Instrumental variable analysis

Professor of Behavioral Economics

Economics Division

Leeds University Business School

Web: https://business.leeds.ac.uk/divisions-economics/staff/126/dr-peter-howley

Natural Experiments

- Terms applied to a wide variety of studies that resemble randomised field experiments discussed in last lecture

- **Natural experiments** (or **quasi-experiments**) are similar to random assignment experiments, **except**:
  - observations fall into treatment and control groups **"naturally"** (because of an **exogenous event**) *instead of being randomly assigned* by the researcher
  - By **"exogenous event"** we mean that the natural event must not be under the control of the researcher

- In other words, we exploit differences in outcomes between a treatment and a control group, just as in a classical experimental design but in the case of a natural experiment, however, treatment status is determined by nature, politics, an accident or some other action beyond a researchers control.

Instrumental variable analysis

- When you can neatly identify a treatment and a control group then as we have seen you can often calculate the causal effect (e.g. new policy) directly

- When the treatment/control group distinction works as a device to manipulate an underlying variable researchers sometimes use an extension of the natural experiment method known as instrumental variables (IV)

  - IV studies are simply a specific way to use natural experiments.

- An instrumental variable is a variable that causes the independent variable to change but does not affect the outcome in any way, other than through the independent variable

  - The best way to learn this is by working our way through some examples which we will do

Endogeneity

- IV methods (like all natural experimental methods) are useful when you are concerned about endogenity and other approaches such as regression adjustment or a pure experimental design (RCT in a field experimental setting) is not possible

- When does endogenity arise? – spurious correlation (we have looked at this a lot)

  - 1. **Omitted variables bias** – confounding (observed association can be misleading as it partly reflects omitted variables that are related with to both variables (X and Y).

  - 2. **Bi-directional causality**

- To that I add a third – 3. **Measurement error** (typically a downward bias)

- When economists are talking about endogeneity in an applied setting they are usually talking about one of those three

  - will focus on omitted variables bias to illustrate the benefits of an Instrumental variables analysis but equally relevant for dealing with measurement error or bi-directional causality (sometimes referred to as simultaneity bias)

Omitted variable bias – what can we do?

- **Regression adjustment** – control for the omitted variables
  - Hard to control for everything
  - Knowledge can be limited and so can be hard to even know what could be the source of omitted variable bias in certain circumstances.

- **Experimental design** - Randomization ensures that the "confounders" are "balanced" between the treatment and the control group.
  - Not always possible to pursue an experimental design (indeed most studies in the social sciences deal with observational as opposed to experimental data)

- Instrumental variable analysis can help us deal with this endogeneity

  - an instrumental variables approach will **generate predicted values for our endogenous variable**, which will be **purged of omitted-variable bias.**

What is an instrumental variable

- Suppose X (e.g. education) and Y (earnings) are our key variables

- An instrumental variable is a third variable **Z which is related with X, but not with Y** except through its association with X

- Not as easy as it sounds to identify these variables

  - not always but often derive from natural experiments, e.g. nature, some law or policy change will often provide you with your instrumental variable

  - think of it like this – some random event will often predict levels of (changes in) your key 'endogenous' explanatory variable but wont be related with your outcome variable

How do they work in practice

- IVs provide a remedy to omitted variable bias by using only that portion of the variability in an independent variable (e.g., education) that is uncorrelated with the omitted variables (e.g., ability) to estimate the causal relationship between the independent variable (education) and the dependent variable (earnings)

- **More technically, a key assumption of standard regression models is that a given treatment or independent variable is uncorrelated with the model's error term.**

  - Textbooks at least economics ones typically talk about endogeneity in this fashion – bias due to correlation with the error term

- How might this correlation with the error term occur?

  Omitted variable bias!

  Bi-directional causality

  Measurement error

Things to be aware of

- **Weak instrument problem**: If a treatment variable/endogenous explanatory variable and your instrumental variable are only weakly correlated (or completely unrelated), then the IV estimate may not be true to the causal effect

- **Instrument validity**: This is key
  - Your instrumental variable Z should be **correlated with X but not with Y**, except through its association with X – commonly referred to as the exclusion restriction
  - Sometimes control variables can help if validity is in doubt, i.e. conditioning on the control variables can you make a strong case that your proposed instrumental variable will not affect the outcome/dependent variable, but will impact exposure to the treatment/explanatory variable in question?

- If these two conditions are met, then instrumental variable analysis is really good at purging your estimates from any omitted variable bias
  - same point when it comes to bias due to bi-directional causality and measurement error

Lets work though some examples

- Evans and Ringel (1999) Maternal Smoking, Cigarette Taxes, and Birth Weight

- Study question was whether smoking affected birthweights

- Can just look at the relationship between smoking and birthweight in the US. However

  - Mothers who smoke likely to be very different to mothers who don't and these differences (omitted variables) may be related with birthweight – selection bias

  - In other words, smoking is likely endogenous to birth weight

  -A field experiment (RCT) is not feasible – we cant force people to smoke and others not to!

- An instrumental variable would help deal with this endogeneity bias

  - Researchers used cigarette taxes as an instrumental variable

  -Cigarette taxes will vary over states and over time

Maternal smoking, cigarette taxes and birthweight

- Is cigarette taxes a good candidate for an instrumental variable in this context?

Maternal smoking, cigarette taxes and birthweight

- Is cigarette taxes a good candidate for an instrumental variable in this context?

1. Is cigarette taxes significantly correlated with the endogenous explanatory variable (smoking)

   - of course, moreover the correlation is unlikely to be weak

Maternal smoking, cigarette taxes and birthweight

- Is cigarette taxes a good candidate for an instrumental variable in this context?

1. Is cigarette taxes significantly correlated with the endogenous explanatory variable (smoking)

   - of course, moreover the correlation is unlikely to be weak

2. Is **Z unrelated to Y** except through its effect on X (e.g. no direct relationship between cigarette taxes and birthweight) or "*Cigarette taxes only affect birth weight indirectly through their effect on maternal smoking*

   - Sounds reasonable as the level of cigarette taxes across states are quite random
   - In this case need to ensure that this somewhat random distribution of taxes is unrelated with birthweights
   - Imagine if taxes were systematically higher where pollution/health expenditure was lower and these differences were not controlled for? – not as easy to find IVs as you think!

Workplace productivity

- Imagine we want to determine the relationship between having children and workplace productivity

  - Likely to observe a negative relationship but estimate will obviously be biased in a regression framework

  - Experimental design is obviously not ethical or feasible

- Need an instrumental variable – gender concordance

- *Gender concordance*: if a couple's first two children are of the same gender, couples are much more likely to have a third child

  - in other words if you were to compare families whose first two children were the same sex, you would find that they would *on average* have larger families, than families whose first two children consisted of a boy and a girl

- Analysists have used gender concordance of a couple's first two children as an instrumental variable (e.g. 1 if first two children are the same, 0 otherwise)

Workplace productivity

- Intuitively is this likely to be a good instrument?

1. Will be related with variable of interest - number of children couples have (endogenous variable of interest)

   - Drawing on other research there are strong reasons to expect that, all things being equal, a couple who's first two children are the same gender will have bigger families

2. Will gender concordance be unrelated to workplace productivity except though its relationship with number of children (i.e. no direct relationship)

   - Very likely: the sex of your children follows a random process and unlikely to be related to workplace productivity

One of the most widely known examples in the economics literature

- *Angrist (1990): American Economic Review Lifetime earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records*

- Research question: Does serving in the military negatively affect future earnings?
  - can't simply compare earnings of those who serve versus those who don't – why?

One of the most widely known examples in the economics literature

- *Angrist (1990): American Economic Review Lifetime earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records*

- Research question: Does serving in the military negatively affect future earnings?

  - can't simply compare earnings of those who serve versus those who don't – why?

  - People who serve in the military and those who don't will differ in a large number of ways and these differences could explain any differences in earnings

    - can try to control for differences using regression analysis ***but difficult to control for everything here***

Draft: Lottery – the process of which placed otherwise similar men into a treatment group (those who drafted) and a control group (those who weren't) by **randomly** selecting birth dates of the year

- *In other words individuals did not choose to join*

This randomisation meant that people who served in the military as a result of being drafted were similar in observed and unobserved characteristics to those who didn't.

- thus the draft served as a useful instrumental variable

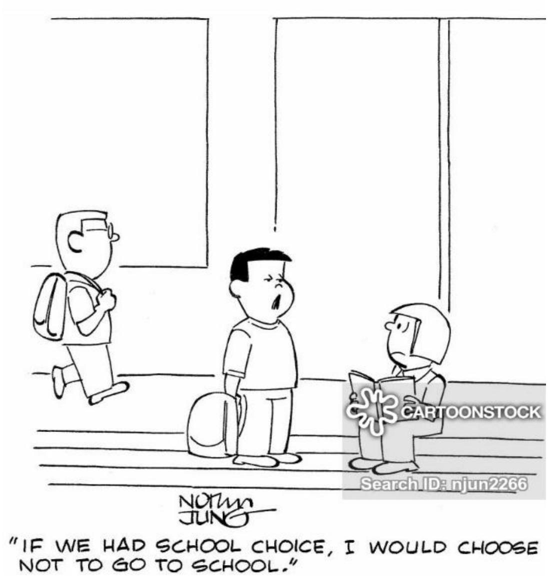
Findings suggest that draftees earn significantly less (15%) than they would have if not drafted

Does compulsory schooling affect earnings?

- Possible any examination of the relationship between education and earnings will be affected by omitted variable bias

  - innate ability, motivation, family connections etc.
  - can 'control' for some of these but not all.

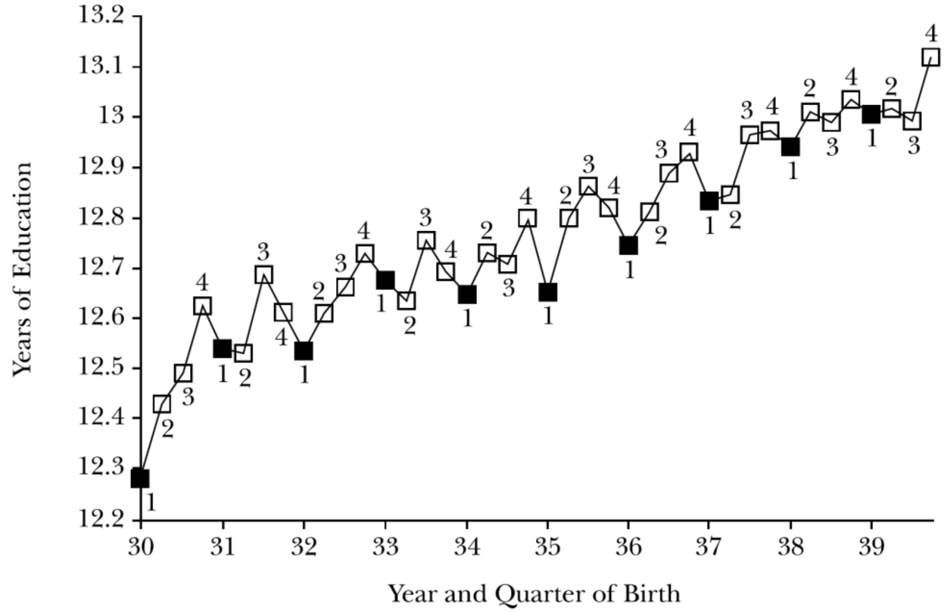
- Angrist and Krueger (1991) used date of birth as an instrumental variable as the combination of school start age policies and compulsory schooling laws created a natural experiment in which children **are compelled** to attend school for different lengths of time **depending on their birthdays**.
  - in short your date of birth predicted how long you were compelled to stay in school



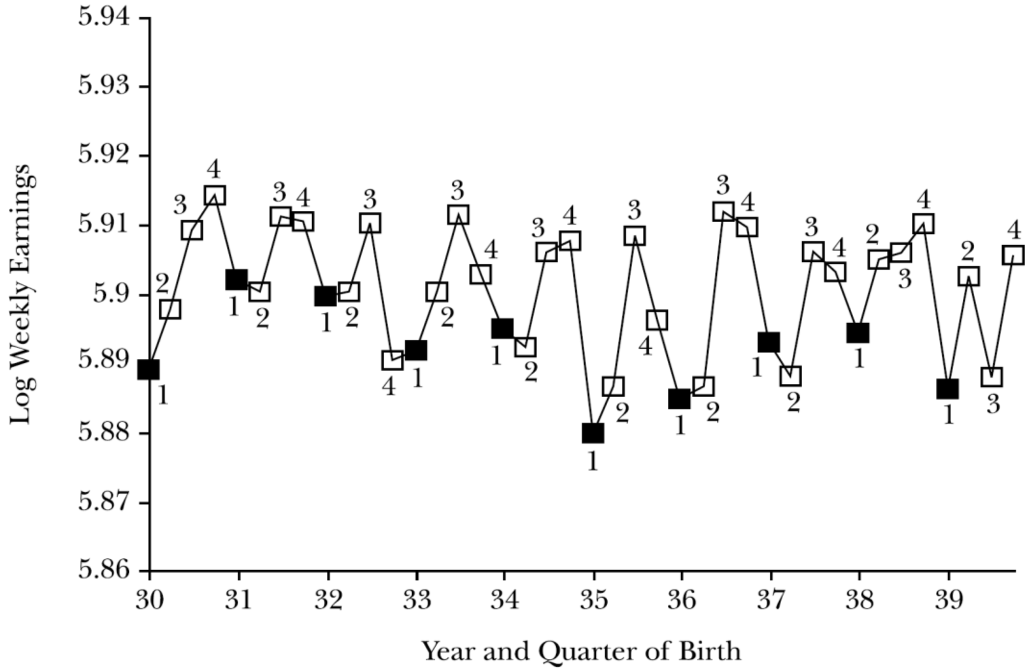
- Because an individual's date of birth is probably unrelated to earnings (e.g. should not be correlated with omitted variables such as a person's innate ability, motivation or family connections) DOB should provide a valid instrument

Angrist and Krueger 1991

Figure 1
Mean Years of Completed Education, by Quarter of Birth



Mean Log Weekly Earnings, by Quarter of Birth



Statistical analysis

- Relatively straightforward analytical procedure once you have the intuition behind the procedure right

- Two stage process (typically called two stage least squares) – sounds complicated but its not

  - **First stage** is that you **model your endogenous variable as a function of your instrumental variable** and a vector of control variables – e.g. regress X (endogenous variable) Z (instrumental variable) other controls. You use this to estimate your predicted values for your endogenous explanatory variables. Think of these predicted as opposed to actual X values as being purged of omitted variable bias or other sources of endogeneity

  - **Second stage** is that you plug these predicted values into your main equation, i.e. the predicted X values purged of endogeneity simply replace your original values for your endogenous variable

- In practice STATA will do all this for you in one step – Ivreg function

  - Lots of material online but we will go through this in seminar

Recap

- An instrumental variable analysis can help you address causality concerns in the face of endogeneity

  - **Omitted variable bias**

  - Bi-directional causality

  - Measurement error

- If you suspect that despite your best efforts (e.g. control variables) endogeneity is still a problem, i.e. your estimate of the relationship between X (e.g. education) and Y (earnings) is affected by, for example, omitted variable bias (e.g. ability) then a valid instrumental variable may allow you to purge your estimates of this bias

- Trick is to find a variable that is correlated with your endogenous explanatory (X) variable but one that is not related to Y except indirectly through its association with X

  - not as easy as it sounds

- Instruments often (not always) derive from natural experiments and as such can be seen as an extension of the natural experimental methods (e.g. Diff-in-Diff we discussed last week

  - often some law, policy or some other 'natural' event (e.g. weather) will provide you with the variable that will be correlated with your key explanatory variable but not with your outcome variable

External reading

- Some very accessible descriptions of Instrumental variable analysis
- Causal Inference via Natural Experiments and Instrumental Variables
- Instrumental variables and the search for identification