Violation of the exogeneity assumption and the 2SLS estimator

Econometrics (35B206), Lecture 5

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2018

LRM, exogeneity assumption

One of the assumptions of the LRM is that ε_i is exogenous. Here we relax this assumption. Relaxing an assumption of a model means changing to a new model that can accommodate the new situation about ε_i. It also means that we need a new estimator to estimate β. We now study the new model, and the new estimator.

LRM, exogeneity assumption

While we relax the exogeneity assumption and study its implications, we retain the other assumptions: linearity, full rank, homoskedasticity, random sampling, normality.

LRM, exogeneity assumption

The strict exogeneity assumption states that

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_k\right] = 0.$$

$\boldsymbol{x}_k$ contains n observations for variable k. It says that the mean of ε_i at observation i is independent of the explanatory variable k observed at any observation, including i.

LRM, exogeneity assumption

The weak exogeneity assumption states that

$$\mathrm{E}\left[\varepsilon_i \mid x_{ik}\right] = 0.$$

x_{ik} is the observation i for variable k. Hence, we do not consider all n observations of variable k, denoted by $\boldsymbol{x}_k$, but just the observation i, denoted by x_{ik}.

LRM, exogeneity assumption

Generalising

$$\mathrm{E}\left[\varepsilon_i \mid x_{ik}\right] = 0$$

to K variables, we consider

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}.$$

LRM, exogeneity assumption

In this lecture we will consider violation of

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0},$$

but still assume that

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_j\right] = \mathbf{0}.$$

That is, we violate weak exogeneity, and not strict exogeneity.

LRM, exogeneity assumption, meaning

A first implication of

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}$$

is that, by the LIE,

$$\begin{aligned}
\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] &= \mathrm{E}_{\boldsymbol{x}_i}\left[\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i \mid \boldsymbol{x}_i\right]\right] \\
&= \mathrm{E}_{\boldsymbol{x}_i}\left[\boldsymbol{x}_i \mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\
&= \mathbf{0}.
\end{aligned}$$

When referring to exogeneity, we will use the latter instead of the former because we can use it when talking about covariance, or because violating it is enough to violate the consistency of the OLS estimator (Lecture 2). Keep in mind that when the latter is ever stated, it is because the former holds.

LRM, exogeneity assumption, meaning

A second implication of

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}$$

is that, by the LIE,

$$\begin{aligned}\mathrm{E}\left[\varepsilon_i\right] &= \mathrm{E}_{\boldsymbol{x}_i}\left[\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\ &= 0.\end{aligned}$$

It says that if the average of ε_i at all slices of the population determined by the values of $\boldsymbol{x}_i$ equals zero, then the average of these zero conditional means must also be zero.

LRM, exogeneity assumption, meaning

A third implication of

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}$$

is that, using the above results,

$$\begin{aligned}
\mathrm{Cov}\left[\varepsilon_i, \boldsymbol{x}_i\right] &= \mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] - \mathrm{E}\left[\varepsilon_i\right] \mathrm{E}\left[\boldsymbol{x}_i\right] \\
&= \mathbf{0} - \mathbf{0}\mathrm{E}\left[\boldsymbol{x}_i\right] \\
&= \mathbf{0}.
\end{aligned}$$

That is, ε_i are $\boldsymbol{x}_i$ are not correlated.

LRM, exogeneity assumption, meaning

So when you see

$$\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] = \mathbf{0},$$

and

$$\mathrm{E}\left[\varepsilon_i\right] = 0,$$

this means that ε_i and $\boldsymbol{x}_i$ are not correlated. Or, when you see

$$\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] \neq \mathbf{0},$$

and

$$\mathrm{E}\left[\varepsilon_i\right] = 0,$$

this means that ε_i and $\boldsymbol{x}_i$ are correlated.

LRM, exogeneity assumption, meaning

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}$$

implies

$$\mathrm{E}\left[\varepsilon_i\right] = 0.$$

This does not mean that the LHS of the two terms are equal to each other. We now assume this so that

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathrm{E}\left[\varepsilon_i\right].$$

It says that the average of ε_i is the same across all slices of the population determined by the values of $\boldsymbol{x}_i$. This then necessarily means that the common average is equal to the average of ε_i over the entire population. That is, values of $\boldsymbol{x}_i$ do not determine the average value of ε_i. We then say that ε_i is mean independent of $\boldsymbol{x}_i$.

LRM, exogeneity assumption, meaning

Consider

$$\mathsf{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathsf{E}\left[\varepsilon_i\right],$$

or

$$\mathsf{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}.$$

Both are statements of the mean independence. What is the econometric interpretation?

LRM, exogeneity assumption, meaning

For any function of $\boldsymbol{x}_i$ and ε_i (flexible case), by the LIE,

$$\begin{aligned}
\mathrm{E}\left[g(\boldsymbol{x}_i)h(\varepsilon_i)\right] &= \mathrm{E}_{\boldsymbol{x}_i}\left[\mathrm{E}\left[g(\boldsymbol{x}_i)h(\varepsilon_i) \mid \boldsymbol{x}_i\right]\right] \\
&= \mathrm{E}_{\boldsymbol{x}_i}\left[g(\boldsymbol{x}_i)\mathrm{E}\left[h(\varepsilon_i) \mid \boldsymbol{x}_i\right]\right] \\
&= \mathrm{E}\left[g(\boldsymbol{x}_i)\mathrm{E}\left[h(\varepsilon_i)\right]\right] \\
&= \mathrm{E}\left[g(\boldsymbol{x}_i)\right]\mathrm{E}\left[h(\varepsilon_i)\right]
\end{aligned}$$

if ε_i and $\boldsymbol{x}_i$ are independent, meaning that

$$\mathrm{E}\left[h(\varepsilon_i) \mid \boldsymbol{x}_i\right] = \mathrm{E}\left[h(\varepsilon_i)\right].$$

It says that all unconditional moments of ε_i are equal to the conditional moments of ε_i. If ε_i is mean independent of $\boldsymbol{x}_i$, that is

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathrm{E}\left[\varepsilon_i\right],$$

the equality for the general function does not hold. Mean independence is weaker than independence!

LRM, exogeneity assumption, meaning

For $\boldsymbol{x}_i$ and ε_i (restrictive case),

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{x}_i \varepsilon_i\right] &= \mathrm{E}_{\boldsymbol{x}_i}\left[\mathrm{E}\left[\boldsymbol{x}_i \varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\ &= \mathrm{E}_{\boldsymbol{x}_i}\left[\boldsymbol{x}_i \mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\ &= \mathrm{E}\left[\boldsymbol{x}_i \mathrm{E}\left[\varepsilon_i\right]\right] \\ &= \mathrm{E}\left[\boldsymbol{x}_i\right] \mathrm{E}\left[\varepsilon_i\right] \end{aligned}$$

if ε_i is mean independent of $\boldsymbol{x}_i$, that is

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathrm{E}\left[\varepsilon_i\right].$$

Mean independence implies that ε_i and $\boldsymbol{x}_i$ are uncorrelated. Mean independence is stronger than uncorrelatedness!

LRM, exogeneity assumption, meaning

For any function of $\boldsymbol{x}_i$, and for ε_i (less restrictive case),

$$\begin{aligned}
\mathrm{E}\left[g(\boldsymbol{x}_i)\varepsilon_i\right] &= \mathrm{E}_{\boldsymbol{x}_i}\left[\mathrm{E}\left[g(\boldsymbol{x}_i)\varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\
&= \mathrm{E}_{\boldsymbol{x}_i}\left[g(\boldsymbol{x}_i)\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right]\right] \\
&= \mathrm{E}\left[g(\boldsymbol{x}_i)\mathrm{E}\left[\varepsilon_i\right]\right] \\
&= \mathrm{E}\left[g(\boldsymbol{x}_i)\right]\mathrm{E}\left[\varepsilon_i\right]
\end{aligned}$$

if ε_i is mean independent of $\boldsymbol{x}_i$, meaning that

$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathrm{E}\left[\varepsilon_i\right].$$

If ε_i and $\boldsymbol{x}_i$ are independent, the equality for the less restrictive function holds. If ε_i and $\boldsymbol{x}_i$ are uncorrelated, the equality for the less restrictive function does not hold. Mean independence is in-between independence and uncorrelatedness!

Violation of the exogeneity assumption

If

$$\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] = 0$$

does not hold, we say that $\boldsymbol{x}_i$ is endogenous. When does it not hold? We consider three cases.

Violation of the exogeneity assumption, OVB case

Consider the LRM

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i.$$

Suppose that

$$\mathrm{E}\left[\varepsilon_i \mid x_{i1}\right] = 0,$$

and

$$\mathrm{E}\left[\varepsilon_i \mid x_{i2}\right] = 0.$$

Hence, the model is correctly specified.

Violation of the exogeneity assumption, OVB case

Now suppose that we do not observe x_{i2} so that it enters the error. Hence, the model becomes

$$y_i = x_{i1}\beta_1 + \varepsilon_i^*$$

where

$$\varepsilon_i^* = x_{i2}\beta_2 + \varepsilon_i.$$

Then,

$$\begin{aligned} \mathrm{E}\left[\varepsilon_i^* \mid x_{i1}\right] &= \mathrm{E}\left[x_{i2}\boldsymbol{\beta}_2 \mid x_{i1}\right] + \mathrm{E}\left[\varepsilon_i \mid x_{i1}\right] \\ &= \boldsymbol{\beta}_2 \mathrm{E}\left[x_{i2} \mid x_{i1}\right] + 0 \\ &\neq 0 \end{aligned}$$

if $\beta_2 \neq 0$ and $\mathrm{E}\left[x_{i2} \mid x_{i1}\right] \neq 0$. The former means that x_{i2} matters and enters the model. The latter means that x_{i1} and x_{i2} are correlated. The exogeneity assumption is violated for ε_i^*!

Violation of the exogeneity assumption, OVB case

What is the implication of

$$\mathrm{E}\left[\varepsilon_i^* \mid x_{i1}\right] \neq 0$$

for b_1 as the OLS estimator of β_1? Derive b_1 when x_{i2} is omitted in the true model:

$$\begin{aligned}
b_1 &= (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{y} \\
&= (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'(\boldsymbol{x}_1\beta_1 + \boldsymbol{x}_2\beta_2 + \boldsymbol{\varepsilon}) \\
&= \beta_1 + (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{x}_2\beta_2 + (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{\varepsilon}.
\end{aligned}$$

Taking the expectation conditional on $\boldsymbol{X}$,

$$\mathrm{E}\left[b_1 \mid \boldsymbol{X}\right] = \beta_1 + (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{x}_2\beta_2$$

since $\mathrm{E}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right] = \boldsymbol{0}$ in the true model.

Violation of the exogeneity assumption, OVB case

$$\mathrm{E}\left[b_1 \mid \boldsymbol{X}\right] = \beta_1 + (\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{x}_2\beta_2.$$

In two cases

$$\mathrm{E}\left[b_1 \mid \boldsymbol{X}\right] = \beta_1,$$

that is, b_1 is unbiased. First, if

$$(\boldsymbol{x}_1'\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1'\boldsymbol{x}_2 = 0,$$

meaning that there is no correlation between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in the sample, realising that the stated expression is the OLS coefficient estimate of $\boldsymbol{x}_1$ from the regression of $\boldsymbol{x}_2$ on $\boldsymbol{x}_1$. Second, if

$$\beta_2 = 0,$$

meaning that $\boldsymbol{x}_2$ does not enter the true model. Otherwise b_1 is subject to the omitted variable bias. The stated equation above is the OVB formula.

Violation of the exogeneity assumption, OVB example

Regress *wage* on *educ* where we ignore *exper* say because it is unobserved:

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

Violation of the exogeneity assumption, OVB example

In the regression we have ignored *exper*. We suspect that b_{educ} is biased. That is, we suspect that b_{educ} would change if we control for *exper* in the regression. Do you expect b_{educ} to have an upward or downward bias? Use the OVB formula to form an expectation:

$$\mathrm{E}\left[b_{educ} \mid \boldsymbol{educ}, \boldsymbol{exper}\right] = \beta_{educ} + (\boldsymbol{educ}'\boldsymbol{educ})^{-1}\boldsymbol{educ}'\boldsymbol{exper}\beta_{exper}.$$

We would expect

$$(\boldsymbol{educ}'\boldsymbol{educ})^{-1}\boldsymbol{educ}'\boldsymbol{exper}$$

to be negative (effect of exper on educ), and

$$\beta_{exper}$$

to be positive (effect of exper on wage). Hence, we should expect b_{educ} to have downward bias when we ignore *exper* in the true regression!

Violation of the exogeneity assumption, OVB example

Regress *wage* on *educ* and *exper*, and observe that b_{educ} increases. This confirms that b_{educ} has downward bias when *exper* is ignored in the previous regression.

```
. regress wage educ exper
```

Source	SS	df	MS
Model	10008.3629	2	5004.18147
Residual	28865.0661	994	29.0393019
Total	38873.429	996	39.0295472

Number of obs	=	997
F(2, 994)	=	172.32
Prob > F	=	0.0000
R-squared	=	0.2575
Adj R-squared	=	0.2560
Root MSE	=	5.3888

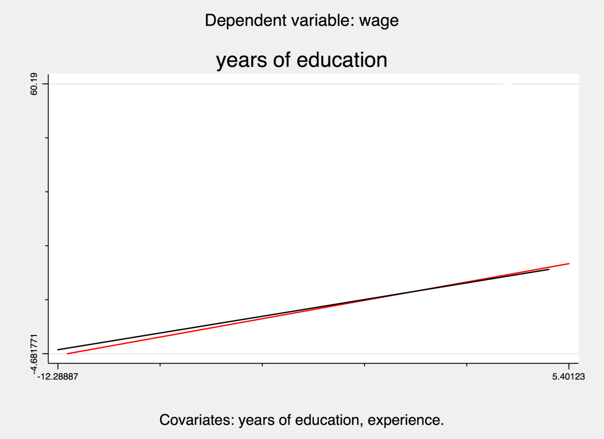
wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.246932	.0702966	17.74	0.000	1.108985	1.384879
exper	.1327808	.0153744	8.64	0.000	.1026108	.1629509
_cons	-8.833768	1.041212	-8.48	0.000	-10.87699	-6.790542

Violation of the exogeneity assumption, OVB example

The fitted line from the regression of *wage* on *educ*. The slope is b_{educ}, and it is biased because we ignore *exper*!

Violation of the exogeneity assumption, OVB example

Adding the fitted line from the regression of *wage* on *educ* after partialling out the effect of *exper* (red line). The slope is b_{educ}, and it is unbiased! The difference in the slopes is the size of the bias due to ignoring *exper* in the regression!



Violation of the exogeneity assumption, ME case

Consider the linear model

$$y_i = x_i^* \beta + \varepsilon_i.$$

Suppose x_i^* is the true variable we do not observe. Suppose we observe x_i, a noisy version of x_i^* with unobserved measurement error ω_i so that

$$x_i = x_i^* + \omega_i.$$

Since we observe only x_i, replace x_i^* in the model to obtain

$$y_i = x_i \beta \underbrace{-\omega_i \beta + \varepsilon_i}_{\varepsilon_i^*}.$$

x_i is correlated with ε_i^* due to ω_i. OLS estimator of β is subject to the measurement error bias.

Violation of the exogeneity assumption, SEM case

Consider the simultaneous equations model

$$y_{i1} = y_{i2}\alpha_1 + z_{i1}\beta_1 + \varepsilon_{i1},$$
$$y_{i2} = y_{i1}\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2}.$$

The constant is ignored in each equation for simplicity. Assume that

$$\mathrm{E}\left[\varepsilon_{i1} \mid z_{i1}, z_{i2}\right] = 0,$$
$$\mathrm{E}\left[\varepsilon_{i2} \mid z_{i1}, z_{i2}\right] = 0,$$

meaning that z_{i1} and z_{i2} are both uncorrelated with ε_{i1} and ε_{i2}. Also assume that

$$\mathrm{E}\left[\varepsilon_{i1}\right] = 0,$$
$$\mathrm{E}\left[\varepsilon_{i2}\right] = 0.$$

Suppose that our interest lies in estimating α_1 in the first equation.

Violation of the exogeneity assumption, SEM case

Solve the two equations for y_{i2}, in terms of z_{i1}, z_{i2}, ε_{i1}, and ε_{i2}. First replace y_{i1} in the equation for y_{i2}, and then solve for y_{i2} as

$$\begin{aligned}
y_{i2} &= y_{i1}\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2} \\
&= \left(y_{i2}\alpha_1 + z_{i1}\beta_1 + \varepsilon_{i1}\right)\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2} \\
&= y_{i2}\alpha_1\alpha_2 + z_{i1}\beta_1\alpha_2 + \varepsilon_{i1}\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2} \\
\left(1-\alpha_1\alpha_2\right) y_{i2} &= z_{i1}\beta_1\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i1}\alpha_2 + \varepsilon_{i2} \\
y_{i2} &= z_{i1}\frac{\beta_1\alpha_2}{1-\alpha_1\alpha_2} + z_{i2}\frac{\beta_2}{1-\alpha_1\alpha_2} + \varepsilon_{i1}\frac{\alpha_2}{1-\alpha_1\alpha_2} \\
&\quad + \varepsilon_{i2}\frac{1}{1-\alpha_1\alpha_2},
\end{aligned}$$

assuming that $\alpha_1\alpha_2 \neq 1$.

Violation of the exogeneity assumption, SEM case

The parameter of interest was α_1 in equation

$$y_{i1} = y_{i2}\alpha_1 + z_{i1}\beta_1 + \varepsilon_{i1},$$

and we have just shown that

$$y_{i2} = z_{i1}\frac{\beta_1\alpha_2}{1-\alpha_1\alpha_2} + z_{i2}\frac{\beta_2}{1-\alpha_1\alpha_2} + \varepsilon_{i1}\frac{\alpha_2}{1-\alpha_1\alpha_2} + \varepsilon_{i2}\frac{1}{1-\alpha_1\alpha_2}.$$

Does

$$\mathrm{E}\left[y_{i2}\varepsilon_{i1}\right] = 0$$

hold?

Violation of the exogeneity assumption, SEM case

$$y_{i2} = z_{i1}\frac{\beta_1\alpha_2}{1-\alpha_1\alpha_2} + z_{i2}\frac{\beta_2}{1-\alpha_1\alpha_2} + \varepsilon_{i1}\frac{\alpha_2}{1-\alpha_1\alpha_2} + \varepsilon_{i2}\frac{1}{1-\alpha_1\alpha_2}.$$

Multiply both sides with ε_{i1}, take expectations, and use the earlier assumption that $\mathrm{E}\left[z_{i1}\varepsilon_{i1}\right] = 0$ and $\mathrm{E}\left[z_{i2}\varepsilon_{i1}\right] = 0$ to obtain

$$\mathrm{E}\left[y_{i2}\varepsilon_{i1}\right] = \mathrm{E}\left[\varepsilon_{i1}\varepsilon_{i1}\right]\frac{\alpha_2}{1-\alpha_1\alpha_2} + \mathrm{E}\left[\varepsilon_{i2}\varepsilon_{i1}\right]\frac{1}{1-\alpha_1\alpha_2}.$$

If

$$\alpha_2 \neq 0, \mathrm{E}\left[\varepsilon_{i2}\varepsilon_{i1}\right] = 0,$$

or if

$$\alpha_2 = 0, \mathrm{E}\left[\varepsilon_{i2}\varepsilon_{i1}\right] \neq 0,$$

we have

$$\mathrm{E}\left[y_{i2}\varepsilon_{i1}\right] \neq 0,$$

and the OLS estimator of α_1 is subject to the simultaneity bias.

Violation of the exogeneity assumption, SEM case

Why does

$$\alpha_2 \neq 0, \mathrm{E}\left[\varepsilon_{i2}\varepsilon_{i1}\right] = 0$$

cause simultaneity? Considering the SEM,

$$\begin{aligned} y_{i1} &= y_{i2}\alpha_1 + z_{i1}\beta_1 + \varepsilon_{i1}, \\ y_{i2} &= y_{i1}\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2}, \end{aligned}$$

this is obvious.

Violation of the exogeneity assumption, SEM case

Why does

$$\alpha_2 = 0, \mathrm{E}\left[\varepsilon_{i2}\varepsilon_{i1}\right] \neq 0,$$

cause simultaneity? Considering the SEM,

$$\begin{aligned} y_{i1} &= y_{i2}\alpha_1 + z_{i1}\beta_1 + \varepsilon_{i1}, \\ y_{i2} &= y_{i1}\alpha_2 + z_{i2}\beta_2 + \varepsilon_{i2}, \end{aligned}$$

ε_{i2} is a determinant of y_{i2}. ε_{i2} is correlated with ε_{i1}. Hence, y_{i2} enters ε_{i1}, making y_{i2} endogenous in the first equation.

Violation of the exogeneity assumption, estimation methods

When

$$\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] \neq 0$$

the OLS estimator is biased and inconsistent. As we will study, there are alternative estimators that are consistent. The **2SLS** and LIML estimators estimate a single equation, and hence are called single-equation methods. The 3SLS, **GMM**, and FIML estimators jointly estimate an entire system of equations, and hence are called system of equations methods. In this lecture we consider the 2SLS estimator. In the next lecture, we consider the GMM estimator.

IV Model, assumptions

Consider the LRM

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i.$$

$\boldsymbol{x}_i$ is $K \times 1$. Assume that all the assumptions of the LRM hold except that

$$\mathrm{E}\left[\varepsilon_i \boldsymbol{x}_i\right] \neq 0.$$

That is, $\boldsymbol{x}_i$ is endogenous.

IV Model, assumptions

Suppose $\boldsymbol{z}_i$ is a $L \times 1$ vector of instrumental variables. $\boldsymbol{z}_i$ satisfies two main assumptions.

IV Model, assumptions

A1.IV. Relevance. The variables in $\boldsymbol{z}_i$ are sufficiently linearly related to the variables in $\boldsymbol{x}_i$. That is,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$$

has full column rank. Consider the dimensions of the expected value. $\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i'$ is $1 \times K$. Hence,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$$

is $L \times K$. Hence its rank should be K. Hence, the assumption imposes a rank condition. What does a rank condition has to do with $\boldsymbol{z}_i$ being related to $\boldsymbol{x}_i$?

IV Model, assumptions, rank condition example

Consider the LRM

$$y_i = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \varepsilon_i$$

so that

$$\boldsymbol{x}_i' = \begin{bmatrix} 1 & x_{i2} & x_{i3} \end{bmatrix}.$$

Suppose that x_{i2} is exogenous but x_{i3} is endogenous. Suppose we have access to instruments z_{i1}, z_{i2}, z_{i3}. 1 and x_{i2} can also be instruments because they can have explanatory power for x_{i3}. The vector of instruments then takes the form

$$\boldsymbol{z}_i = \begin{bmatrix} 1 \\ x_{i2} \\ z_{i1} \\ z_{i2} \\ z_{i3} \end{bmatrix}.$$

IV Model, assumptions, rank condition example

Then,

$$\boldsymbol{z}_i \boldsymbol{x}_i' = \begin{bmatrix} 1 \\ x_{i2} \\ z_{i1} \\ z_{i2} \\ z_{i3} \end{bmatrix} \begin{bmatrix} 1 & x_{i2} & x_{i3} \end{bmatrix} = \begin{bmatrix} 1 & x_{i2} & x_{i3} \\ x_{i2} & x_{i2}x_{i2} & x_{i2}x_{i3} \\ z_{i1} & z_{i1}x_{i2} & z_{i1}x_{i3} \\ z_{i2} & z_{i2}x_{i2} & z_{i2}x_{i3} \\ z_{i3} & z_{i3}x_{i2} & z_{i3}x_{i3} \end{bmatrix}.$$

Taking the expectation,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \\ \mathrm{E}\left[z_{i2}\right] & \mathrm{E}\left[z_{i2}x_{i2}\right] & \mathrm{E}\left[z_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i3}\right] & \mathrm{E}\left[z_{i3}x_{i2}\right] & \mathrm{E}\left[z_{i3}x_{i3}\right] \end{bmatrix}.$$

IV Model, assumptions, rank condition example

Consider a case where we do not have access to any $\boldsymbol{z}_i$. Then,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \end{bmatrix}.$$

Assume that individual expectations are such that

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The matrix

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$$

cannot have full column rank. That is, rank cannot be K which is 3. Matrix has fewer rows than columns. First and third columns are linearly dependent. Rank condition is not satisfied. $\boldsymbol{\beta}_3$ is under identified. Not surprising: $\mathrm{E}\left[x_{i2}x_{i3}\right] - \mathrm{E}\left[x_{i2}\right]\mathrm{E}\left[x_{i3}\right] = 0$. x_{i2} and x_{i3} are not correlated! x_{i2} cannot be an instrument.

IV Model, assumptions, rank condition example

Consider a case where we have access to only z_{i1} of $\boldsymbol{z}_i$. Then,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

Assume that individual expectations are such that

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

The matrix

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$$

has full column rank if

$$\mathrm{E}\left[z_{i1}x_{i3}\right] \neq 0.$$

That is, if z_{i1} and x_{i3} are correlated! Columns do not add up.
Rank condition is satisfied. β_3 is exactly identified.

IV Model, assumptions, rank condition example

Consider a case where we have access to all $\boldsymbol{z}_i$. Then,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \\ \mathrm{E}\left[z_{i2}\right] & \mathrm{E}\left[z_{i2}x_{i2}\right] & \mathrm{E}\left[z_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i3}\right] & \mathrm{E}\left[z_{i3}x_{i2}\right] & \mathrm{E}\left[z_{i3}x_{i3}\right] \end{bmatrix}.$$

Assume that individual expectations are such that

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & \mathrm{E}\left[z_{i1}x_{i3}\right] \\ 0 & 0 & \mathrm{E}\left[z_{i2}x_{i3}\right] \\ 0 & 0 & \mathrm{E}\left[z_{i3}x_{i3}\right] \end{bmatrix}.$$

The matrix

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$$

has full column rank if one of the expectations $\neq 0$. $\boldsymbol{\beta}_3$ is exactly identified. Or if two or more of them $\neq 0$. $\boldsymbol{\beta}_3$ is overidentified.

IV Model, assumptions, rank condition example

In the examples above, we have assumed values for the individual expectations. However, some of the assumptions we made for certain expectations are not arbitrary but intentional. Now we change one these assumptions, and study the consequences. This exercise provides insights into the implications of the rank condition.

IV Model, assumptions, rank condition example

Consider again the case where we have only z_{i1} of $\boldsymbol{z}_i$. Then,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

Assume that individual expectations are such that

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

Compared to the earlier example, the difference is that 1 was 0. We have full column rank if $\mathrm{E}\left[z_{i1}x_{i3}\right] = 1$. However, this setup is wrong. If $\mathrm{E}\left[z_{i1}x_{i2}\right] \neq 0$ and $\mathrm{E}\left[z_{i1}x_{i3}\right] \neq 0$, then $\mathrm{E}\left[x_{i2}x_{i3}\right] = 0$ cannot be true: x_{i2} and x_{i2} are correlated through z_{i1}. Hence, let us assume that $\mathrm{E}\left[x_{i2}x_{i3}\right] = 1$ in the next example.

IV Model, assumptions, rank condition example

Again, if we have access to only z_{i1} of $\boldsymbol{z}_i$,

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

Assume that individual expectations are such that

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

We wish that $\mathrm{E}\left[z_{i1}x_{i3}\right] = 1$. However, in this case column rank is not 3. Columns add up. But this is surprising because if $\mathrm{E}\left[z_{i1}x_{i3}\right] = 1$, that is if z_{i1} and x_{i3} are correlated, we would expect the rank condition to hold. What is wrong?

IV Model, assumptions, rank condition example

We have

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i3}\right] \\ \mathrm{E}\left[x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i2}\right] & \mathrm{E}\left[x_{i2}x_{i3}\right] \\ \mathrm{E}\left[z_{i1}\right] & \mathrm{E}\left[z_{i1}x_{i2}\right] & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

and

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & \mathrm{E}\left[z_{i1}x_{i3}\right] \end{bmatrix}.$$

If $\mathrm{E}\left[z_{i1}x_{i3}\right] = 1$, then $\mathrm{E}\left[z_{i1}x_{i3}\right] = \mathrm{E}\left[x_{i2}x_{i3}\right]$. This says that z_{i1} is correlated with x_{i3}, but this correlation is the same as the correlation between x_{i2} and x_{i3}. This means that z_{i1} does not bring new information for x_{i3}! z_{i1} cannot be an instrument! z_{i1} should bring new information for x_{i3} that is different than the information x_{i2} brings!

IV Model, assumptions, rank condition example

There is another, perhaps a more explicit way of seeing this if you are willing to consider another assumption we make. Consider the assumption $\mathrm{E}\left[z_{i1}\right] = \mathrm{E}\left[x_{i2}\right]$. It implies that $z_{i1} = x_{i2} + \nu_i$ where $\mathrm{E}\left[\nu_i\right] = 0$. Furthermore, note that $\mathrm{E}\left[z_{i1}x_{i3}\right] = \mathrm{E}\left[x_{i2}x_{i3}\right]$ implies that $\mathrm{E}\left[\left(z_{i1} - x_{i2}\right)x_{i3}\right] = \mathrm{E}\left[\nu_i x_{i3}\right] = 0$. That is, υ_i is not correlated with x_{i3}. This means that z_{i1} does not bring new information for x_{i3} through υ_i. z_{i1} brings information for x_{i3} through x_{i2} because $\mathrm{E}\left[x_{i2}x_{i3}\right] \neq 0$. But we already know that x_{i2} is an instrument for x_{i3}. Hence, z_{i1} does not bring new information for x_{i3}. z_{i1} cannot be an instrument.

IV Model, assumptions

A2.IV. Exogeneity. ε_i is uncorrelated with each variable in $\boldsymbol{z}_i$. That is,

$$\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] = \mathbf{0}.$$

This assumption imposes an orthogonality condition. What does this mean?

IV Model, assumptions

A2.IV. Exogeneity. ε_i is uncorrelated with each variable in $\boldsymbol{z}_i$.
That is, Two vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ are orthogonal to each other if their dot product is zero, that is, if

$$\boldsymbol{m}'\boldsymbol{n} = 0.$$

Two vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ with random components are orthogonal to each other if

$$\mathrm{E}\left[\boldsymbol{m}'\boldsymbol{n}\right] = 0.$$

This means that the random components of $\boldsymbol{m}'\boldsymbol{n}$ may be positive, negative, or zero, but the average of them is 0. If two random vectors are orthogonal, this does not mean that they are independent. It also does not mean that they are uncorrelated. They are uncorrelated if one of the vectors has zero mean.

IV Model, assumptions

A2.IV. Exogeneity. ε_i is uncorrelated with each variable in $\boldsymbol{z}_i$. That is, Hence, the assumption

$$\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] = \mathbf{0}$$

implies an orthogonality condition. There are L such conditions since $\boldsymbol{z}_i$ is $L \times 1$.

IV Model, assumptions

A3.IV. $\boldsymbol{x}_i, \boldsymbol{z}_i, \varepsilon_i$, for $i = 1, \ldots, n$, are an i.i.d. sequence of random variables.

IV Model, assumptions

A4.IV. $\text{Var}\left[\varepsilon_i \boldsymbol{z}_i\right] = \text{E}\left[\varepsilon_i \varepsilon_i \boldsymbol{z}_i \boldsymbol{z}_i'\right]$ is a finite positive definite matrix.

IV Model, assumptions

$\boldsymbol{z}_i$ is $L \times 1$. The instruments in $\boldsymbol{z}_i$ satisfy the rank and exogeneity assumptions. $\boldsymbol{x}_i$ is $K \times 1$. Suppose that $L > K$ so that there are more instruments than there are endogenous variables. In this case we say that the system is overidentified. We can consider a linear combination of the instruments, and estimate β consistently. This is done in two stages and leads to the two-stage least squares estimator: $\boldsymbol{b}_{2SLS}$, which we study now.

IV Model, assumptions

$\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i$ is $K \times 1$. If $L > K$ there are more instruments than there are endogenous variables. That is, we have more information than we need to proxy a given endogenous variable. Should we then just use an arbitrary selection of K instruments, and throw away the remaining $L - K$ instruments? No. Throwing away useful information leads to an inefficient estimator: $\boldsymbol{b}_{IV}$. Linear combinations of the L instruments also satisfy the rank and exogeneity assumptions. This choice leads to an efficient estimator: $\boldsymbol{b}_{2SLS}$. Here efficiency refers to a smaller variance of the estimator used.

IV Model, 2SLS estimator

Consider the LRM

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i.$$

Suppose that $\boldsymbol{x}_i'$ contains two variables which are both endogenous. To derive our estimator, it is enough if one of them is endogenous. We assume that we have two endogenous variables instead of one only to keep the derivation general.

IV Model, 2SLS estimator

Stage one. For each endogenous regressor, estimate by OLS

$$x_{ik} = \boldsymbol{z}_i'\boldsymbol{\pi}_k + \upsilon_{ik}.$$

$\boldsymbol{z}_i'$ contains the instruments. $1 \times L$. $\boldsymbol{\pi}_k$ contains the parameters for $\boldsymbol{z}_i'$. $L \times 1$. Obtaining the prediction $\hat{x}_{ik}$, and generalising to n observations,

$$\hat{\boldsymbol{x}}_k = \boldsymbol{P}_Z \boldsymbol{x}_k = \boldsymbol{Z} \underbrace{\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{x}_k}_{\hat{\boldsymbol{\pi}}_k}.$$

$\hat{\boldsymbol{x}}_k$ contains n predictions. $n \times 1$. $\boldsymbol{Z}$ contains L instruments, each with n observations. $n \times L$. $\hat{\boldsymbol{\pi}}_k$ contains L parameter estimates, for variable k. $L \times 1$. Generalising to K endogenous variables,

$$\hat{\boldsymbol{X}} = \boldsymbol{Z} \underbrace{\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{X}}_{\hat{\boldsymbol{\pi}}}.$$

$\hat{\boldsymbol{\pi}}$ contains L parameter estimates, for K endogenous variables. $L \times K$. $\hat{\boldsymbol{X}}$ is $n \times K$.

IV Model, 2SLS estimator

Stage two. Using the predictions as regressors, estimate by OLS the single equation

$$y_i = \hat{\boldsymbol{x}}'_i \boldsymbol{\beta} + \varepsilon_i^*$$

where

$$\varepsilon_i^* = \hat{\upsilon}'_i \boldsymbol{\beta} + \varepsilon_i.$$

$\hat{\boldsymbol{x}}'_i$ is the vector of predicted endogenous variables, for individual i. It is $1 \times K$. Generalising to n observations, the OLS estimator of this model is

$$\begin{aligned} \boldsymbol{b} &= \left(\hat{\boldsymbol{X}}' \hat{\boldsymbol{X}}\right)^{-1} \hat{\boldsymbol{X}}' \boldsymbol{y} \\ &\equiv \boldsymbol{b}_{2SLS}. \end{aligned}$$

This estimator, obtained in two stages, is the two-stage least squares (2SLS) estimator.

IV Model, 2SLS estimator

In case that you are curious about how we end up with

$$\varepsilon_i^* = \hat{\upsilon}_i' \boldsymbol{\beta} + \varepsilon_i.$$

Considering that there is only one endogenous variable,

$$x_i = z_i \pi + \upsilon_i.$$

Then,

$$x_i = \hat{x}_i + \hat{\upsilon}_i.$$

Replacing x_i in

$$y_i = x_i \beta + \varepsilon_i,$$

we have

$$y_i = \hat{x}_i \beta + \hat{\upsilon}_i \beta + \varepsilon_i$$

and

$$\varepsilon_i^* \equiv \hat{\upsilon}_i \beta + \varepsilon_i.$$

IV Model, 2SLS estimator

Why $\boldsymbol{b}_{2SLS}$ is in fact the OLS estimator in the model considered?
First take note of the following facts.

$$\hat{\boldsymbol{X}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}.$$

$$\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'.$$

$$\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{P}'_{\boldsymbol{Z}}\boldsymbol{P}_{\boldsymbol{Z}}.$$

$$\boldsymbol{P}'_{\boldsymbol{Z}} = \boldsymbol{P}_{\boldsymbol{Z}}.$$

IV Model, 2SLS estimator

$$
\begin{aligned}
\boldsymbol{b}_{2SLS} &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}}'\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^{*}\right)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y}
\end{aligned}
$$

where

$$\boldsymbol{X}^{*} \equiv \boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}.$$

That is, the 2SLS estimator is indeed the OLS estimator: in the first stage $\boldsymbol{X}^*$ is constructed, and in the second stage the OLS estimator is applied on $\boldsymbol{y}$ and $\boldsymbol{X}^*$ (transformed $\boldsymbol{X}$). In the first stage, $\boldsymbol{P}_{\boldsymbol{Z}}$ projects $\boldsymbol{X}$ on to the space spanned by $\boldsymbol{Z}$ which is orthogonal to $\boldsymbol{\varepsilon}$, because

$$\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right] = \mathbf{0}.$$

The first stage has removed the endogeneity problem!

IV Model, 2SLS estimator

$\boldsymbol{b}_{2SLS}$ takes an alternative form. Using $\hat{\boldsymbol{X}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}$,

$$\begin{aligned}
\boldsymbol{b}_{2SLS} &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\
&= \left(\hat{\boldsymbol{X}}'\boldsymbol{X}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y}.
\end{aligned}$$

IV Model, 2SLS estimator

For future reference, note that

$$\begin{aligned}\boldsymbol{b}_{2SLS} &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\ &= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y}.\end{aligned}$$

IV Model, 2SLS estimator

$\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i$ is $K \times 1$. Suppose $L = K$. The number of instruments is equal to the number of endogenous variables. The system is exactly identified. Then,

$$\begin{aligned}
\boldsymbol{b}_{2SLS} &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&\equiv \boldsymbol{b}_{IV}.
\end{aligned}$$

If $L = K$, $\boldsymbol{Z}'\boldsymbol{X}$ is a $K \times K$ square matrix with full rank. Square matrices are nonsingular and invertible if they have full rank. $\boldsymbol{Z}'\boldsymbol{X}$ is invertible. $\boldsymbol{b}_{2SLS} = \boldsymbol{b}_{IV}$. If $L > K$, $\boldsymbol{Z}'\boldsymbol{X}$ is $L \times K$ with rank $K < L$. $\boldsymbol{Z}'\boldsymbol{X}$ is not invertible. $\boldsymbol{b}_{2SLS} \neq \boldsymbol{b}_{IV}$.

IV Model, 2SLS estimator, statistical properties

Small sample properties of $\boldsymbol{b}_{2SLS}$ can not be established analytically. Using simulation analysis it can be shown that $\boldsymbol{b}_{2SLS}$ is in general biased. Hence, we rely on the large sample properties of $\boldsymbol{b}_{2SLS}$. $\boldsymbol{b}_{2SLS}$ is consistent and asymptotically normally distributed. We prove these later.

IV Model, 2SLS estimator, example

```
. reg lwage educ age age2 black

      Source |       SS           df       MS      Number of obs   =     2,220
-------------+----------------------------------   F(4, 2215)      =    143.09
       Model |  88.0908302         4  22.0227076   Prob > F        =    0.0000
    Residual |  340.908673     2,215  .153909108   R-squared       =    0.2053
-------------+----------------------------------   Adj R-squared   =    0.2039
       Total |  428.999503     2,219  .193330105   Root MSE        =    .39231

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0385118   .0032895    11.71   0.000      .032061    .0449627
         age |   .1326507   .0555628     2.39   0.017     .0236901    .2416113
        age2 |  -.0015523   .0009674    -1.60   0.109    -.0034494    .0003448
       black |  -.2127221   .0232691    -9.14   0.000    -.2583537   -.1670906
       _cons |   3.315457   .7883061     4.21   0.000     1.769561    4.861354
------------------------------------------------------------------------------
```

IV Model, 2SLS estimator, example

```
. ivregress 2sls lwage (educ = motheduc fatheduc) age age2 black, first
```

First-stage regressions

Number of obs	=	2,220
F(5, 2214)	=	157.81
Prob > F	=	0.0000
R-squared	=	0.2628
Adj R-squared	=	0.2611
Root MSE	=	2.2244

educ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.9804534	.314502	3.12	0.002	.3637036	1.597203
age2	-.0160649	.0054764	-2.93	0.003	-.0268043	-.0053256
black	-.1607076	.1376706	-1.17	0.243	-.4306846	.1092694
motheduc	.1975247	.0201066	9.82	0.000	.1580948	.2369545
fatheduc	.2230658	.0167964	13.28	0.000	.1901275	.2560042
_cons	-5.389924	4.472077	-1.21	0.228	-14.15983	3.379979

IV Model, 2SLS estimator, example

```
Instrumental variables (2SLS) regression          Number of obs   =      2,220
                                                  Wald chi2(4)    =     503.26
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1900
                                                  Root MSE        =     .39564
```

lwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ	.0600324	.0069201	8.68	0.000	.0464692	.0735955
age	.1094726	.0564143	1.94	0.052	-.0010974	.2200426
age2	-.0011585	.0009819	-1.18	0.238	-.003083	.0007659
black	-.1833938	.0248831	-7.37	0.000	-.2321638	-.1346237
_cons	3.354017	.7950635	4.22	0.000	1.795721	4.912313

```
Instrumented:  educ
Instruments:   age age2 black motheduc fatheduc
```