR Experiment

- A strong and lasting payoff to smaller classes (Finn and Achilles 1990; Krueger 1999)
- Cost about $12 million
- Implemented for a cohort of kindergartners in 1985/86.
- Ran for four years until the original cohort was in the third grade
- About 11,600 children.
- The average class size in regular classes: 22.3
- Treatment group: small classes with 13-17 children
- A positive effect of small classes relative to regular-size classes of about 5 to 6 percentile points

Observational Data

- Collected through surveys, by observing people's behavior in the naturally occurring settings, as a by-product of some other business activity.
- You collect data about what happened previously
- The researcher is a **passive** actor in the processes creating the data itself
- She observes actions and results but is not in a position to **interfere** with the environment

Quasi-Experiments

- Good empirical research using observational data uses data to answer specific causal questions, as if in a randomized controlled trial
- In the absence of a real experiment, we look for well-controlled comparisons and/or natural quasi-experiments

Quasi-Experimental Approach to Class Size

- Angrist and Lavy (1999)
- Observational data is analyzed in an experimental spirit
- In Israel, class size is capped at 40
- Compare cohorts of size **40 and 41**
- Likely to be similar on other dimensions such as ability and family background
- The difference between 40 and 41 students enrolled is "as good as random"

Angrist and Lavy (1999) Results

- Angrist and Lavy (1999) results show a strong link between class size and achievement
- In contrast with naive analyses based on simple comparisons between those enrolled in larger and smaller classes
- Students in smaller classes seem to do worse on standardized tests

Research Design Questions

Research Design Questions

1. The relationship of interest
2. The ideal experiment
3. The identification strategy
4. The mode of inference

Relationship of Interest

- Useful for making predictions about the consequences of changing circumstances or policies
- What would happen in alternative (or "counterfactual") worlds
- For example, the causal effect of schooling on wages (Card, 1999)
- It is useful for predicting the earnings consequences of changing the costs of attending college or strengthening compulsory attendance laws
- This relationship is also of theoretical interest

Not Just Individuals

- Causal questions can be asked about individuals, firms, countries
- Acemoglu, Johnson, and Robinson (2001), Acemoglu et al. (2019) research on the effect of colonial institutions on economic growth
- Whether countries that inherited more democratic institutions from their colonial rulers later enjoyed higher economic growth as a consequence
- Implications for understanding of history and for the consequences of contemporary development policy

Ideal Experiment

- The experiment that could ideally be used to capture the causal effect of interest
- For example, we can imagine offering potential dropouts a reward for finishing school
- We might like to go back in time and randomly assign different government structures to former colonies on their Independence Days
- Ideal experiments are most often hypothetical
- Worth contemplating because they help us pick fruitful research topics

Old Idea

Haavelmo (1944)

A design of experiments ... is an essential appendix to any quantitative theory. And we usually have some such experiment in mind when we construct the theories, although-unfortunately-most economists do not describe their design of experiments explicitly. If they did, they would see that the experiments they have in mind may be grouped into two different classes, namely, (1) experiments that we should like to make to see if certain real economic phenomena when artificially isolated from "other influences" would verify certain hypotheses, and (2) the stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. In both cases the aim of the theory is the same, to become master of the happenings of real life.

Good Research Question

- A good research question **can be answered**, the answer will **improve your understanding** of how the world works
- It's possible to find some evidence in the world that your question would have a believable answer to
- The research question, once answered, should tell you about something broader than itself
- Takes us from theory to hypothesis, helps improve your **why** explanation

Identification Strategy

- aka **empirical strategy**
- Angrist and Krueger (1999) used the term **identification strategy** to describe the manner in which a researcher uses observational data to approximate a real experiment
- Angrist and Krueger (1991) used the interaction between compulsory attendance laws in American schools and students' season of birth as a natural experiment
- Season of birth affects the degree to which high school students are constrained by laws allowing them to drop out on their birthdays

Mode of Inference

- Rubin (1991): what is your mode of statistical inference?
- The population to be studied, the sample to be used, and the assumptions made when constructing standard errors
- Sometimes straightforward (Census micro-data samples to study the American population)
- Often more complex, especially with data that are clustered or grouped

Data Generating Process and Identification

DGP

- What is the data generating process?
- The laws that work behind the scenes
- We can't see them directly, but we see the data that result from them
- We can learn what these laws are from empirical observation

Variation

- Thinking about the data generating process gives us two ideas
- The first is the idea of **looking for variation**
- DGP shows us all the different processes working behind the scenes that give us our data
- But we are interested in part of that variation
- How can we find the variation we need and focus just on that?

More on Variation

- How to answer our research question = where your variation is
- Unlikely that the variation in the raw data answers the question (correlation is not causation)
- What variation needs to be removed?

Identification

- How can we use the DGP to be sure that the variation we are digging out is the right variation?
- **Identification** is finding where the variation you're interested in is and isolating **just that part** so you know that you're answering your research question
- It's called identification because we've ensured that our calculation identifies a single theoretical mechanism of interest

Identification vs. Statistical Problems

Manski (1995)

It is useful to separate the inferential problem into statistical and identification components. Studies of identification seek to characterize the conclusions that could be drawn if one could use the sampling process to obtain an unlimited number of observations. Identification problems cannot be solved by gathering more of the same kind of data. The study of identification logically comes first. Negative identification findings imply that statistical inference is fruitless: it makes no sense to try to use a sample of finite size to infer something that could not be learned even if a sample of infinite size were available. Positive identification findings imply that one should go on to study the feasibility of statistical inference.

From Hypothesis to Data

- A research question takes us from theory to hypothesis
- Identification takes us from hypothesis to the data

Context

- Identification requires statistical procedures in order to properly get rid of the kinds of variation we don't want.
- It relies on theory and assumptions about the DGP
- We need to make a claim about what we **already know** in order to have any hopes of learning something new

Identification Process

1. Using theory, paint the most accurate picture possible of what the data generating process looks like
2. Use that data generating process to figure out the reasons our data might look the way it does that don't answer our research question
3. Find ways to block out those alternate reasons and dig out the variation we need

R Demo: Human Capital and Ability Bias

Let's Create Our Own DGP

1. Ability is uniformly distributed on [0, 1]
2. Each unit of ability increases your income by 20 thousand euros per year
3. To decide whether you go to college or not, you first flip a coin (heads = go to college, tails = do not go to college). However, if it is tails, you then check your ability. If it is higher than a threshold of 0.5, you go to college anyway.
4. College degree increases your income by 10 thousand euros per year
5. The base income (no college, zero ability) is normally distributed with a mean of 50 thousand euros per year and a s.d. of 20 thousand
6. We observe a sample of 5000 individuals

Sensitivity