Economic Questions and Data

Ming-Ching Luoh

2022.2.15.

Economic Questions

- Economics suggests important **relationships**, often with policy implications, but virtually never suggests **quantitative magnitudes** of **causal effects**.

Examples of economic questions

Question #1: Does reducing class size improve elementary school education?

- Reducing class size costs money: It requires hiring more teachers and building more classrooms.
- To weigh costs and benefits, the decision maker must have a precise quantitative understanding of the likely benefits.

Question #2: Is there racial discrimination in the market for home loans?

- By law, U.S. lending institutions cannot take race into account when deciding to grant or deny a request for a mortgage.
- Using data from early 1990s, researchers found that 28% of black applicants are denied mortgages, while only 9% of white applicants are denied.
- Do these data indicate that there is racial bias in mortgage lending? If so, how large is it?

Question #3: How much do cigarette taxes reduce smoking?

- One of the most flexible tools for cutting smoking consumption is to increase taxes on cigarettes.
- The percentage change in the quantity demanded resulting from a 1% increase is the *price elasticity of demand*.

Causal Effects and Idealized Experiments

- In common usage, an action is said to cause an outcome if the outcome is the direct result, or consequence, of that action.
- Touching a hot stove caused you to get burned.
- Putting fertilizer on your tomato plants causes them to produce more tomatoes.

Estimation of Causal Effects

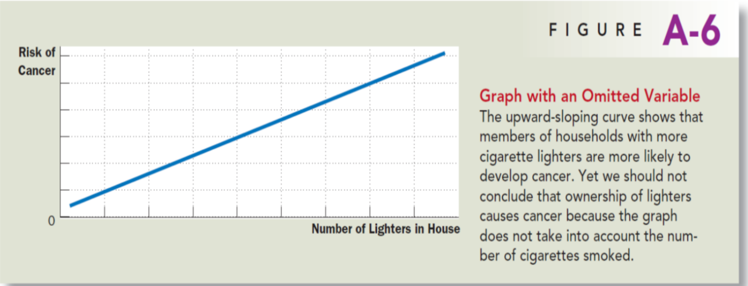
- How might we measure the causal effect on tomato yield (measured in kilograms) of applying a certain amount of fertilizer, say 100 grams of fertilizer per square meter?
- One way to measure this causal effect is to conduct an experiment.
  - Plant many plots of tomatoes.
  - Each plot is tended identically, with one exception: Some plots get 100 grams of fertilizer per square meter, while the rest get none.
  - Whether a plot is fertilized or not is determined randomly by a computer.

- The difference between the average yield per square meter of the treated and untreated plots is the effect on tomato production of the fertilizer treatment.
- This is an example of a idealized controlled experiment. It is controlled in the sense that there are both a control group that receives no treatment and a treatment group that receives the treatment.
- The **causal effect** is defined to be the effect on an outcome of a given action or treatment, as measured in an ideal randomized controlled experiment.

Cause and Effect

- When graphing data from the real world, it is often more difficult to establish how one variable affects another.
- In other words, seeing correlation (相關性) between two variables does not necessarily imply the existence of causality (因果關係).
- Two problems requires us to proceed with caution when using graphs to draw conclusions about causes and effects.
    - Omitted variables (遺漏變數).
    - Reverse causality.

Omitted Variables



- Figure A-6 shows a strong relationship between two variables: the number of cigarette lighters that a household owns and the probability that someone in the household will develop cancer.

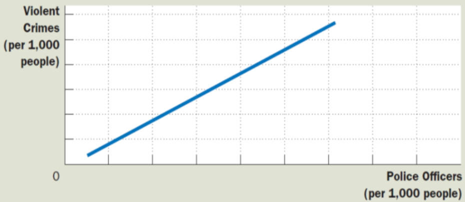
- What should the government do to reduce the number of deaths from cancer?
  1. Discourage the ownership of cigarette lighters by taxing their sale?
  2. Require warning label on cigarette lighters: "This lighter is dangerous to your health" ?
- Are these two suggestions valid?
- Has Figure A-6 held constant every relevant variable except the one under consideration? No.
- An easy explanation for Figure A-6 is that people who own more cigarette lighters are more likely to smoke cigarettes and that cigarettes, not lighters, cause cancer.

- If Figure A-6 does not hold constant the amount of smoking, it does not tell us the true effect of owning a cigarette lighter.
- This story illustrate an important principle: When you see a graph used to support an argument about cause and effect, it is important to ask whether the movements of an omitted variable could explain the results you see.

Reverse Causality

A-7 FIGURE

Graph Suggesting Reverse Causality
The upward-sloping curve shows that cities with a higher concentration of police are more dangerous. Yet the graph does not tell us whether police cause crime or crime-plagued cities hire more police.



- Figure A-7 plots the number of violent crimes per thousand people in major cities against the number of police officers per thousand people.

- The anarchists note the curve's upward slope and argue that because police increase rahter than decrease the amount of urban violence, law enforcement should be abolished.
- Is this a correct argument?
- Figure A-7 only tells that more dangerous cities have more police officers.
- The explanation for this may be that more dangerous cities hire more police.
- In otehr words, rather than police causing crime, crime may cause police.

- If we could run a controlled experiment, we could avoid the danger of reverse causality.
- To run an experiment, we would set the number of police officers in different cities randomly and then examine the correlation between police and crime.
- Unfortunately, controlled experiments are rare in social science.

- It might seem that an easy way to determine the direction of causality is to examine which variable moves first.
- If we see crime increase and then the police force expand, we reach one conclusion.
- If we see the police force expand and then crime increase, we reach the other.
- Yet there is also a flaw with this approach: Often, people change their behavior in response to a change in their *expectations* of future conditions.

- A city that expects a major crime wave in the future, for instance, might hire more police now.
- This problem is easier to see in the case of babies and minivan.
- Couples often buy a minivan in anticipation of the birth of a child.
- The minivan comes before the baby, but we wouldn't want to conclude that the sale of minivans cause the population to grow!

Randomized Trials

- *Potential outcome* Y_{1i} is the outcome for an individual under a potential *treatment*, while *potential outcome* Y_{0i} is the outcome for an individual under a potential *non-treatment*.
- For those *treated*, we observe Y_{1i}. For those not treated, we observe Y_{0i}.
- Let $D_i = 1$ denotes being treated (treatment group), while $D_i = 0$ denotes not being treated (control group).

- The (causal) trearment effect we want to measure is $Y_{1i} - Y_{0i}$ for individual i. Or the average treatment effect

$$E(Y_{1i} - Y_{0i}) = E(Y_i|D_i = 1) - E(Y_i|D_i = 0).$$

Unfortunately, we can not observe Y_{1i} and Y_{0i} at the same time.

- We only observe one group that are treated, and another group that are not treated, and compare their mean differences.

$$\begin{aligned} & Avg_n(Y_i|D_i = 1) - Avg_n(Y_i|D_i = 0) \\ = \quad & Avg_n(Y_{1i}|D_i = 1) - Avg_n(Y_{0i}|D_i = 0) \end{aligned}$$

where $Avg_n(Y_{0i}|D_i = 0) = \frac{1}{n_0} \sum_{i=1}^{n_0} Y_i$.

- For simplicity, suppse that $Y_{1i} = Y_{0i} + \kappa$, or $Y_{1i} - Y_{0i} = \kappa$ for consant treatment effect. Then

$$
\begin{aligned}
& Avg_n(Y_{1i}|D_i = 1) - Avg_n(Y_{0i}|D_i = 0) \\
= \quad & \{\kappa + Avg_n(Y_{0i}|D_i = 1)\} - Avg_n(Y_{0i}|D_i = 0) \\
= \quad & \kappa + \{Avg_n(Y_{0i}|D_i = 1) - Avg_n(Y_{0i}|D_i = 0)\} \\
= \quad & \kappa + \underline{\text{selection bias}}
\end{aligned}
$$

- When n is large enough, $Avg_n(Y_{0i}|D_i = 1) \xrightarrow{p} \mathrm{E}(Y_{0i}|D_i = 1)$ and $Avg_n(Y_{0i}|D_i = 0) \xrightarrow{p} \mathrm{E}(Y_{0i}|D_i = 0)$ (Law of Large Numbers, LLN).

- When D_i is randomly assigned, e.g. by experiments,

$$\begin{aligned} & \text{selection bias} \\ = \quad & Avg_n(Y_{0i}|D_i = 1) - Avg_n(Y_{0i}|D_i = 0) \\ \xrightarrow{p} \quad & \mathrm{E}(Y_{0i}|D_i = 1) - \mathrm{E}(Y_{0i}|D_i = 0) \\ = \quad & 0 \end{aligned}$$

In other words, random assignment can eliminate selection bias.

Data: Sources and Types

Experimental versus Observational Data

- **Experimental data** come from experiments designed to evaluate a treatment or policy to investigate a causal effect.
- Because of financial, practical, and ethical problems, experiments in economics are rare. Most economic data are obtained by observing real-world behavior.
- Data obtained by observing actual behavior outside an experimental setting are called **observational data**.

- In the real world, levels of "treatment" are not assigned at random, so it is difficult to sort out the effect of the "treatment" from other relevant factors.
- Much of econometrics is devoted to methods for meeting the challenges encountered when real-world data are used to estimate causal effects.

Cross-Sectional Data:

California School District Data

TABLE 1.1 Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student-Teacher Ratio	Expenditure per Pupil ($)	Percentage of Students Learning English
1	690.8	17.89	$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
⋮	⋮	⋮	⋮	⋮
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

Note: The California test score data set is described in Appendix 4.1.

Time Series Data:

Growth Rate of GDP and Term Spread in the United States, Quarterly Data, 1960:Q1-2017:Q4.

TABLE 1.2 Selected Observations on the Growth Rate of GDP and the Term Spread in the United States: Quarterly Data, 1960:Q1–2017:Q4

Observation Number	Date (year: quarter)	GDP Growth Rate (% at an annual rate)	Term Spread (percentage points)
1	1960:Q1	8.8%	0.6
2	1960:Q2	−1.5	1.3
3	1960:Q3	1.0	1.5
4	1960:Q4	−4.9	1.6
5	1961:Q1	2.7	1.4
⋮	⋮	⋮	⋮
230	2017:Q2	3.0	1.4
231	2017:Q3	3.1	1.2
232	2017:Q4	2.5	1.2

Note: The United States GDP and term spread data set is described in Appendix 15.1.

Panel Data:

Cigaratte Sales, Prices, and Taxes by State and Year for U.S. States, 1985-95.

TABLE 1.3 Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	$1.022	$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
⋮	⋮	⋮	⋮	⋮	⋮
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
⋮	⋮	⋮	⋮	⋮	⋮
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
⋮	⋮	⋮	⋮	⋮	⋮
528	Wyoming	1995	112.2	1.585	0.360

Note: The cigarette consumption data set is described in Appendix 12.1.

Big Data: (大數據)

- How big is Big?
- What is it?
- ex: 教育部使用大數據分析畢業生主修科系與勞保投保薪資的關係?