DSE - Data-Driven Economic Analysis
Econometrics Module

Lecture 1 - The Problems of Causal Inference

Michele De Nadai
michele.denadai@unimi.it

Trimester II, 2023

UNIVERSITÀ
DEGLI STUDI
DI MILANO

What is Econometrics?

- Different researchers would give different answers, e.g.
- *"Econometrics is the study of the application of statistical methods to the analysis of economic phenomena" (Tintner, 1953)*
- *"Econometrics is the art and science of using statistical methods for the measurement of economic relations" (Chow, 1985)*
- *"Econometrics is the application of statistical and mathematical methods to the analysis of economic data, with a purpose of giving empirical content to economic theories and verifying them or refuting them" (Maddala, 1992)*

A definition

- In general, Econometrics is the application of Statistics to economic data in order to find **quantitative relations** or to make **forecasts**
- Typical goals of econometric analysis
  - Estimating relationships between economic variables
  - Testing economic theories & hypotheses
  - Forecasting economic variables
  - Evaluating government & business policy

Economic questions

Components of econometric analysis:

- Economic model
- Econometric model
- Appropriate data

Economic models:

- Make assumptions about agents behaviour
- Establish relationships between economic variables
- Provides a priori expectations
- Example: consumer demand theory $\rightarrow$ demand equation where we expect price to be inversely related to quantity demanded

Example: Job training and worker productivity

What is the effect of additional training on worker productivity?

- Formal economic theory is available but not readily implementable

$$\text{wage} = f(\text{educ}, \text{experience}, \text{training})$$

- Functional form of relationship not specified
- Data not specified
- Other factors may be relevant

Example: Job training and worker productivity

To map an economic model into an **econometric model** we need:

- Data
- Make assumptions about the function $f(\cdot)$, select observable variables,...

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{experience} + \beta_3 \text{training} + u$$

- Make assumptions about error term (u)
- Econometric models can then be used for:
  - Testing a priori expectations.
  - Interpreting and quantifying relationships between variables.
  - Are these relationships **causal**?

Data

- Data derive from experiments or observations
  - **Experimental** data come from experiments designed to evaluate a treatment
  - **Observational** (non-experimental) data are collected by observing the "state of nature"
- A dataset is made of N observations over K variables
- The unit of analysis may be an individual, a household, a firm, a country, etc.

Structure of a dataset

- **Cross-section**
  - e.g., GDP of EU states observed in a given quarter
  - e.g., Income of a sample of households observed in a given month
- **Time series**
  - e.g., GDP of a state observed quarterly over a time period
  - e.g., Monthly income of a representative household over a time period
- **Panel**
  - e.g., GDP of EU states observed quarterly over a time period
  - e.g., Monthly income of a sample of households over foor periods

Cross-sectional data

- Data on **multiple units** collected at a **single time period**.
- The number of units coincides with the number of observations, N.
- Example: data on test scores in California school districts

	A	B	C	D
1	District	Mean test score	STR	Exp./student
2	1	691	17.9	6.39
3	2	661	21.5	5.1
4	3	644	18.7	5.5
5	4	648	17.4	7.1
6	...	...	...	...

Time series data

- Data on a **single unit** collected at **multiple periods**.
- The number of observations coincides with the number of time periods.
- To emphasize the time reference this number is often denoted with T rather than N.
- Example: data set on rates of inflation and unemployment in US from the second quarter of 1959 to the fourth quarter of 2004

	A	B	C	D
1	Observation	Date	CPI infl. Rate (%)	Unempl. Rate (%)
2	1	1959	0.7	5.1
3	2	1959	2.1	5.3
4	3	1959	2.4	5.6
5	4	1960	0.4	5.1
6	...	...	...	...

Panel data

- Data on **multiple units** collected at **multiple periods**.
- If there are N units and each unit is observed at T time periods, then the number of observations is $\frac{N}{T}$.
- Example: data set on cigarette consumption, price and taxes for each of the 50 U.S. states for 11 years (time periods)

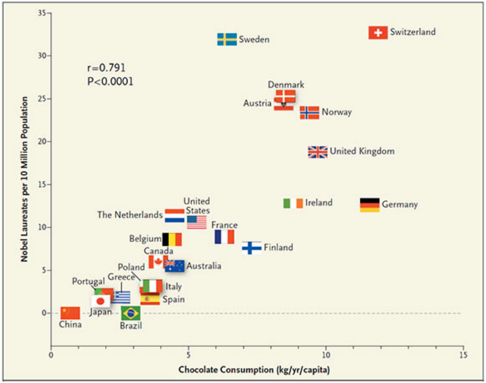
	A	B	C	D
1	State	Date	Consumption	Taxes
2	1	1985	327	14
3	1	1986	351	13
4	...	...	...	...
5	2	1985	253	8
6	...	...	...	...

Causality

- In applied research we are often interested in **causal effects**, as they typically answer policy relevant questions.
- This is different, and sometimes in contrast to, the notion of correlation.
- We (hopefully) all know that **correlation is not causation**.
- **Correlation**: relatively easy to measure, but it is merely a description of the state of nature
- **Causation**: hard to measure, but typically has meaningful policy implications

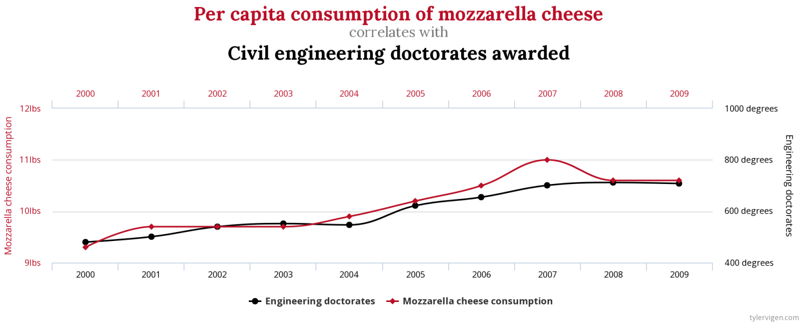
Correlation vs. Causality

- Messerli (*New England Journal of Medicine*, 2012) argues that eating chocolate favors intelligence.
- To support his theory he finds high **correlation** (0.791) between the number of Nobel prize winners in a country and the per-capita annual consumption of chocolate.



Causality

What about this correlation?



Potential Outcome Framework

- Say we are interested in the (causal) effect of college attendance on wages.
- Define a binary treatment variable D_i, e.g. college attendance.
- Potential outcomes (potential wages) are:

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

- These represent the value of wages for individual i in both possible states.
- The **Fundamental Problem of Causal Inference** is that we can never observe both Y_0 and Y_1 for the same individual.

Potential Outcome Framework

- The observed outcome (wage) can we written as:

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

- $Y_{i1} - Y_{i0}$ is the **individual treatment effect**, but we never observe both Y_{i0} and Y_{i1}.
- This implies that the individual treatment effect is **not identified** without additional assumptions.
- What can we identify about the effect of college attendance on wages, given the observable information?
- We might be interested in $E(Y_{i1} - Y_{i0})$, i.e. the **average treatment effect**

Parameters of interest

- Policy relevant questions are not all answered by the same parameter.
- One typical parameter of interest is the **average treatment effect**

$$ATE = E(Y_{i1} - Y_{i0})$$

  that is the average difference in wages in the population with and without college attendance.
- One alternative parameter might be the average effect of college attendance among those who went to college:

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

  this is the **average treatment effect on the treated**.

ATT and Selection Bias

- Given enough information about Y_i and D_i we might aim at estimating some of these causal parameters by looking at the difference between $E[Y_i|D_i = 1]$ and $E[Y_i|D_i = 0]$:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0] \\ &+ E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 1] \\ &= E[Y_{i1} - Y_{i0}|D_i = 1] \\ &+ \underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{\text{selection bias}} \end{aligned}$$

- Meaning: Observed difference in wages = ATT + selection bias
- Causal interpretation is not possible unless $E[Y_{i0}|D_i = 1] = E[Y_{i0}|D_i = 0]$.

Randomization

- In thiese contexts, **randomization** of treatment choices is considered the **gold standard**.
- If treatment D_i were (mean) **independent** of potential outcomes then

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

  which would solve the selection problem.
- In fact independence implies

$$(Y_{i0}, Y_{i1}) \perp\!\!\!\perp D_i$$

- Independent treatment is ruled out whenever individuals **self-select** into treatment based on private (or common) information.
- This is the rule rather than the exception in most practical applications.

Properties of Expectation Operator

- $E[X] = \int_{\mathcal{X}} x f(x) dx$ if X is continuous
- $E[X] = \sum_{k=1}^{K} x_k \Pr(X = x_k)$ if X is discrete
- $E[a] = a$
- $E[a + bX] = a + bE[X]$
- $E[X + Y] = E[X] + E[Y]$
- $E[Y] = E[E[Y|X]]$ (Law of Iterated Expectations)

Selection on Observables

- If individuals self-select into treatment based on **observable characteristics** $(\mathbf{X}_i)$ we can still assume that, conditional on $\mathbf{X}_i$, treatment is "as good as random".
- In general, in many contexts we might assume that

$$(Y_{i0}, Y_{i1}) \perp\!\!\!\perp D_i | \mathbf{X}_i$$

- The conditioning set might include potentially many variables
- This identification strategy is typically referred as **selection on observables**

ATE under Selection on Observables

- Under selection on observables the difference $E(Y_i|D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_i|D_i = 0, \mathbf{X}_i = \mathbf{x})$ identifies:

$$
\begin{aligned}
& E[Y_{i1}|D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_{i0}|D_i = 0, \mathbf{X}_i = \mathbf{x}] \\
+ \quad & E[Y_{i0}|D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_{i0}|D_i = 1, \mathbf{X}_i = \mathbf{x}] \\
= \quad & E[Y_{i1} - Y_{i0}|D_i = 1, \mathbf{X}_i = \mathbf{x}] \\
+ \quad & E[Y_{i0}|D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_{i0}|D_i = 0, \mathbf{X}_i = \mathbf{x}] \\
= \quad & E[Y_{i1} - Y_{i0}|D_i = 1, \mathbf{X}_i = \mathbf{x}]
\end{aligned}
$$

- Knowledge of these differences for each possible value of $\mathbf{x}$ ensures identification of ATT and ATE via the **Law of Iterated Expectations**.

ATT and LIE

- For example if $\mathbf{X}_i$ is gender then:

$$
\begin{aligned}
ATT &= E(Y_{i1} - Y_{i0} | D_i = 1) \\
&= E(Y_{i1} - Y_{i0} | D_i = 1, \mathbf{X}_i = \text{male}) \times \\
&\quad \Pr(\mathbf{X}_i = \text{male} | D_i = 1) \\
&+ E(Y_{i1} - Y_{i0} | D_i = 1, \mathbf{X}_i = \text{female}) \times \\
&\quad \Pr(\mathbf{X}_i = \text{female} | D_i = 1)
\end{aligned}
$$

- Similar arguments can be made for ATE from knowledge of $E(Y_{i1} - Y_{i0} | D_i = 1)$ and $E(Y_{i1} - Y_{i0} | D_i = 0)$.

Conditional Expectation Function

- Identification results are expressed in terms of population quantities, such as the **Conditional Expectation Function**, which we can in general write as as

$$E[Y_i|X_i = x] = m(x)$$

  where x can be discrete or continuous.

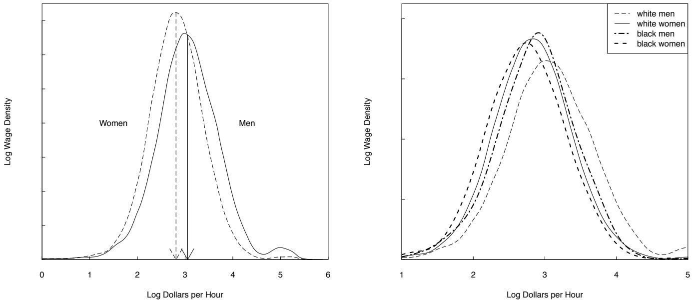
- By properties of expectations we always have

$$Y_i = E[Y_i|X_i = x] + u_i$$

  with u_i mean independent of X_i, i.e. $E[u_i|X_i = x] = 0$.

CEF of wages

- Does the wage distribution vary across subpopulations (gender, race, etc)?
- Can be answered by *Conditional Expectation Function*.



(a) Women and Men

(b) By Sex and Race

Figure 2.3: Log Wage Density by Sex and Race

Linear CEF when X is binary

- If X_i is binary **CEF is linear**, i.e.

$$\text{CEF} = m(x) = a + bx$$

- Example:

$$X_i = \begin{cases} 0 & \text{if individual i is man} \\ 1 & \text{if individual i is woman} \end{cases}$$

- Let $E[Y_i|X_i = 0] = m(0) = \mu_0$ and $E[Y_i|X_i = 1] = m(1) = \mu_1$.
- CEF can be written as:

$$E[Y_i|X_i = x] = \underbrace{\mu_0}_{a} + (\underbrace{\mu_1 - \mu_0}_{b})x$$

Linear vs. non-linear CEF

- We can do similar constructions for variables that can take more than two (but still finite) values.
- BOTTOM LINE: We are interested in CEF. As long as the regressors take a finite set of values, the CEF can be written as a **linear** (in parameters) CEF.
- When the number of regressors is large, we have to count all the cases. This may not be practical.
- When X_i is **continuous** CEF might **not be linear** at all. Still, assumption of linearity is convenient.