Two stage least squares and IV

EC420 MSU

Justin Kirkpatrick
Last updated March 17, 2020

Recall on Monday we:

- Covered continuou × ntinuous interactions

"What's in the error"

Introduced the Instrumental Variable (IV)

D

ariation D

Which lets *ident* the effect D Y

3 conditions for a valid IV

Today we will

- ~~Review the concept of IV~~
- What does "as-good-as-randomly assigned" mean?
- Estimate an IV using Two-Stage Least Squares (2SLS)
- Testing our IV requirements
- Calculate the $se(\hat{\beta}^{IV})$
  - That's the "IV estimator of β"
  - We could call our original OLS estimator $\hat{\beta}^{OLS}$ Biased

When we want an *unbiased* estimate of:

$$ATE = E[Y_1 - Y_0]$$

But all we have is observed data:

$$E[Y_1|D = 1] - E[Y_0|D = 0]$$

And we're worried about selection bias if we use our observed data because D is

- Possibly correlated with something in the error term
- The potential outcomes $(Y_{i0}, Y_{i1}) \not\perp D$
  - Both of these mean the same thing / are the same problem

Concept of IV

We introduced the term *endogenous*

- "Endogenous" means "determined within the system"
- When treatment may be affected by something within our model, we say it may be "endogenous"
  - When $(Y_{i0}, Y_{i1}) \not\perp D$, then D is endogenous
  - Y_{i0}, Y_{i1} is *within the system

And the term *exogenous*

- "Exogenous" means "determined outside the system"
- Nothing in our regression helped determine D
  - Not Y_{i1} or Y_{i0} or anything in u

Conditionally exogenous is fine:

- D could be conditionally exogenous, conditional on some controls X

Exogenous $\Rightarrow$ as good as randomly assigned

And note that:

Any x can be endogenous or exogenous. This isn't limited to D. Whether or not endogeneity is a problem depends on context.

$E[Y|D=1] - E[Y|D=0]$ is the same as β_1 in:

$$y = \beta_0 + \beta_1 D + u$$

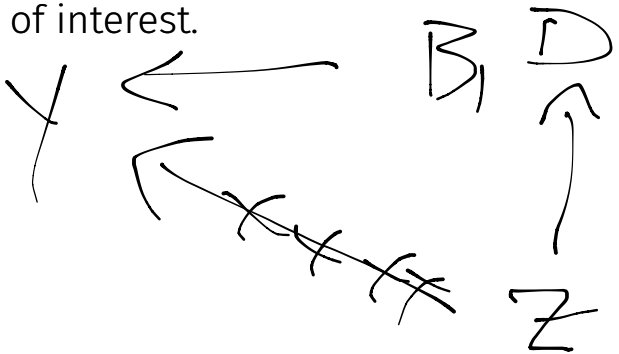
- D is our variable of interest, our *treatment*
- Y is our outcome

But we are afraid that something in u is correlated with getting treatment, D

- $E[u|D] \neq 0$
- D is endogenous

Then, to get an unbiased estimate of β_1 we need an instrument, Z that:

- Determines or affects D
- Is "as good as (conditionally) randomly assigned" (is conditionally exogenous)
- But does not correlate with u or only affects y through D (same thing)
  - When its uncorrelated, then $E[Y_0|Z=1] - E[Y_0|Z=0]$ is 0, and no selection bias
  - Except this just tells us about the potential outcomes over Z, not D, our variable of interest.



Recall our KIPP example:

We were interested in calculating the *{Effect of attending KIPP on test scores}*

This would be:

$$ATE = E[TestScore_{i1} - TestScore_{i0}]$$

We could calculate:

$$E[TestScore|KIPP == 1] - E[TestScore|KIPP == 0]$$

But, in our selection bias section, we learned that this is not going to get us the actual effect of KIPP because of self-selection into KIPP.

But we know that:

{Effect of winning lottery on test scores} =

{Effect of winning on attending KIPP} × **{Effect of attending KIPP on test scores}**

Written another way that may make more sense:

$$\underbrace{\frac{\text{Change in test scores}}{\text{Change in lottery win}}}_{\text{Observed and Unbiased}} = \underbrace{\frac{\text{Change in test scores}}{\text{Change in attendance}}}_{\text{Thing we're trying to measure without bias}} \times \underbrace{\frac{\text{Change in attendance}}{\text{Change in lottery win}}}_{\text{Observed and Unbiased}}$$

But what is that first term? It is:

$$E[TestScore|Lottery == 1] - E[TestScore|Lottery == 0]$$

We can take the sample analog of E and calculate the mean of $TestScore$ for those who won the lottery and those who didn't (ignoring attendance). This is the same as a regression:

$$TestScore = \phi_0 + \phi_1 Lottery + v$$

To clarify that point:

$$y = \phi_0 + \phi_1 Z + v$$

- $y = TestScore$
- $Z = Lottery$
- $D = KIPP$ (attendance)

Is just different notation for:

$$TestScore = \phi_0 + \phi_1 Lottery + v$$

And:

$$\begin{aligned}\beta_1 &= E[TestScore|Lottery == 1] - E[TestScore|Lottery == 0] \\ &= E[y|Z == 1] - E[y|Z == 0]\end{aligned}$$

When Z is as good as randomly assigned, which we'll get to shortly.

We have unbiased estimate of:

$$E[y|Z == 1] - E[y|Z == 0]$$

Which is the same as ϕ_1 in:

$$y = \phi_0 + \phi_1 Z + v$$

And we have **unbiased**

$$E[D|Z == 1] - E[D|Z == 0]$$

Which we can also get from γ_1 in a regression:

$$D = \gamma_0 + \gamma_1 Z + w$$

Because Z is *as good as randomly assigned* (more on that in a second)

Concept of IV

If we take ϕ_1 as $\frac{\text{Change in test scores}}{\text{Change in lottery win}}$:

y = test scores
D = KIPP Attendance
Z = Lottery

$$y = \phi_0 + \phi_1 Z + v$$

And if we take γ_1 as $\frac{\text{Change in attendance}}{\text{Change in lottery win}}$

$$D = \gamma_0 + \gamma_1 Z + w$$

Then we can calculate

$$\frac{\text{Change in test scores}}{\text{Change in attendance}} = \frac{\text{Change in test scores}}{\text{Change in lottery win}} \Big/ \frac{\text{Change in attendance}}{\text{Change in lottery win}}$$

ϕ_1, γ_1

as:

$$\frac{\phi_1}{\gamma_1} = \beta_1^{IV}$$

$y = \beta_0 + \beta_1 D + u$

As good as random

IV requires that Z is "as good as randomly assigned, conditional on x's"

What does "as good as randomly assigned" mean?

- In our KIPP example, Z was randomly assigned because it was the result of a lottery.
- But can we have a Z that is not totally randomly assigned?

Yes, we can "Put into regression as x"

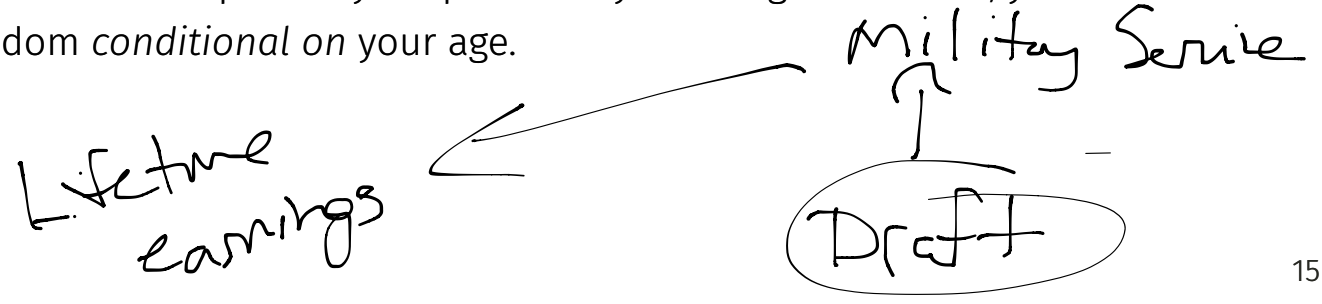
- As long as we *control* for all the things that might not be random.
- That is, once we have statistical controls in our regression that explain part of how Z affects D, the rest of the variation in Z is uncorrelated with anything else in our regression.
- Example on next slide:

As good as random

Papers by the MM authors looked at the effect of military service on lifetime earnings. Since "people who may be inclined to join the army" may have some unobserved characteristics that might also affect military service (e.g. selection bias, $E[u|X] \neq 0$), the authors used people's draft numbers during Vietnam to *instrument* for likelihood of serving in the military.

Draft numbers were randomly assigned. A lower draft number meant you were more likely to be drafted. Some people with low draft numbers joined right away (you got to choose your assignment if you enlisted voluntarily); some always joined (always-takers!).

If drafts were done within age groups - that is, the draft board worked it's way up the numbers until each age group's quota was filled - then your age plus your number would also predict your probability of being drafted. So, your draft number was random *conditional on* your age.



As good as random

so...

$$Drafted = \gamma_0 + \gamma_1 LowDraftNumber + \gamma_2 AgeGroup + w$$

Conditional on age, draft number was as good as randomly assigned.

- D is $Drafted$
- Z is $LowDraftNumber$
- $AgeGroup$ is x's.

And "conditional on x's, Z is as good as randomly assigned"

$$Earnings = \beta_0 + \beta_1 Drafted/Service$$

Define "Identification"

The term **identification** is used a lot in econometrics. With our β^{IV}, we would say we have *identified* β_1.

identification means we can write the parameter of interest, β, in terms of population moments.

- $\frac{\phi}{\gamma}$ is in terms of population moments because it is $\frac{\frac{Cov(Y,Z)}{Var(Z)}}{\frac{Cov(D,Z)}{Var(Z)}}$
- *identification* is a population-moments concept. It is a statement about the population parameter, β.
- Of course, if the population moment isn't identified, then our sample analog, $\frac{\hat{\phi}}{\hat{\gamma}}$ is useless.

On that last slide:

$$\beta_1^{IV} = \frac{\frac{Cov(Y,Z)}{Var(Z)}}{\frac{Cov(D,Z)}{Var(Z)}} = \frac{\phi_1}{\gamma_1}$$

If we cancel things out...

$$\beta_1^{IV} = \frac{Cov(Y,Z)}{Cov(D,Z)}$$

Remember our first requirement: *relevant first stage*

If Z has no effect on D, what is $Cov(D,Z)$?

So we see where that comes from! We'll point out the other two IV requirements when we run across them

$\frac{\phi}{\gamma}$ is not useful when you have multiple x's

So we have a different *method* of estimating β^{IV}:

First Stage: regress D on Z, X:

$$D = \gamma_0 + \gamma_1 Z + \gamma_2 x_1 + \cdots + w$$

Since w is uncorrelated with things correlated with Z, ϕ_1 is unbiased.

Next, generate $\hat{D}$:

$$x_1 = \hat{\delta}_0 + \hat{\delta}_1 x_2 + \hat{V}$$
$$\hat{V} = x_1 - \hat{\delta}_0 - \hat{\delta}_1 x_2$$

Take $\hat{\phi}$, your estimates and get the predicted value of D, $\hat{D}$

$$\hat{D} = \hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{\gamma}_2 x_1 + \cdots$$

This has a "partialling out" flavor

Exogenous

- w contains all the variation in D that does not explained by Z.
- And likewise $\hat{D}$ contains all the variation in D that **is** explained by Z (w is not in $\hat{D}$)

Remember

- IMPORTANT: our problem in the first place was that D was correlated with something unobserved in the error.
- But $\hat{D}$ is entirely from Z and other exogenous statistical controls, and Z is not correlated with this unobserved problem by *the exclusion restriction*

For our second stage:

$$Y = \beta_0 + \beta_1 \hat{D} + \beta_2 x_2 + \cdots + u$$

And the estimate of β_1 here is $\hat{\beta}_1^{IV}$, our coefficient of inerest!

β_1 is unbiased!

We estimated β_1 using *only variation in D associated with Z, and Z is (assumed to be) uncorrelated with u, solving our initial problem.

What if we have multiple endogenous variables?

Wooldridge uses y_1 for the outcome, and $y_2, \cdots$ on the right hand side for an endogenous variable that we need to instrument. Wooldridge uses z for all exogenous variables, instruments or not. I prefer to use x since that leaves z for instruments only

In the regression:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 x_1 + \beta_4 x_2 + u$$

The problems are y_2 and y_3. Since there are two, **we need two valid instruments**, z_1 and z_2.

In our first stage, we also include the x's from the main regression:

$$y_2 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 x_1 + \gamma_4 x_2 + v$$

Gives us $\hat{y}_2$, while:

$$y_3 = \kappa_0 + \kappa_1 z_1 + \kappa_2 z_2 + \kappa_3 x_1 + \kappa_4 x_2 + w$$

Which gives us $\hat{y}_3$. Finally, we get a second stage β^{IV}:

$$y_1 = \beta_0 + \beta_1 \hat{y}_2 + \beta_2 \hat{y}_3 + \beta_3 x_1 + \beta_3 x_2 + u$$

Those are the same x's in each regression, no hats.

A variable is:
1. Endogenous (D, y_2 & y_3)
2. Exogenous & a control — x_1, x_2
   - all these are "as good as randomly assigned"
   - But are not instruments b/c they affect y directly
3. Exogenous & Instruments (z_1, z_2)

No z's in 2nd Stage

A couple notes:

- The instruments do not appear in the second stage ($\hat{y}_2$ and $\hat{y}_3$ are perfectly colinear with Z).
- The exogenous statistical controls, x_1, x_2 appear in both stages.
  - In fact, if they are used in the first, then left out of the second, the result can be biased.

We have to have at least one exogenous instrument z for every endogenous variable.

When we have 2 endogenous variables, we need 2 instruments with two first-stage regressions that both meet the *relevant first stage* requirement.

We can have more than one instrument per endogenous variable!

We say we are *overidentified* when this is the case. We can test this (Hausman Test - W15.5)

Overidentification is not necessarily a bad thing, despite the name.

First, Z must have a causal effect on D

The instrument, Z, has to have a casual effect on the variable of interest, D.

- In our example, this means Z has to *change enrollment* (in expectation)
- This is the *relevant first stage* requirement or condition

Testing this requirement

- We need a test that tells us if all the variables in our model of $D = \gamma_0 + \gamma_1 Z + w$ are any better at predicting D than just guessing $\gamma_0 = \bar{D}$.
- Sound familiar?

Relevant First Stage test

- In a single variable regression (one instrument), then we can just test if $\gamma_1 = 0$
.
- If we have two instruments or more, then we use the F test for:
$D = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + w.$
- Rule of thumb: F-stat needs to be > 10 (Stock and Yogo)

$D = \gamma_0 + u$ (no Z's)

"Weak Instrument" = maybe no "relevant 1st stage"

"conditionally random" is ok.

Second, Z must be as good as randomly assigned

The instrument, Z, cannot be determined by the omitted variable / selection bias we're trying to get out.

- This is the *independence requirement*

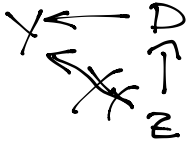
This one, we can't test for.

This is for the same reason that we can't just test for $E[u|X] = 0$, there might still be something out there that is correlated with Z and is also correlted with D, which means Z is not randomly assigned.

Instead, we have to make a compelling argument for this to be true.

Third, Z must affect the outcome only through the variable of interest

The instrument, Z, cannot have a direct affect on the outcome



- This is the *exclusion restriction*.

This one, we can't directly test for either

We have to make a compelling argument for this to be true.

If we have multiple instruments, we can sort of test this, though...

(over identification)

But, if we have many instruments available (more than we have endogenous variables), we can choose which one is best

We do this by comparing the β^{IV} estimated with one instrument, then the other. They *should* give the same result if they both meet the requirements, right? If they're different (statistically speaking), then we have a problem. One (or both) doesn't meet the *exclusion restriction*.

Don't assume failing-to-reject this test means the instruments meet the *exclusion restriction*.

The overidentification (Hausman) Test

First, take the $\hat{u}$ from your second stage

Remember, this is $y - \hat{y}$ where $\hat{y}$ is from the 2nd stage.

If all of your instruments meet the *exclusion restriction*, then they should not be correlated with this residual, $\hat{u}$.

$$\hat{u} = \eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 x_1 + \cdots + \varepsilon$$

The R^2 tells us whether or not all of the (hopefully exogenous) z's and x's on the RHS are unrelated to $\hat{u}$.

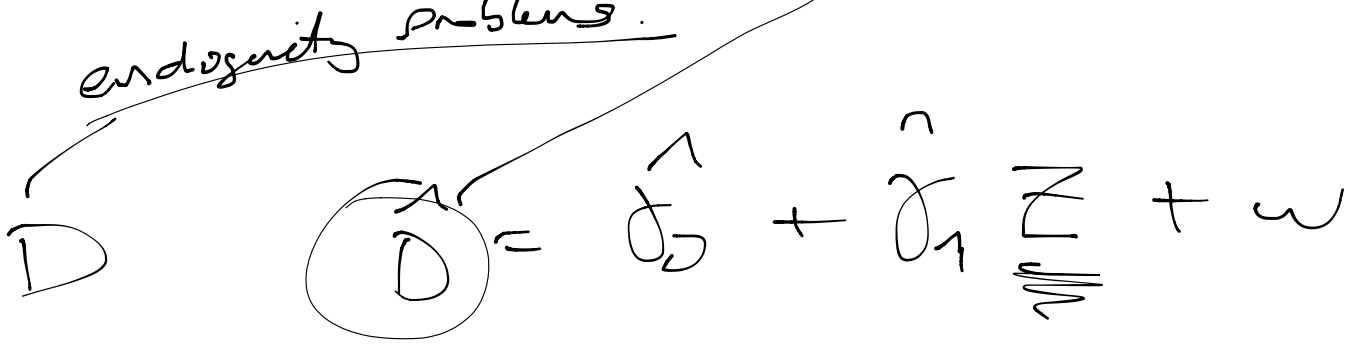
$nR^2 \sim \chi^2_q$ where q is the number of instruments minus the number of endogenous variables.

What about R^2 of the second stage?

$$y = \beta_0 + \beta_1 \hat{D} + \beta_2 x_1 + \cdots + u$$

As Wooldridge states, R^2 from IV is not very useful; we aren't trying to explain more variation, we're trying to use a specific subset of the variation to get at a causal estimate.

We ignore it.



se of IV estimator

The whole point of this was to do inference on an *unbiased* estimate of $\hat{\beta}$, which is $\hat{\beta}^{IV}$.

But we used two stages, so what is the se(\hat{\beta^{IV}})? We can't use just the 2nd stage, right?

- (No, we can't)

First, we assume:

Something-like-MLR5: $Var(u|z) = \sigma^2$

- Same as before, but z instead of x
- Still homoskedasticity
- Still have a robust version

se of IV estimator

For one endogenous x, and one instrument z

$$se(\hat{\beta}_1^{IV}) = \frac{\sigma_u^2}{n\sigma_x^2\rho_{x,z}^2}$$

- $n\sigma_x^2$ looks a lot like SST_x since $\hat{\sigma}_x^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2$

What is $\rho_{x,z}$?

It is the *correlation coefficient* of x and z:

$$\rho_{x,z} = \frac{Cov(X,Z)}{\sqrt{Var(X)}\sqrt{Var(Z)}}$$

if X, Z perfectly corr
$\therefore Cov(X,Z) = Var(X)$
$\rho = \frac{Var(X)}{\sqrt{Var(X)} * \sqrt{Var(X)}} = 1$
$\rightarrow \rho_{xz}^2 = 1$

se of IV estimator

Another way of writing $se(\hat{\beta}_1^{IV})$ is:

$$se(\hat{\beta}_1^{IV}) = \frac{\hat{\sigma}^2}{SST_x R^2_{x,z}}$$

Where $R^2_{x,z}$ is the R^2 in the **first-stage** regression.

- If that $R^2_{x,z}$ is exactly 1, then this is the same as $\hat{\beta}^{OLS}$.
  - What would it mean if $R^2_{x,z} = 1$?
- If $R^2_{x,z}$ is small, then what happens?
- Does that make sense?

The AER package in R has an excellent function

```
install.packages('AER')
```

```
ivreg(Y ~ D + X1 | X1 + Z1 + Z2, data=df)
```

- The first part of the forumla is just like an OLS regression
- Here, I'm following lecture notation:
  - Y is the outcome of interest.
  - $D$is the endogenous variables of interest
  - $X1$ is an exogenous statistical control
  - $Z1$ and $Z2$ are the instruments

$$X_1 = \beta_0 + \beta_1 X_1 + u$$

$$\beta_1 = 1$$

Note that this has more than one instrument (Z1, Z2)

Note that X1 *instruments for itself*. R will instrument everything to the left of the |

$$\hat{X}_1 = X_1$$

In our KIPP example:

```
IVmodel = ivreg(TestScores ~ KIPP + year | year + Lottery,
data=df)
```

I've added *parentalIncome* as an exogenous variable. Time is always exogenous. All instruments will be used to instrument $KIPP$ and $year$.

We can still use our robust standard errors

```
coeftest(IVmodel, vcov = vcovHC(IVmodel, 'HC1')
```

$R^2 = .33$, "The model explains 33% of the Variation in X_1"

R^2 from reg. of

$$x_1 = \delta_0 + \delta_1 x_2 + v$$

$$se(\hat{\beta}_1) = \sqrt{\frac{\sigma_u^2}{SST_x}}$$

$$se(\hat{\beta}_1) = \sqrt{\frac{\sigma_u^2}{SST_{x_1}(1 - R_j^2)}}$$

when: $y = \beta_0 + \beta_1 x_1 + u$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

$$se(\bar{x}) = \sqrt{\sigma_y^2 / n} = \sigma_x / \sqrt{n}$$

$\bar{x}$: mean