MQE: Economic Inference from Data: Module 3: Instrumental Variables

Claire Duquennois

6/9/2020

You can't always get what you want

Even with fixed effects, certain types of unobservables can still bias our estimates.

For OVB to not be a problem, we want a treatment variable x_i where we know that there does not exist some omitted variable x_{ov} such that

- $cor(x_i, x_{ov}) \neq 0$
- and $cor(y_i, x_{ov}) \neq 0$.

This is a tall order...

You can't always get what you want

But if you try <u>sometimes</u>,

You can't always get what you want

But if you try <u>sometimes</u>,

you just <u>might</u> find,

You can't always get what you want

But if you try sometimes,

you just might find,

you get what you need: a good instrumental variable.

An instrument for what?

I am interested in the relationship between y and x_1.

The true data generating process looks like this:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- x_i and x_2 are uncorrelated with ϵ
- x_i and x_2 they are correlated with each other such that $Cov(x_1, x_2) \neq 0$

So whats the problem?

- you don't actually observe x_2.

Uh oh.

The problem:

The naive approach (but you of course know better then to do this. . .)

Regress y on just x_1:

$$y_i = \beta_0 + \beta_1 x_1 + \nu$$

where

$$\nu = \beta_2 x_2 + \epsilon.$$

The problem:

$$
\begin{aligned}
\hat{\beta}_{1,OLS} &= \frac{cov(x_1, y)}{var(x_1)} \\
&= \frac{cov(x_1, \beta_0 + \beta_1 x_1 + \nu)}{var(x_1)} \\
&= \frac{cov(x_1, \beta_0) + cov(x_1, \beta_1 x_1) + cov(x_1, \nu)}{var(x_1)} \\
&= \frac{\beta_1 var(x_1) + cov(x_1, \nu)}{var(x_1)} \\
&= \beta_1 + \frac{cov(x_1, \nu)}{var(x_1)}
\end{aligned}
$$

$cov(x_1, \nu) \neq 0 \Rightarrow \hat{\beta}_{1,OLS}$ is biased.

All is not lost!

An **instrumental variable** (IV) is a variable that

- is correlated with the "good" or "*exogenous*" variation in x_1
- is unrelated to the "bad" or "*endogenous*" or "*related-to-*x_2" variation in x_1.

Formally

An IV is a variable, z that satisfies two important properties:

- $Cov(z, x_1) \neq 0$ (the first stage).
- $Cov(z, \nu) = 0$ (the exclusion restriction).

The First Stage

$Cov(z, x_1) \neq 0$

- z and x_1 are correlated
- the IV is useless without a first stage.

We are trying to get a $\hat{\beta}_1$ such that $E[\hat{\beta}_1] = \beta_1$. If our instrument is totally unrelated to x_1, we won't have any hope of using it to get at β_1.

The exclusion restriction

$Cov(z, \nu) = 0$

- z has to affect y **only** through x_1.
- $\Rightarrow Cov(z, \epsilon) = 0$ (because we've already assumed that x_2 is uncorrelated with ϵ).

The IV estimator

$$
\begin{aligned}
\hat{\beta}_{1,IV} &= \frac{cov(z,y)}{cov(z,x)} \\
&= \frac{cov(z,\beta_0+\beta_1 x_1+\nu)}{cov(z,x_1)} \\
&= \beta_1 \frac{cov(z,x_1)}{cov(z,x_1)} + \frac{cov(z,\nu)}{cov(z,x_1)} \\
&= \beta_1 + \frac{cov(z,\nu)}{cov(z,x_1)}.
\end{aligned}
$$

With the exclusion restriction: $cov(z,\nu)=0 \Rightarrow E[\hat{\beta}_{1,IV}]=\beta_1$

Huzzah! We have an unbiased estimator!