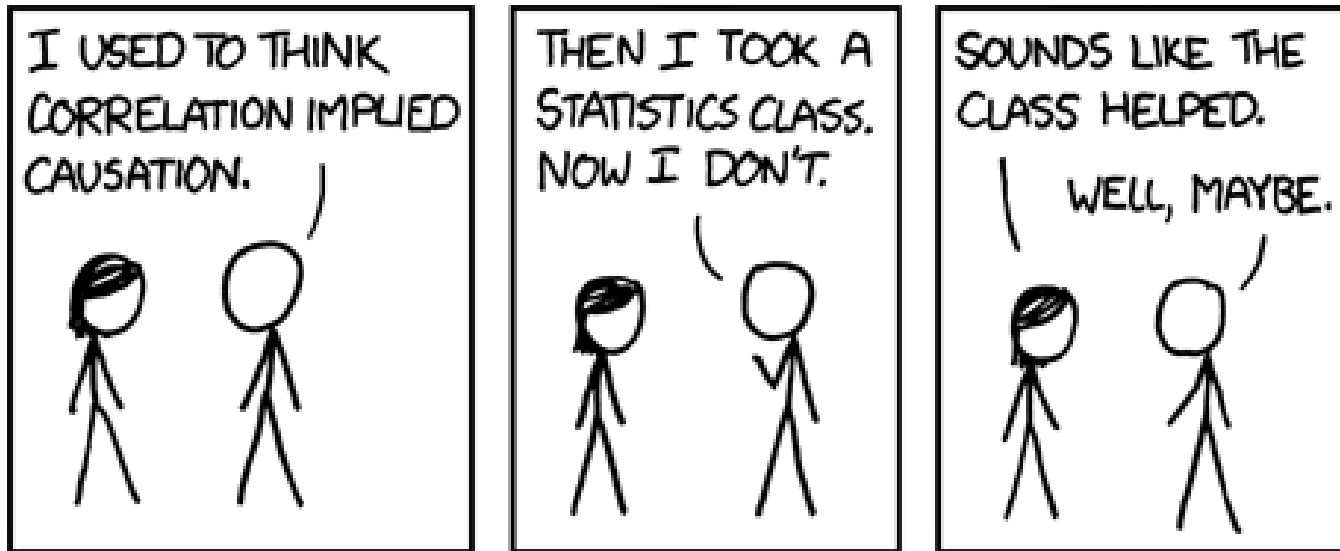


Empirical Micro: Program Evaluation

Ian Crawford

Department of Economics & New College



Sources:

- J.D. Angrist and J-S Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009
- C. Manski, "Nonparametric Bounds on Treatment Effects", *American Economic Review*, 80 (2), 1990, pp.319-323.
- G. Imbens and J. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation", *Journal of Economic Literature*, 47(1) 2009
- D. A. Cobb-Clark and T. Crossley, "Econometrics for Evaluations: An Introduction to Recent Developments," *The Economic Record*, 79(247), 2003
- R. Blundell and M. Costa Dias, "Evaluation Methods for Non-Experimental Data," *Fiscal Studies* 21, 2000
- G. W. Imbens and J. D. Angrist, "Identification and Estimation of Local Average Treatment Effects" *Econometrica* 62(2), 1994

Outline

1. The Evaluation problem/framework
2. Selection
3. Experimental Data
4. Observational Data
 - (a) Bounds
 - (b) The CIA
 - (c) IV (Eye-Vee)

Do hospitals make people healthier?

Although the answer is likely to be "yes" it's arguable either way for some patient groups (e.g. the poor, elderly population that uses A&E for primary care).

The natural approach to an empirical-minded person is probably to get some data on hospital visits and compare the health status of those who have been to the hospital and those who have not.

The National Health Interview Survey (NHIS) contains both types of information.

It contains a question “*During the past 12 months, was the respondent a patient in a hospital overnight?*” which we can use to identify recent hospital visitors.

It also asks “*Would you say your health in general is excellent, very good, good, fair, poor?*” (assigning a 5 to excellent health and a 1 to poor health)*

*It’s actually the other way around in NHIS but I’ve flipped it so that high numbers are better.

Here are the results:

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.21	0.014
No Hospital	90049	2.93	0.003

The difference in mean health status is -0.72 and a statistical test shows that this result is extremely unlikely to have been generated by chance through random sampling.

This result suggests that *going to the hospital makes people sicker*, and the effect is strongly statistically significant

Its not impossible this is the right answer: hospitals are full of sick people who might infect us, and sharp instruments that might hurt us.

The Potential Outcomes Framework

To describe this problem more precisely suppose that we have data on N individuals from an *iid* random sample.

The data records, for each individual $i = 1, \dots, N$

$$\{Y_i, D_i\}$$

The outcome variable is their health status $Y_i \in \{1, \dots, 5\}$

The treatment variable $D_i \in \{0, 1\}$ takes the value 1 if they have been hospitalised and 0 otherwise.

For each individual i we postulate the existence of two *potential outcomes*[†]:

$$Y_i(0) \text{ and } Y_i(1).$$

$Y_i(0)$ denotes the outcome that would have been realised if the person had not received the treatment.

$Y_i(1)$ denotes the outcome that would have been realised if the person had received the treatment.

That is we must be able to imagine what might have happened to someone who went to the hospital if they had not gone and *vice versa*.

[†]Rubin (1974, 1977), and Holland (1986). this (btw) is sheer genius - never underestimate the importance of working out a simple notation/way of writing down a problem.

$$\text{Potential Outcome} = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases}$$

We would like to know the difference between $Y_i(0)$ and $Y_i(1)$, which can be said to be the **causal effect** of the treatment for individual i .

This is what we would measure if we could go back in time and change a person's treatment status or had a parallel universe at our disposal.

If we knew this we could average over the sample and the CLT would let us make inferences about the average effect of treatment in the population.

The problem is that for no individual do we observe both.

- if individual i is hospitalised we observe $Y_i(1)$; and $Y_i(0)$ remains an unobserved counterfactual.
- if individual i is not hospitalised we observe $Y_i(0)$; and $Y_i(1)$ remains an unobserved counterfactual.

The *observed* outcome, Y_i , can be linked to *potential* outcomes as follows

$$Y_i = Y_i(1) D_i + Y_i(0) (1 - D_i)$$

The causal effect we are after is $Y_i(1) - Y_i(0)$

What do we want to know?

Since impacts can be different for different individuals and we never see both $Y_i(1)$ and $Y_i(0)$ we have very little chance of being able to figure out individual impacts.

Nor are we necessarily very interested in the effects on specific individuals. We are much more interested in:

1. The average effect of the treatment on the population (ATE).
2. The average effect of the treatment on the treated population (ATT).

The Average Treatment Effect (ATE)

This is simply the average of the average treatment effect in the population of treatment versus the average outcome under non-treatment.

$$ATE = E [Y_i (1)] - E [Y_i (0)]$$

This parameter tells us what the expected effect of the intervention would be on average for the entire population.

The Average Effect of Treatment on the Treated (ATT)

We may not always be interested in the effect of the treatment on those who did not – and might never – receive the treatment.

We are often more interested in the average effect of the treatment on the treated:

$$ATT = E [Y_i (1) | D_i = 1] - E [Y_i (0) | D_i = 1]$$

The most common evaluation context is one of ex-post evaluation, where we wish to know what change in outcomes an intervention delivered for those who were subject to the intervention.

The average effect of the treatment on the treated answers this question and also tells us what the effect of the treatment is likely to be if similar groups of individuals were to receive the same treatment.

Calculation of Treatment Effects

We know what we want *ATE* or *ATT*.

We have some data $\{Y_i, D_i\}_{i=1, \dots, N}$.

Let's get busy?

Not So Fast

We need to be clear about what the data can tell you and what it cannot.

To see this we need the *law of iterated expectation*[‡]. This says that the population mean can be expressed as the weighted means of sub-populations.

$$E [Y_i (0)] = E [Y_i (0) | D_i = 1] P (D_i = 1) + E [Y_i (0) | D_i = 0] P (D_i = 0)$$

$$E [Y_i (1)] = E [Y_i (1) | D_i = 1] P (D_i = 1) + E [Y_i (1) | D_i = 0] P (D_i = 0)$$

[‡]Proved by Bent Nielsen in his lectures in Week 1. I suggest you also make up some data and check it for yourself.

The *ATE* is

$$ATE = E [Y_i (1)] - E [Y_i (0)]$$

and if we expand out the two expected values into their treated/untreated components we get:

$$ATE = \underbrace{E [Y_i (1) | D_i = 1] P (D_i = 1) + E [Y_i (1) | D_i = 0] P (D_i = 0)}_{E[Y_i(1)]} - \underbrace{E [Y_i (0) | D_i = 1] P (D_i = 1) + E [Y_i (0) | D_i = 0] P (D_i = 0)}_{E[Y_i(0)]}$$

What can the data tell us?

$$\begin{aligned} ATE &= E[Y_i(1) | D_i = 1] P(D_i = 1) + E[Y_i(1) | D_i = 0] P(D_i = 0) \\ &\quad - E[Y_i(0) | D_i = 1] P(D_i = 1) + E[Y_i(0) | D_i = 0] P(D_i = 0) \end{aligned}$$

1. The treatment-assignment probabilities: $P(D_i = 1)$ and $P(D_i = 0)$.
2. The average outcome for the treated: $E[Y_i(1) | D_i = 1]$.
3. The average outcome for the untreated: $E[Y_i(0) | D_i = 0]$.

What can't the data tell us?

$$\begin{aligned} ATE &= E[Y_i(1) | D_i = 1] P(D_i = 1) + E[Y_i(1) | D_i = 0] P(D_i = 0) \\ &\quad - E[Y_i(0) | D_i = 1] P(D_i = 1) + E[Y_i(0) | D_i = 0] P(D_i = 0) \end{aligned}$$

1. The average treatment outcome for the untreated: $E[Y_i(1) | D_i = 0]$
2. The average no-treatment outcome for the treated: $E[Y_i(0) | D_i = 1]$

This means that **the ATE cannot be calculated directly from the data.**

So what *do* we get if we do what we did with the hospital data and just compare the outcomes of the treated/untreated?

That is

$$E [Y_i (1) | D_i = 1] - E [Y_i (0) | D_i = 0]$$

[This is readily observable and I simply can't tell you how tempting this sort of comparison is to politicians, civil servants, journalists, and gurus of all kinds. Often the temptation is irresistible. Sadly many of these people read PPE. That is why you are now doing this course. Blame them. Please send me any examples you come across - the more ridiculous the better.]

Selection bias

The comparison of average reported health conditional on hospitalisation status is formally linked to the average causal effect of hospitalisation by the identity below:

$$E [Y_i (1) | D_i = 1] - E [Y_i (0) | D_i = 0] =$$
$$\underbrace{E [Y_i (1) | D_i = 1] - E [Y_i (0) | D_i = 1]}_{\text{Ave. effect of treatment on the treated}} + \underbrace{E [Y_i (0) | D_i = 1] - E [Y_i (0) | D_i = 0]}_{\text{Selection effect}}$$

Obs. difference in average health = ATT + Selection effect

So the observed difference in health status adds to the causal effect a term we call the *selection effect (bias)*.

$$E [Y_i (0) | D_i = 1] - E [Y_i (0) | D_i = 0]$$

The selection bias is the difference in counterfactual health status in the no-hospital scenario between those who were and were not hospitalised.

If $E [Y_i (0) | D_i = 1] = E [Y_i (0) | D_i = 0]$ (i.e. the no-hospital outcome is the same for those who went to hospital and those who didn't) then no problem.

But the data cannot tell you this as $E [Y_i (0) | D_i = 1]$ is not observable.

But if that's what you think then great, the simple comparison of treated and untreated outcomes is fine (stupid and wrong, but fine, if that's what you believe).

But

Perhaps the sick are more likely than the healthy to seek or be referred for treatment?

This is a **behavioural assumption**.

That means that the expected non-hospital outcome is worse for the group who went to hospital than it is for the group who didn't. So

$$E [Y_i (0) | D_i = 1] < E [Y_i (0) | D_i = 0]$$

Or

$$E [Y_i (0) | D_i = 1] - E [Y_i (0) | D_i = 0] < 0$$

$$\text{Selection bias} < 0$$

If we consider the effect on our comparison of outcomes for the treated versus untreated we get

$$\text{Obs. difference in average health} = ATT + \text{selection effect}$$

$$\text{Obs. difference in average health} = ATT + \text{a negative number}$$

Therefore

$$\text{Observed difference in ave. health} < ATT$$

By comparing them to group who were arguable more healthy (they didn't go to hospital after all) selection bias makes those who were hospitalised seem worse off, even though they are probably better than they otherwise would have been - hence the -0.72 difference in the hospital data.

Selection effects are endemic in observational microeconomic data. They make causal effects hard to measure.

Experiments Solve the Selection Problem

If D_i is randomly assigned it is *independent* of potential outcomes.

$$D_i \perp \{Y_i(0), Y_i(1)\}$$

This means that the probability of assignment to treatment does not vary with the potential outcomes - those more likely to benefit are not more likely to go to hospital under random assignment.

It *does not mean* that the actual outcome is independent of treatment. It means that that the **potential** outcomes are independent of the treatment.

They simply sit there, in parallel universes, and wait for the treatment assignment mechanism to reveal one or the other.

Experiments Solve the Selection Problem

Random assignment helps because it implies that

$$E[Y_i(0) | D_i = 0] = E[Y_i(0) | D_i = 1]$$

and

$$E[Y_i(1) | D_i = 0] = E[Y_i(1) | D_i = 1]$$

Go back to our decomposition of the observed difference in the average outcomes in to the ATT and selection bias

$$\begin{aligned} E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 0] = \\ \underbrace{E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]}_{\text{Ave. effect of treatment on the treated}} \\ + \underbrace{E[Y_i(0) | D_i = 1] - E[Y_i(0) | D_i = 0]}_{\text{Selection effect}} \end{aligned}$$

The fact that under random assignment $E[Y_i(0) | D_i = 1] = E[Y_i(0) | D_i = 0]$ means that the selection bias is zero

Intuition

If people aren't selected into treatment on the basis of their outcomes that means that we would expect the no-treatment outcome for the treated to be the same on average as the no-treatment outcome for the untreated:

$$E [Y_i (0) | D_i = 1] = E [Y_i (0) | D_i = 0]$$

A more formal argument

The concept of independence is symmetric[§]: if A independent of B then B is independent of A .

So if assignment is independent of potential outcomes, then potential outcomes are independent of assignment.

$$P[D|Y(0)] = P(D) \Rightarrow P[Y(0)|D] = P(Y(0))$$

So

$$P(Y(0)|D=1) = P(Y(0)|D=0) = P(Y(0))$$

[§]We did this in the Probability lectures - go and remind yourself if you are unsure.

Conditional Expectations are formed like this[¶] for continuous outcomes (A)

$$E(A|B = b) = \int A P(A|B = b) dA$$

In this case

$$E(Y_i(0) | D_i = 1) = \int Y(0) P(Y(0) | D = 1) dY(0)$$

$$E(Y_i(0) | D_i = 0) = \int Y(0) P(Y(0) | D = 0) dY(0)$$

[¶]We did this too.

We've just shown that

$$P(Y(0) | D = 1) = P(Y(0) | D = 0) = P(Y_i(0))$$

So

$$\begin{aligned} E(Y_i(0) | D_i = 1) &= \int Y(0) P(Y(0)) dY(0) \\ E(Y_i(0) | D_i = 0) &= \int Y(0) P(Y(0)) dY(0) \end{aligned}$$

The LHS's are identical so

$$E[Y_i(0) | D_i = 1] = E[Y_i(0) | D_i = 0]$$

which are the terms in the selection bias

A randomised controlled experiment means that the average effect of treatment on the treated (*ATT*) is the same as the observed difference in the average outcomes for the treatment and control group.

$$E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 0] = \underbrace{E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]}_{\text{Ave. effect of treatment on the treated}}$$

Also it means that the average treatment effect (*ATE*) is equal to the average effect of treatment on the treated (*ATT*)

$$ATE = ATT$$

which means that you can make statements about the effects of an intervention over the whole population.

[I've left you to show this in the suggested tutorial material]

Experiments Solve the Selection Problem...

The main thing, however, is that random assignment of treatment eliminates selection bias for the group of experimental subjects.

This does not mean that randomised trials are problem-free, but in principle they solve the most important problem that arises in empirical micro.

Randomised controlled trials are therefore usually considered the best possible approach to the study of causal effects.

Experiments Solve the Selection Problem...

Experimental evaluations have traditionally been rare in economics but they are becoming very popular.

In many cases ethical considerations, as well as the reluctance of administrators to deny services to randomly selected individuals after they have been deemed eligible, have made it difficult to get approval for, and implement, randomised evaluations.

Nevertheless, the few experiments that have been conducted, including some of the labour market training programs, have generally been influential, sometimes extremely so.

Experiments Solve the Selection Problem...

One area where the importance of random assignment is growing rapidly is education research.

The *2002 Education Sciences Reform Act* in the US mandates the use of rigorous experimental or quasiexperimental research designs for all federally-funded education studies.

E.g. the Tennessee STAR experiment designed to estimate the effects of placing primary school children in smaller classes.

E.g. the evaluation of the EMA in England and Wales.

...But they don't solve everything

Even randomised experiments rely to some extent on substantive knowledge. In particular the experimental setup only works if there are no interaction effects between subjects.

Only once the researcher is willing conceptually to rule out interactions between units is it the case that randomisation can establish causal effects.

In settings with potentially unrestricted interactions between units, randomisation by itself cannot solve the identification problems required for establishing causality.

One example of such spillovers was an experiment which offered a capital grant to small businesses in Sri Lanka. In one sector (bamboo furniture) the recipients

of the capital grant used it to purchase all the raw material bamboo on the market, *de facto* driving their competitors out of business.

In the economics literature randomisation has played a much less prominent role than in bio-statistics.

Part of this may be due to the fact that for the treatments of interest to economists, e.g., education and labor market programs, it is generally impossible to do blind or double-blind experiments, creating the possibility of *placebo* effects that compromise the internal validity of the estimates.

The evidence in medical trials is that placebo effects are very influential.

Contamination

People in the control group access the treatment anyway. In other words, some of the untreated turn out to be treated too.

This generates a bias against finding a significant difference between the treated and untreated (the treatment effect is underestimated).

Non-compliance

Individuals who are offered a treatment refuse to take it. In other words, some of the treated turn out to be untreated.

In economic experiments, someone may be offered vocational training but fail to take advantage of it.

Working with Observational Data

Observational data are data which are not generated by a randomised controlled trial.

We often work with observational data in a search for causal effects (experiments being expensive and tricky to set up and also at risk from contamination/compliance problems which re-introduce selection bias).

How can we infer causal effects from these kinds of data where selection bias is a problem once more?

(1) Bounds (2) The CIA (3) Instrumental Variables

1. Bounds

"Ranges are for cattle - give me a number!" Lyndon B. Johnson

Selection bias can be positive or negative.

If you know the likely direction of the bias than you have a one-sided bound on the *ATT*.

Obs difference in averages = *ATT* + Selection

If (selection bias > 0) then (Obs difference in averages $> ATT$)

If (selection bias < 0) then (Obs difference in averages $< ATT$)

Supposing that the outcomes $Y_i(0)$ and $Y_i(1)$ are bounded, we can put bounds on the *ATE*.

For example,

- it might be a test score $Y_i \in [0\%, 100\%]$
- it might be binary like "finds a jobs/doesn't find a job" $Y_i \in \{0, 1\}$

Either way

$$Y_i(0) \in [a_0, b_0] \text{ and } Y_i(1) \in [a_1, b_1]$$

These values are known as the *support* of the RV.

All conditional means lie in the support of the RV^{||}.

That means that we can put bounds on the tricky counterfactual terms:

$$\begin{aligned} E [Y_i (0) | D = 1] &\in [a_0, b_0] \\ E [Y_i (1) | D = 0] &\in [a_1, b_1] \end{aligned}$$

We can use these to bound $E [Y_i (0)]$ and $E [Y_i (1)]$ then on the *ATE*.

^{||}See for yourself - generate some numbers and see if you can get the mean to lie outside the range.

For example

$$E [Y_i (0)] = E [Y_i (0) | D_i = 0] P (D_i = 0) + E [Y_i (0) | D_i = 1] P (D_i = 1)$$

Given

$$Y_i (0) \in [a_0, b_0]$$

That means that

$$E [Y_i (0)] \in [\underline{Z}_0, \bar{Z}_0]$$

where

$$\underline{Z}_0 = E [Y_i (0) | D_i = 0] P (D_i = 0) + a_0 P (D_i = 1)$$

$$\bar{Z}_0 = E [Y_i (0) | D_i = 0] P (D_i = 0) + b_0 P (D_i = 1)$$

Likewise we can put bounds on $E[Y_i(1)]$.

$$E[Y_i(0)] \in [\underline{Z}_0, \bar{Z}_0]$$
$$E[Y_i(1)] \in [\underline{Z}_1, \bar{Z}_1]$$

As a result, the bounds on the ATE are

$$ATE = E[Y_i(1)] - E[Y_i(0)] \in [\underline{Z}, \bar{Z}]$$

where

$$\underline{Z} = \underline{Z}_1 - \bar{Z}_0 \quad (\text{bottom to top})$$
$$\bar{Z} = \bar{Z}_1 - \underline{Z}_0 \quad (\text{top to bottom})$$

The width of the bound is

$$(b_1 - a_1) P(D = 0) + (b_0 - a_0) P(D = 1)$$

So how precise you are able to be depends on the treatment-assignment probabilities (which you can observe) and the width of the support on the outcome variables.

This bound might be wide and it might cover zero - that's life.

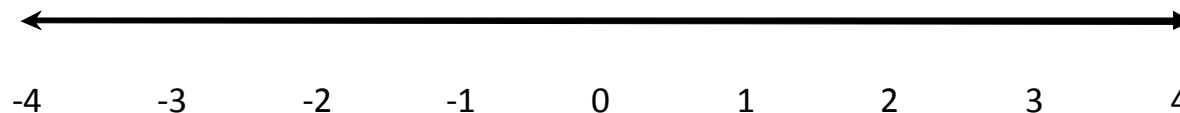
Furthermore **more data (a bigger sample) won't help** - that will only help you estimate the bound with better statistical precision.

An Example: The hospitalisation data

From our data we can estimate the treatment assignment probabilities and the average outcomes for the treated and untreated

$$\begin{aligned} P(D_i = 1) &= 0.0795 \text{ and } P(D_i = 0) = 0.9205 \\ E[Y_i(1) | D_i = 1] &= 2.21 \text{ and } E[Y_i(0) | D_i = 0] = 2.93 \end{aligned}$$

If we know nothing other than the fact that health is on a five point scale then all we can say is that $ATE \in [-4, +4]$



But since health $Y_i(0) \in [1, 5]$ we know that the bounds on $E[Y_i(0)]$ are

$$\begin{aligned} \underline{Z}_0 &= E[Y_i(0) | D_i = 0] P(D_i = 0) + 1 \times P(D_i = 1) \\ &= 2.93 \times 0.9205 + 1 \times 0.0795 = 2.777 \end{aligned}$$

$$\begin{aligned} \bar{Z}_0 &= E[Y_i(0) | D_i = 0] P(D_i = 0) + 5 \times P(D_i = 1) \\ &= 2.93 \times 0.9205 + 5 \times 0.0795 = 3.095 \end{aligned}$$

So we know that

$$E[Y_i(0)] \in [2.777, 3.095]$$

Likewise, the bounds on $E[Y_i(1)]$ are

$$\begin{aligned}\underline{Z}_1 &= E[Y_i(1) | D_i = 0] P(D_i = 0) + E[Y_i(1) | D_i = 1] P(D_i = 1) \\ &= 1 \times 0.9205 + 2.21 \times 0.0795 = 1.096\end{aligned}$$

$$\begin{aligned}\bar{Z}_1 &= E[Y_i(1) | D_i = 0] P(D_i = 0) + E[Y_i(1) | D_i = 1] P(D_i = 1) \\ &= 5 \times 0.9205 + 2.21 \times 0.0795 = 4.778\end{aligned}$$

So we know that

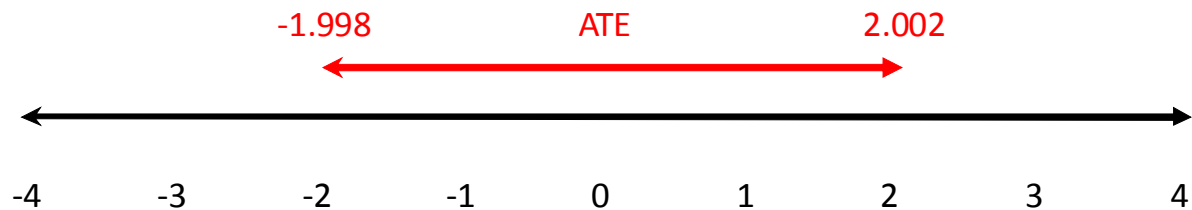
$$E[Y_i(1)] \in [1.096, 4.778]$$

Given

$$E[Y_i(0)] \in [2.777, 3.095]$$

$$E[Y_i(1)] \in [1.096, 4.778]$$

the ATE of hospitalisation is in the range $ATE \in [-1.998, 2.002]$



We have ruled out precisely half of the outcome space - so the cup is half full/half empty according to taste.

We can improve if we think about our selection bias argument. **NB this requires a behavioural assumption.**

There we said that we would expect sicker-than-average people to seek hospital treatment.

So we would expect the no-hospital outcomes for people who go to hospital to be worse on average than the no-hospital outcomes for those who don't go to hospital. That is

$$E [Y_i (0) | D_i = 1] \leq E [Y_i (0) | D_i = 0]$$

This means that

$$E [Y_i (0)] \in [1, E [Y_i (0) | D_i = 0]]$$

$$E [Y_i (0)] \in [1, 2.93]$$

Is there anything we can say about the other counterfactual $E [Y_i (1) | D_i = 0]$, i.e. the health outcomes of people who didn't go to hospital. What if they *had*?

What would we expect their outcomes to be, compared to the people who did present themselves at hospital? Presumably they would do even better because they've been to hospital and treated for potential/latent problems and they are not even sick at the moment. *If* that's the case then

$$E [Y_i (1) | D_i = 0] > E [Y_i (1) | D_i = 1]$$

In our data this means that

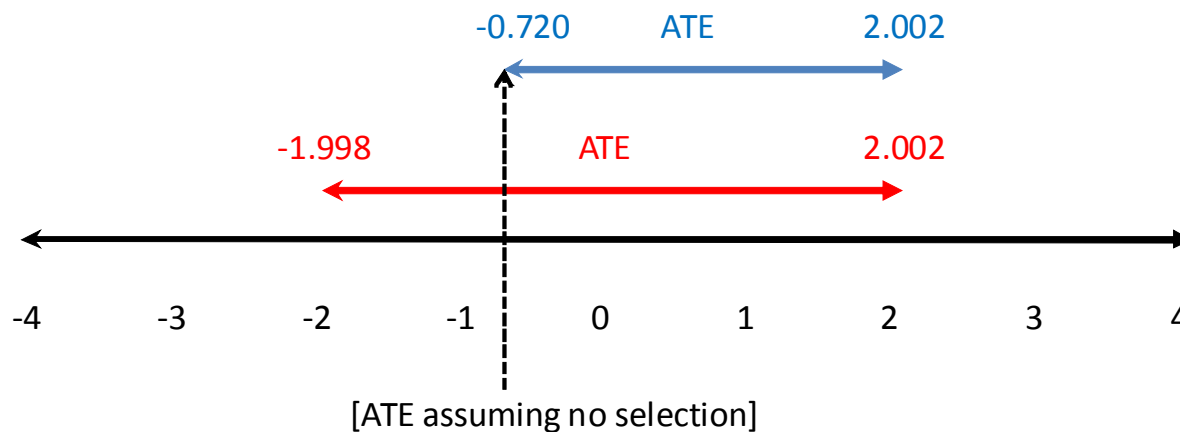
$$E [Y_i (1)] \in [2.21, 5]$$

Using these arguments which derive from behavioural assumptions/models

$$E[Y_i(0)] \in [1, 2.93]$$

$$E[Y_i(1)] \in [2.21, 5]$$

then using the same methods as before we get $ATE \in [-0.720, 2.002]$



Now I'm the first to admit that this isn't brilliantly useful (LBJ would be *very* unhappy). We've managed to tie down the *ATE* to lie somewhere between the statements:

- “going to hospital reduces average health scores in the population by 0.72 points”.
- “going to hospital raises health scores by 2.002 points”.

But it does tie us down to a range where

- we have ruled out nearly two thirds of the potential outcome space.
- the simple difference -0.72 is shown to be right at the low extreme of the possible range.
- we know that the effect can't be more than about 2 points.

The Role of Economic Theory

We improved our knowledge of the ATE by using a behavioural model which said that sicker-than-average people tend to go to hospital and healthier-than-average people tend not to go. *There was nothing in the data which could tell us this.*

Microeconomic theory is full of behavioural models (most of them based on marginal cost/marginal benefit arguments) which supply us with insights into selection bias:

- Workers only accept wages above their reservation wage.
- Consumers only buy marginal units if their willingness to pay exceeds the marginal cost (price).
- Players only select strategies which have better pay-off than the alternatives.

The Role of Economic Theory

At heart, quantitative economics concerns the interplay between data and behavioural models.

1. Stronger behavioural assumptions allow tighter identification.
2. Better data (as opposed to simply *more* data) allows identification under weaker assumptions.

Knowing how robust identification is to your explicit and implicit assumptions is essential (and surprisingly challenging).

A sermon – on Incredible Econometrics

[From the Gospel according to Chuck]

Powerful incentives influence those doing policy evaluation to maintain assumptions far stronger than they can persuasively defend, in order to draw strong conclusions.

(a) The academic community rewards those who produce unambiguous findings.

(b) The public/journalists/politicians reward those who offer simple analyses leading to unequivocal policy recommendations.

But research findings based on incredible assumptions are not much use.

The objective of a policymaker should be to choose the best policy, not the policy which would be best if incredible assumptions were to hold.

Independence again

The key to point identification of treatment effects is to make sure that *assignment to treatment is independent of potential outcomes*.

We saw that explicit randomisation of assignment to treatment is **one** way of guaranteeing this.

In observational data subjects have not been randomised into treatment.

We are going to look at two new strategies which work with observational data.

Both of them work by finding ways of reintroducing randomisation into the data - *conditional independence* and *instrumental variables*.

2. Conditional Independence/Unconfoundedness

So far the only data we have observed has been outcomes and treatment.

Suppose that we can now observe a list of other things about the individual.

$\mathbf{X}_i = \{X_i^1, \dots, X_i^K\}$ (“covariates” in the lingo).

These could be made up of personal characteristics, information on educational attainment, past work history etc.

Formally \mathbf{X}_i denotes a vector but generally you’ll be fine thinking of it as a scalar. Now the data are

$$\{Y_i, D_i, \mathbf{X}_i\}$$

The original independence assumption was that assignment to treatment was independent of potential outcomes

$$D_i \perp \{Y_i(0), Y_i(1)\}$$

Randomisation delivered this.

The conditional independence assumption (CIA) says assignment to treatment is independent of potential outcomes *conditional on covariates*.

$$D_i \perp \{Y_i(0), Y_i(1)\} | \mathbf{X}_i$$

In other words outcomes don't vary with treatment *holding other factors about the individual fixed*.

Identifying treatment effects under Conditional Independence/Unconfoundedness

Given CIA you just concentrate your attention on particular values of the covariates \mathbf{X} and hold those fixed.

$$\underbrace{E[Y_i(1) | D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_i(0) | D_i = 0, \mathbf{X}_i = \mathbf{x}]}_{\text{Obs. difference in averages at } \mathbf{X}_i = \mathbf{x}} =$$

$$\underbrace{E[Y_i(1) | D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_i(0) | D_i = 1, \mathbf{X}_i = \mathbf{x}]}_{\text{Ave. effect of treatment on the treated at } \mathbf{X}_i = \mathbf{x}}$$

$$+ \underbrace{E[Y_i(0) | D_i = 1, \mathbf{X}_i = \mathbf{x}] - E[Y_i(0) | D_i = 0, \mathbf{X}_i = \mathbf{x}]}_{\text{Selection effect at } \mathbf{X}_i = \mathbf{x}}$$

Everything is conditional on the covariates taking particular fixed values.

Using *exactly the same logic as before*, what conditional independence buys us is

$$E [Y_i (0) | D_i = 1, \mathbf{X}_i = \mathbf{x}] = E [Y_i (0) | D_i = 0, \mathbf{X}_i = \mathbf{x}]$$

but this time we are holding other covariates fixed. In fact all we need to get *average treatment effects is conditional mean independence (CIA \Rightarrow CMI)*.

This makes the the conditional selection bias term disappear and

$$\text{Obs. difference in averages} | \mathbf{X}_i = \mathbf{x} = ATT(\mathbf{x})$$

This identifies the ATT (and the ATE) at each fixed value of the covariates.

Given CIA you can recover both $ATE(\mathbf{x})$ and $ATT(\mathbf{x})$.

If our sample is large enough and \mathbf{x} only takes few values, we can just divide the data into cells and take sample averages of outcomes for treated and untreated observations within each cell to get a consistent estimate of each $ATE(\mathbf{x})$.

Getting ATE and ATT is then just a matter of averaging over cells.

$$E[ATE] = E[ATE(\mathbf{x})]$$

$$E[ATT] = E[ATT(\mathbf{x})]$$

When is the CIA incredible?

The best case scenario for the CIA is when treatment is assigned randomly conditional on X .

This is the case, for example, in the study in Black, *et al* (*American Economic Review*, 2003). They looked at a mandatory retraining program for the unemployed.

They wanted to know whether it raised subsequent earnings.

Eligibility for the program was determined by a number of personal characteristics and past employment/training history.

Workers were divided into groups on the basis of these variables.

Some groups were ineligible.

Others were required to take training.

[So far there is no randomisation and a comparison of these apples/pears would be seriously confounded]

But ... when some of the groups due to receive mandatory training contained more workers than there were training places available, training opportunities were allocated by lottery.

Hence training was assigned randomly *conditional on the covariates used to assign people to groups*.

If randomisation isn't obviously inherent in the data detailed institutional knowledge can help to some extent.

An example of this is Angrist (*Econometrica*, 1998) which looks at whether men who volunteered for the US military were economically better off in the long run.

Voluntary military service is not randomly assigned.

But the US military screens applicants on the basis of observed covariates like age, schooling and test scores.

Conditional on these covariates Angrist looked at a comparison of those qualified applicants who failed to enlist at the last minute with those who actually turned up at Boot Camp. He claimed that this was random with respect to potential outcomes.

This is less clean than the training example where we *know* that, conditional on covariates, assignment to treatment was random.

In order to believe the CIA in this context you need to think that the things which lead a qualified applicant to drop out of the enlistment process right at the last moment were independent of his future earnings.

You have to decide if this is incredible.

E.g. What if the characteristic which lead applicants to drop out at the last minute was "*Flakiness*"**

Flakiness is plausibly correlated with both potential earnings (Angrist's outcome variable) and whether or not applicants could get it together enough to turn up a basic training.

If the screening the military carried out did not effectively screen for flakiness then assignment to treatment was not random and CIA is incredible.

**Thanks to Clare Leaver (Queens) for this suggestion.

CIA Implementation - support

When we apply the CIA we need to calculate things like

$$E [Y_i (1) | D_i = 1, \mathbf{X}_i = \mathbf{x}] \text{ and } E [Y_i (0) | D_i = 0, \mathbf{X}_i = \mathbf{x}]$$

i.e. the average outcomes for the treated and the untreated in a particular slice of the population (defined by the covariates taking certain values).

You need both of these objects.

CIA Implementation - support

I have, throughout this lecture, essentially assumed that we are working with the population rather than a sample and under this assumption then there is no problem - you will always be able to compute both (and if not it's a non-issue).

The problem is that when we are working with sample data if we need to specify the slice of the population very tightly in order to make the CIA plausible we may be left with very few observations in one of the groups - maybe even none.

The Curse of Dimensionality.

Suppose that \mathbf{X} contains 5 covariates and we have data on 100 individuals independently uniformly distributed within a 5-dimensional unit cube $[0, 1]^5$

What is the probability of having some points in a cell of a reasonable size, say a cube with side 0.2? It's $0.2^5 = 0.00032$

The expected number of observations in such a neighbourhood is 0.032.

To get an expected number of 5 would require a cube whose side was 0.55, i.e. more than half the range in each dimension (and this would cause problems with the credibility of any CIA-based argument based on small differences being negligible).

In this situation more data helps.

3. Instrumental Variables (IV)

The CIA approach restores the situation by looking at very closely defined subgroups and arguing that once you have conditioned on these covariates what remains is essentially random with respect to outcomes.

The Instrumental Variables approach works by arguing that there is randomisation at work, but in a mechanism which is further back up a slightly longer causal chain.

The data now consists of

1. the outcome of interest Y_i
2. the treatment indicator $D_i \in \{0, 1\}$
3. an *instrument* $Z_i \in \{0, 1\}$ it doesn't have to be binary but we will work with a binary instrument.

What is an “Instrument”?

It is something which initiates a causal chain:

Z_i affects D_i (the treatment variable of interest).

D_i affects Y_i (the outcome of interest)

The instrument needs to be

1. randomly assigned
2. affect D_i
3. to only affect Y_i through D_i

An example: Angrist (*American Economic Review*, 1990)

Y_i earnings

D_i military service

Z_i "success" in the Vietnam draft-eligibility lottery

The instrument is:

1. randomly assigned
2. affects the treatment
3. does not directly affect the outcome.

To model this we use a potential outcomes notation for the receipt of the treatment (D_i) as well as for the outcome itself (Y_i).

Denote the level of the treatment received if the instrument takes on the values 0 and 1 respectively by $\{D_i(0), D_i(1)\}$

As before denote the potential values for the outcome of interest $\{Y_i(0), Y_i(1)\}$

Observed outcomes and treatments are linked to their potential counterparts by

$$\begin{aligned} Y_i &= Y_i(1) D_i + Y_i(0) (1 - D_i) \\ D_i &= D_i(1) Z_i + D_i(0) (1 - Z_i) \end{aligned}$$

Independence here requires that the instrument is independent of both potential treatment and potential outcomes

$$Z_i \perp \{Y_i(0), Y_i(1), D_i(0), D_i(1)\}$$

As before don't get confused: this does not mean that the instrument does not effect observed outcomes.

It means that the potential outcomes for each person sit "out there" in parallel universes until the assignment mechanism selects one or other of them to be observed.

So the model is composed of two bits:

$$Y_i = Y_i(1) D_i + Y_i(0) (1 - D_i) \quad (1a)$$

$$D_i = D_i(1) Z_i + D_i(0) (1 - Z_i) \quad (1b)$$

$$Z_i \perp \{Y_i(0), Y_i(1), D_i(0), D_i(1)\} \quad (2)$$

1. captures an exclusion restriction that there is no direct effect of the instrument on the outcome (Z_i is absent in the definition Y_i).
2. captures random assignment of Z_i so that causal effects of the instrument on both Y_i and D_i can be estimated consistently.

Assumption 1 is not implied by randomisation so has to be argued on a case-by-case basis.

If the randomly assigned instrument completely determined D_i then we can just use Z_i as the treatment (they are the same) and we are immediately in business as we have a nice randomised controlled trial.

In observational data D_i is **not** randomly assigned.

This means that just because $Z_i = 1$, it is not necessarily the case that $D_i = 1$.

Nor that just because $Z_i = 0$, it is necessarily the case that $D_i = 0$.

Example. In the case of the Vietnam war the draft lottery determined *eligibility* for drafting - but you might still not get drafted for medical reasons or you could go to Canada (the Economics Dep'ts of Canadian Universities are rammed with US.draft-dodgers who are now nearing retirement)

In the case of medical trials assignment to the treatment/control group might be random but you might not take the medicine/access the treatment when you weren't supposed to or you might access it anyway.

This makes D_i a choice variable ("endogenous" in the lingo) and causes selection bias.

To deal with contamination/compliance problems like this it's useful to think of the “*compliance type*” of an individual (Imbens and Angrist, *Econometrica*, 1994).

The type of an individual describes the level of the treatment that an individual would receive given each value of the instrument.

In other words, it is captured by the pair of values

$$\{D_i(0), D_i(1)\}$$

{Treatment status if the instrument is 0, Treatment status if the instrument is 1}

With both the treatment and instrument binary, there are four types of responses for the potential treatment.

$$T_i = \begin{cases} \text{Never Taker if } D_i(0) = D_i(1) = 0 \\ \text{Complier if } D_i(0) = 0, D_i(1) = 1 \\ \text{Defier } D_i(0) = 1, D_i(1) = 0 \\ \text{Always Taker if } D_i(0) = 1, D_i(1) = 1 \end{cases}$$

Never takers will never sign up for Vietnam regardless of their lottery number. Always takers would enlist in any case. Compliers would go if their number came up but would stay safe at home otherwise. Defiers are a bit crazy.

Monotonicity

We need one more ingredient before we can sort out the causal tangle..

We need to make an assumption - that there are *No Defiers* in the data (i.e. people who would only go to Vietnam if their number didn't come up in the draft lottery, but would refuse to go if it did).

This assumption is also called "*monotonicity*" (in the lingo) and a more general statement is that increasing the level of the instrument does not decrease the level of the treatment.

You have to decide if it's incredible.

Incredible IV

IV will work with observational data

1. when you have a valid instrument which means the instrument is:
 - (a) randomly assigned (or as good as)
 - (b) effects the treatment
 - (c) does not directly affect outcome.

2. when the no-defiers/monotonicity requirement holds.

If either one of these requirements do not hold then IV is *incredible*.

Incredible IV - An Example

We are interested in the treatment of admission to *Dimwit College, Oxbridge*:
 $D_i \in \{0, 1\}$.

The outcome is future earnings Y_i .

Applicants are interviewed by two Tutors who make offers of places.

Applicants are randomly assigned to one or the other Tutor for interview $Z_i \in \{0, 1\}$.

Since the identity of the Admissions Tutor is plausibly immaterial to the final outcome, the instrument is valid - it is randomly assigned, effects treatment (offer of a place), but not the outcome (future earnings).

Incredible IV - An Example

But what about the no defiers/monotonicity requirement?

Remember that this is the general requirement that increasing the instrument does not decrease the treatment.

In this example it requires that one tutor will always accept any applicant who the other would have accepted.

This is pretty unlikely.

So the IV strategy is incredible even though the instrument is good.

LATE - Local Average Treatment Effects

Given these data and the no-defiers assumption Imbens and Angrist (1994) show that we can identify the effect of the treatment for the subpopulation of Compliers but only them.

Why not the Never-takers and Always-takers?

Never-takers are never observed receiving the treatment, $E[Y_i(1)|T_i = n]$ is not identified.

Always-takers are never observed not receiving the treatment, $E[Y_i(0)|T_i = a]$ is not identified.

Only Compliers are observed in both treatment groups, so only for this group is there any chance of identifying the average treatment effect.

Intuition for Identification for the LATE

The Instrument is randomly assigned (or as good as).

Only for the Compliers does the random assignment of the instrument translate into the random assignment of the treatment (they do exactly as they were supposed to).

That means that for Compliers the treatment is randomly assigned. Therefore there is no selection bias for this subpopulation.

We therefore need to compute

$$LATE = E [Y_i | D_i = 1, T_i = c] - E [Y_i | D_i = 0, T_i = c]$$

and interpret it as a causal effect (there's no selection bias).

But the problem is you cannot easily identify Compliers

		$Z_i =$	
		0	1
$D_i =$	0	Never Taker/Complier	Never Taker
	1	Always Take	Always Taker/Complier

$E[Y_i | D_i = 1, T_i = c] \neq E[Y_i | D_i = 1, Z_i = 1]$ - this mixes Compliers with Always-Takers

$E[Y_i | D_i = 0, T_i = c] \neq E[Y_i | D_i = 0, Z_i = 0]$ - this mixes Compliers with Never-Takers

So we're going to have to be cunning^{††}

^{††}At least "as cunning as a fox who has just been appointed professor of cunning at Oxford University"

First we need to identify the sample proportions of each type: P_a, P_n, P_c (you'll see why in due course).

The key to doing this is that the instrument Z_i is randomly assigned. Hence the proportion of each type is the same in the $Z_i = 0$ and the $Z_i = 1$ groups.

That means that we can figure out P_a just by looking at the slice of the data with $Z_i = 0$.

Within this subpopulation we observe $D_i = 1$ only for always-takers. Therefore^{‡‡}

$$P_a = P(D_i = 1 | Z_i = 0)$$

^{‡‡}You saw the MLE estimator for population proportions in Bent Nielsen's lectures.

Similarly, in the subpopulation with $Z_i = 1$ we observe $D_i = 0$ only for never-takers.

Hence

$$P_n = P(D_i = 0 | Z_i = 1)$$

The population share of compliers is then obtained by subtracting the population shares of never-takers and always-takers from one:

$$P_c = 1 - P_n - P_a.$$

So now we have $\{P_a, P_n, P_c\}$.

Now take $E(Y_i|D_i = 1, Z_i = 1)$. We know that it is the weighted average over Compliers and Never-Takers.

$$E(Y_i|D_i = 1, Z_i = 1) = E(Y_i|D_i = 1, T_i = a) \frac{P_a}{P_a + P_c} + E(Y_i|D_i = 1, T_i = c) \frac{P_c}{P_a + P_c}$$

We want to know $E(Y_i|D_i = 1, T_i = c)$

We already have the various probabilities and can compute both $E(Y_i|D_i = 1)$ and $E(Y_i|D_i = 1, T_i = a)$ straight from the data (because we know that everyone with $D_i = 1, Z_i = 0$ is an Always-Taker).

So we can back out $E(Y_i|D_i = 1, T_i = c)$ - we know everything else.

Similarly we can back $E(Y_i|D_i = 0, T_i = c)$ out of $E(Y_i|D_i = 0, Z_i = 0)$:

$$E(Y_i|D_i = 0, Z_i = 0) = E(Y_i|D_i = 0, T_i = c) \frac{P_c}{P_n + P_c} + E(Y_i|D_i = 0, T_i = n) \frac{P_n}{P_n + P_c}$$

We want to know $E(Y_i|D_i = 0, T_i = c)$

We can compute $E(Y_i|D_i = 0, T_i = n)$ straight from the data because we know that everyone with $D_i = 0, Z_i = 1$ is a Never-Taker.

So we can back out $E(Y_i|D_i = 0, T_i = c)$ - we know everything else.

Now we've got both parts and can calculate the *LATE*

$$LATE = E(Y_i | D_i = 1, T_i = c) - E(Y_i | D_i = 0, T_i = c)$$

and this has a causal interpretation.

Why? (Reminder) : Because the random assignment of the instrument implies random assignment of the treatment for compliers (and only compliers).

Pretty cunning, eh?

