Violation of the exogeneity assumption, the IV estimator, and the GIV estimator

Econometrics (35B206), Lecture 5

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2019

SLM, error is exogenous

The SLM assumes that ε_i is strictly exogenous, i.e., $\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_k\right] = 0$.

SLM, error is exogenous

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 observation i and also at any other observation j.

SLM, error is exogenous

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}.

SLM, error is exogenous

Generalising

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

to K variables, we consider

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

SLM, error is exogenous

In this lecture we allow

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

That is, we violate weak exogeneity. But we still assume that

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

SLM, error is exogenous, implications

$\mathrm{E}[\varepsilon_i \mid \boldsymbol{x}_i] = \mathbf{0}$ has a number of implications.

SLM, error is exogenous, implication one

First,

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

implies that

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

by the LIE. Keep in mind that when the latter is ever stated, it is because the former holds.

SLM, error is exogenous, implication one

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

implies that

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

When referring to 'exogeneity', we will use the latter statement instead of the former. There are at least two reasons for doing this. First, we can use the latter when talking about covariance: more on this below. Second, the latter is what we need for showing the consistency of the OLS estimator: see the earlier lecture on this.

SLM, error is exogenous, implication, two

Second,

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

implies that

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

by the LIE. 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.

SLM, error is exogenous, implication three

Third,

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

implies that

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

using the above results. That is, ε_i are $\boldsymbol{x}_i$ are uncorrelated.

SLM, error is exogenous, meaning of mean independence

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

implies

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

This does not mean that the left hand sides of the two terms are equal to each other per se. But now assume that

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

This equality tells that the average of ε_i at all slices of the population defined by the different values of $\boldsymbol{x}_i$ is the same as the average of ε_i. That is, values of $\boldsymbol{x}_i$ have no influence on the average value of ε_i. Then, we say that ε_i is mean independent of $\boldsymbol{x}_i$.

SLM, error is exogenous, meaning of mean independence

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 mean independence. Let us clarify the position of mean independence in-between independence and uncorrelatedness.

SLM, error is exogenous, meaning of mean independence

For any function of $\boldsymbol{x}_i$ and ε_i,

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

if ε_i and $\boldsymbol{x}_i$ are independent, since this ensures that

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

It says that all unconditional moments of ε_i are equal to the all 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 first equality for the general function of ε_i does not hold. Mean independence is weaker than independence!

SLM, error is exogenous, meaning of mean independence

For $\boldsymbol{x}_i$ and ε_i,

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

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

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

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

SLM, error is exogenous, meaning of mean independence

For any function of $\boldsymbol{x}_i$, and for ε_i,

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

if ε_i is mean independent of $\boldsymbol{x}_i$, since this ensures 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 first equality still holds. If ε_i and $\boldsymbol{x}_i$ are uncorrelated, the first equality does not hold. Mean independence is in-between independence and uncorrelatedness!

SLM, error is endogenous

Violate

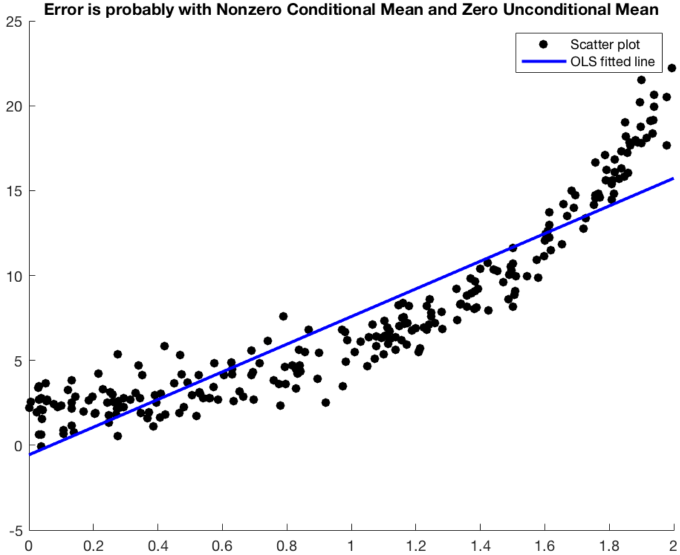
$$\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = 0$$

so that

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

which makes $\boldsymbol{x}_i$ endogenous. When does this happen?

SLM, error is endogenous, case of model misspecification



SLM, error is endogenous, case of model misspecification

The fitted line is based on the standard linear model. The vertical difference between an observation and the fitted line is a residual. The overall mean of the residuals is 0. This is true by construction as long as the regression includes a constant. But for specific ranges of x_i the mean is not 0. So, given the sample data, in the population, is

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

likely to hold? Yes. Is

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

likely to hold? No.

SLM, error is endogenous, case of model misspecification

From the plot, which is based on sample data, we can infer that a linear model is not a good approximation of the true model. We cannot defend the zero conditional mean assumption. In the plot, the sample data for y_i is in fact simulated using the true data generating process

$$y_i = 1 + e^{1.5x_i} + \varepsilon_i$$

where x_i and ε_i take random values from given distributions. This true model is not observed by the researcher. If this was the model we have been using to explain the data, we could defend the zero conditional mean assumption, and hence the conditional expectation function

$$\mathrm{E}\left[y_i \mid x_i\right] = 1 + e^{1.5x_i}.$$

SLM, error is endogenous, case of OVB

Consider the linear model

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

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.

SLM, error is endogenous, case of OVB

Suppose that we do not observe x_{i2} so that it enters the error. 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 $\mathrm{E}\left[x_{i2} \mid x_{i1}\right] \neq 0$. $\beta_2 \neq 0$ means that x_{i2} should enter the model. $\mathrm{E}\left[x_{i2} \mid x_{i1}\right] \neq 0$ means that x_{i1} and x_{i2} are correlated. The zero conditional mean assumption is violated for ε_i^*.

SLM, error is endogenous, case of OVB

What is the implication of

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

for the OLS estimator $\hat{\beta}_1$? The formula for $\hat{\beta}_1$ when x_{i2} is omitted in the true model, while it should not have been, is

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

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

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

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

SLM, error is endogenous, case of OVB

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

In two cases the estimator 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. Realise that the stated expression is the OLS estimate of the coefficient 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 the estimator is subject to the omitted variable bias. The equation stated above is the omitted variable bias formula.

SLM, error is endogenous, case of OVB, emp. example

Regress *wage* on *educ* but ignore *exper* because it is, say, unobserved:

```
. regress wage educ
```

Source	SS	df	MS
Model	7842.35455	1	7842.35455
Residual	31031.0745	995	31.1870095
Total	38873.429	996	39.0295472

Number of obs	=	997
F(1, 995)	=	251.46
Prob > F	=	0.0000
R-squared	=	0.2017
Adj R-squared	=	0.2009
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

SLM, error is endogenous, case of OVB, emp. example

Regress *wage* on *educ* and *exper*, and observe that $\hat{\beta}_{educ}$ increases. This suggests that $\hat{\beta}_{educ}$ has downward bias when *exper* is ignored in the previous regression. How do we reach this conclusion?

```
. 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

SLM, error is endogenous, case of OVB, emp. example

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

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

We would expect

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

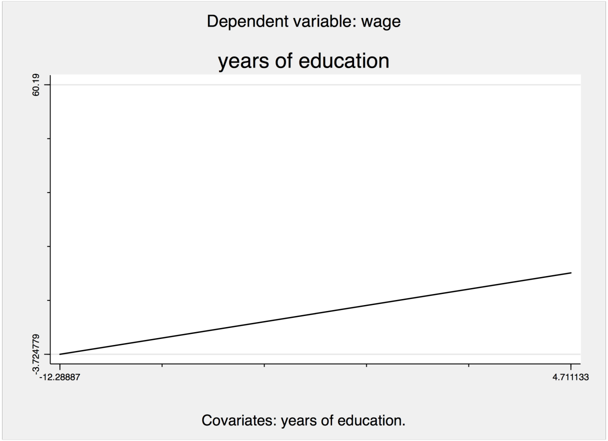
to be negative (effect of exper on educ), and

$$\beta_{exper}$$

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

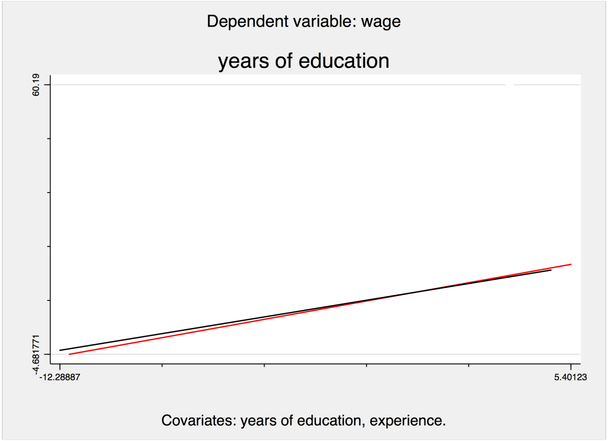
SLM, error is endogenous, case of OVB, emp. example

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



SLM, error is endogenous, case of OVB, emp. example

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



SLM, error is endogenous, case of OVB, sim. example

Suppose that we do not observe x_{i2} so that it enters the error. The model becomes

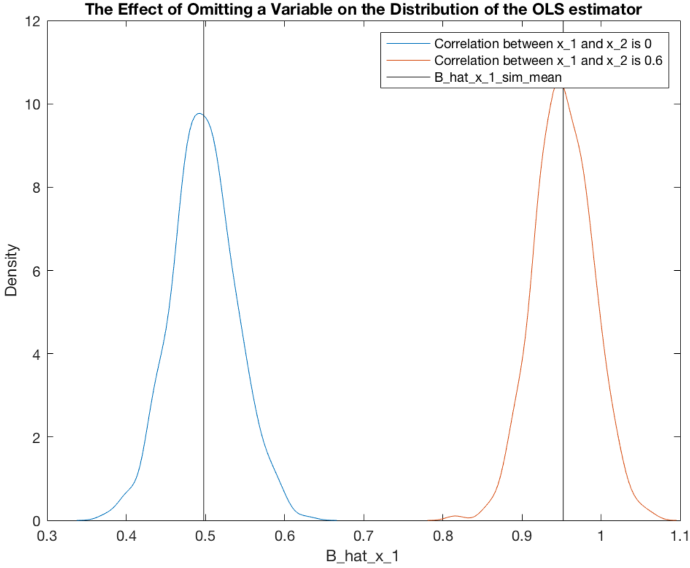
$$y_i = x_{i1}\beta_1 + \varepsilon_i^*$$

where

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

Assume that the true value of β_1 is 0.5. Consider two cases. In the first case, the correlation between the two regressors is 0. In the second case, it is 0.6. Using Monte Carlo simulation, let's check the sampling distribution of $\hat{\beta}_1$ in these two cases.

SLM, error is endogenous, case of OVB, sim. example



SLM, error is endogenous, case of ME

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.

SLM, error is endogenous, case of SEM

Consider the simultaneous equations model given by

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

In each equation the constant is ignored for simplicity. Assume that

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

and that

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

Hence, z_{i1} and z_{i2} are uncorrelated with ε_{i1} and ε_{i2}. Suppose that the interest lies in estimating α_1 in the first equation.

SLM, error is endogenous, case of SEM

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

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

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

SLM, error is endogenous, case of SEM

The parameter of interest was α_1 in the equation

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

and we have just shown that

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

Remember that we need

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

to hold to consistently estimate α_1! Does it hold?

SLM, error is endogenous, case of SEM

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

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

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

If

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

or

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

we have

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

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

SLM, error is endogenous, what to do?

When

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

the OLS estimator is biased and inconsistent. We need a new estimator that has at least the desirable large sample properties. For example, a consistent but biased estimator is already better than the OLS estimator.

SLM, error is endogenous, what to do?

There are in fact different estimators that are consistent. The **IV** 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 study the GIV estimator. Later we cill study the GMM estimator.

IV Model

Consider the linear model

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

where $\boldsymbol{x}_i$ is $K \times 1$. Suppose that

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

IV Model, assumptions, linearity

A1.IV. Linearity. The model is linear in the parameters.

IV Model, assumptions, linearity

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

IV Model, assumptions, rank condition

A2.IV. Relevance. That is,

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

has full column rank. $\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i'$ is $1 \times K$. $\boldsymbol{z}_i \boldsymbol{x}_i'$ is $L \times K$. Hence, the rank of $\boldsymbol{z}_i \boldsymbol{x}_i'$ should be K. Hence, the assumption imposes a rank condition. This condition implies that the variables in $\boldsymbol{z}_i$ are sufficiently linearly related to the variables in $\boldsymbol{x}_i$. 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 linear model

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

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}. Then, the vector of instruments takes the form

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

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

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

does not have full column rank. Rank is not K which is 3. Matrix has fewer rows than columns. First and third columns are linearly dependent. Rank condition is not satisfied. Should we be surprised? $\mathrm{Cov}\left[x_{i2}x_{i3}\right] = \mathrm{E}\left[x_{i2}x_{i3}\right] - \mathrm{E}\left[x_{i2}\right]\mathrm{E}\left[x_{i3}\right] = 0$. x_{i2} and x_{i3} are not correlated! x_{i2} cannot be an instrument. β_3 is under identified.

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! First and third columns are not the same. Rank condition is satisfied. $\boldsymbol{\beta}_3$ is exactly identified.

IV Model, assumptions, rank condition, example

Consider a case where we have access to $\boldsymbol{z}_i$. 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 exp. $\neq 0$. $\boldsymbol{\beta}_3$ is exactly ide. Or if two or more of them $\neq 0$. $\boldsymbol{\beta}_3$ is overid. Rank condition is satisfied.

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 will provide further insights to 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 $\mathrm{E}\left[x_{i2}x_{i3}\right] \neq 0$. That is, x_{i2} and x_{i2} must be correlated through z_{i1}. But $\mathrm{E}\left[x_{i2}x_{i3}\right] = 0$. Hence, let's 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 $\boldsymbol{z}_i$,

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

Assume that individual expectations are such that

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

We wish that $\mathrm{E}[z_{i1}x_{i3}] = 1$. However, in this case column rank is not 3. The first and the second columns add up to the third. But this is surprising because if $\mathrm{E}[z_{i1}x_{i3}] = 1$, 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} and x_{i3} are correlated, 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 $\mathrm{E}[z_{i1}] = \mathrm{E}[x_{i2}]$. It implies that $z_{i1} = x_{i2} + \nu_i$ where $\mathrm{E}[\nu_i] = 0$. Furthermore, note that $\mathrm{E}[z_{i1} x_{i3}] = \mathrm{E}[x_{i2} x_{i3}]$ implies that $\mathrm{E}[(z_{i1} - x_{i2}) x_{i3}] = \mathrm{E}[\nu_i x_{i3}] = 0$. 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 $\mathrm{E}[x_{i2} x_{i3}] \neq 0$. 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, orthogonality condition

A3.IV. Exogeneity. ε_i is uncorrelated with each variable in $\boldsymbol{z}_i$.

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

The assumption imposes an orthogonality condition. There are L such conditions since $\boldsymbol{z}_i$ is $L \times 1$. What does this mean?

IV Model, assumptions, orthogonality condition

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.

IV Model, assumptions, orthogonality condition

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, spherical errors

A4.IV. Errors are homoskedastic and non-autocorrelated. That is,

$$\mathrm{Var}\left[\varepsilon_i \mid \boldsymbol{z}_i\right] = \sigma^2, \; \forall \, i.$$

and

$$\mathrm{Cov}\left[\varepsilon_i, \varepsilon_j \mid \boldsymbol{z}_i\right] = 0, \; \forall \, i \neq j.$$

In the lecture on GMM, we will relax this assumption.

IV Model, assumptions, random sampling

A5.IV. Random sampling. $(\boldsymbol{x}_i, \boldsymbol{z}_i, \varepsilon_i)$, $i = 1, \ldots, n$ are an i.i.d. sequence of random variables.

IV estimator

$\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i$ is $K \times 1$. Suppose that $L = K$. Hence, there are as many instruments as there are endogenous variables. This leads to the $\hat{\boldsymbol{\beta}}_{IV}$ estimator.

IV estimator

$L = K$ has an implication for A2.IV. Since $L = K$, $\boldsymbol{z}_i \boldsymbol{x}_i'$ is a square matrix. This matrix has full rank as A2.IV requires. Square matrices with full rank are invertible. Hence, the inverse of $\mathrm{E}[\boldsymbol{z}_i \boldsymbol{x}_i']$ exists. Or, the inverse of $\boldsymbol{Z}'\boldsymbol{X}$ exists. More on this later.

IV estimator

$$
\begin{aligned}
y_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \\
\boldsymbol{z}_i y_i &= \boldsymbol{z}_i \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i \varepsilon_i \\
\mathrm{E}\left[\boldsymbol{z}_i y_i\right] &= \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\boldsymbol{\beta}\right] + \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] \\
\mathrm{E}\left[\boldsymbol{z}_i y_i\right] &= \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\boldsymbol{\beta} \\
\left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i y_i\right] &= \left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\boldsymbol{\beta} \\
\left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i y_i\right] &= \boldsymbol{\beta}
\end{aligned}
$$

W used two assumptions. First, we used A3.IV so that $\mathrm{E}\left[\boldsymbol{z}_i \varepsilon\right] = 0$. Second, we used A2.IV so that the inverse of $\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$ exists.

IV estimator

$$
\begin{aligned}
\boldsymbol{\beta} &= \left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1} \mathrm{E}\left[\boldsymbol{z}_i y_i\right] \\
&= \left(\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i y_i .
\end{aligned}
$$

Expected value terms are unobserved. We can estimate them using sample data, which gives the IV estimator:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{IV} &= \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i y_i \\
&= \left(\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \sum_{i=1}^{n} \boldsymbol{z}_i y_i \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1} \boldsymbol{Z}'\boldsymbol{y} .
\end{aligned}
$$

IV estimator, motivation

What motivates the estimator is that $\mathrm{E}\left[\boldsymbol{z}_i \varepsilon\right] = 0$ allows us to solve for $\boldsymbol{\beta}$. We obtain K equations in K unknowns in the expression for β. Otherwise we cannot solve for $\boldsymbol{\beta}$, and construct an estimator based on it. See Greene, page 267, for a full treatment of this motivation. We will discuss additional motivation later in this lecture.

IV estimator, finite sample properties

It can be shown that $\hat{\beta}_{IV}$ is biased. Therefore we need to rely on the large sample properties of this estimator and hope that they are satisfactory.

IV estimator, large sample properties, consistency

$\hat{\boldsymbol{\beta}}_{IV}$ is consistent if A1, A2, A3, and A5 hold. Prove this!

IV estimator, large sample properties, asy. normality

We can express the estimator as

$$\hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{Z}'\varepsilon,$$

and then as

$$\hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i.$$

IV estimator, large sample properties, asy. normality

Add $\frac{1}{n}$ in the two sum terms, and multiply both sides of the equation with $\sqrt{n}$ to obtain

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \varepsilon_i.$$

IV estimator, large sample properties, asy. normality

Assuming that $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are i.i.d. (A5), so that we can use the WLLN (Greene, Theorem D.5), and assuming that $\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$ has full column rank (A2),

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1}.$$

Convergence in probability implies convergence in distribution. Hence,

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1} \xrightarrow{d} \left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1}.$$

IV estimator, large sample properties, asy. normality

Assuming that $\boldsymbol{z}_i$ and ε_i are i.i.d. (A5), so that we can use the CLT (Greene, Theorem D.19A), assuming that $\mathrm{E}[\boldsymbol{z}_i\varepsilon_i] = 0$ (A3), assuming that the errors are homoskedastic (A4), and using the LIE,

$$
\begin{aligned}
\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i &\xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\left(\varepsilon_i\boldsymbol{z}_i\right)'\right]\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\varepsilon_i^2\boldsymbol{z}_i\boldsymbol{z}_i'\right]\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\mathrm{E}\left[\varepsilon_i^2\boldsymbol{z}_i\boldsymbol{z}_i' \mid \boldsymbol{z}_i\right]\right]\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\mathrm{E}\left[\varepsilon_i^2 \mid \boldsymbol{z}_i\right]\right]\right] \\
&\xrightarrow{d} N\left[\mathbf{0}, \sigma^2\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\right]
\end{aligned}
$$

We did not assume that ε_i is normal. We are enjoying the CLT.

IV estimator, large sample properties, asy. normality

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}\right)=\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\right)^{-1}}_{\xrightarrow{d}\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}}\underbrace{\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i}_{\xrightarrow{d}N\left[\mathbf{0},\sigma^2\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\right]}.$$

Using the product rule of limiting distributions (Greene, Theorem D.16),

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}\right)\xrightarrow{d}\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}N\left[\mathbf{0},\sigma^2\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\right].$$

IV estimator, large sample properties, asy. normality

$$\begin{aligned}
\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}\right) &\xrightarrow{d} \left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1} N\left[\boldsymbol{0},\sigma^2\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\right] \\
&\xrightarrow{d} N\left[\boldsymbol{0},\sigma^2\left(\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\right)\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\left(\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\right)'\right] \\
&\xrightarrow{d} N\left[\boldsymbol{0},\sigma^2\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\right)^{-1}\right]
\end{aligned}$$

using the property that the transpose and inverse operations commute from the second to the third line.

IV estimator, large sample properties, asy. normality

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\right)^{-1}\right]$$

Assuming that this limiting distribution holds approximately for finite n,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}\right) \xrightarrow{a} N\left[\mathbf{0}, \sigma^2\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\right)^{-1}\right],$$

which leads to

$$\hat{\boldsymbol{\beta}}_{IV} \overset{a}{\sim} N\left[\boldsymbol{\beta}, \sigma^2\frac{1}{n}\left(\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right]\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\right)^{-1}\right].$$

IV estimator, large sample properties, asy. normality

$$\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}_{IV}\right] = \sigma^2 \frac{1}{n} \left(\text{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1} \text{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right] \left(\text{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right]\right)^{-1}.$$

σ^2 and the expected value terms are unobserved. We need to estimate them.

IV estimator, large sample properties, asy. normality

We can estimate

$$\sigma^2$$

with

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_{IV} \right)^2 .$$

IV estimator, large sample properties, asy. normality

We can estimate

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] = \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'$$

with

$$\frac{1}{n} \boldsymbol{Z}' \boldsymbol{X}.$$

We can estimate

$$\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right] = \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i'$$

with

$$\frac{1}{n} \boldsymbol{Z}' \boldsymbol{Z}.$$

IV estimator, large sample properties, asy. normality

$$\begin{aligned}\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{IV}\right] &= \hat{\sigma}^2 \frac{1}{n}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\\ &= \hat{\sigma}^2\left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}.\end{aligned}$$

GIV estimator

$\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i$ is $K \times 1$. Suppose that $L > K$. Hence, 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: $\hat{\boldsymbol{\beta}}_{IV}$. Linear combinations of the L instruments also satisfy the rank and exogeneity assumptions. This leads to an estimator at least as efficient as the $\hat{\boldsymbol{\beta}}_{IV}$ estimator: $\hat{\boldsymbol{\beta}}_{GIV}$.

GIV estimator

$L > K$ has an implication for A2.IV. Since $L > K$, $\boldsymbol{z}_i \boldsymbol{x}_i'$ is $L \times K$. It is not a square matrix. However, it has full column rank which is K as A2.IV requires. But the inverse of $\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]$ does not exist. Or, the inverse of $\boldsymbol{Z}'\boldsymbol{X}$ does not exist. More on this later.

GIV estimator

Consider the linear model

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

We consider that $\boldsymbol{x}'_i$ contains two endogenous variables, instead of only one, to keep the derivation of the $\hat{\boldsymbol{\beta}}_{GIV}$ estimator general.

GIV estimator

The $\hat{\boldsymbol{\beta}}_{GIV}$ estimator is derived, and used, in two stages.

GIV 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 $n \times K$.

GIV 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

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

$\hat{\boldsymbol{x}}'_i$ 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

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

GIV estimator

The estimator is obtained in two stages. Therefore textbooks often call it the two-stage least squares estimator denoted as TSLS.

GIV estimator

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

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

GIV estimator

Why $\hat{\boldsymbol{\beta}}_{GIV}$ is in fact the OLS estimator in the model considered?

GIV estimator

First take note of the following facts.

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

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

$$\hat{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}.$$

GIV estimator

$$\hat{\beta}_{GIV} = \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y}$$
$$= \left((\boldsymbol{P_Z X})'\boldsymbol{P_Z X}\right)^{-1}(\boldsymbol{P_Z X})'\boldsymbol{y}$$

The GIV estimator is the OLS estimator on $\boldsymbol{y}$ and transformed $\boldsymbol{X}$! In the first stage, $\hat{\boldsymbol{X}}$ is constructed. In the second stage the OLS estimator is applied on $\boldsymbol{y}$ and $\hat{\boldsymbol{X}}$.

GIV estimator

In the first stage $\hat{\boldsymbol{X}}$ is constructed. What is happening here?

$$\hat{\boldsymbol{X}} = \boldsymbol{P_Z X}.$$

$\boldsymbol{P_Z}$ projects $\boldsymbol{X}$ on to the space spanned by $\boldsymbol{Z}$. Remember that

$$\boldsymbol{Z} \perp \boldsymbol{\varepsilon}$$

because

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

The first stage removes the endogeneity problem by replacing $\boldsymbol{X}$ by its linear projection on the space spanned by the instruments $\boldsymbol{Z}$, which are, by construction, orthogonal to the error term.

GIV estimator, small sample properties

$\hat{\boldsymbol{\beta}}_{GIV}$ is biased in a finite sample, like the $\hat{\boldsymbol{\beta}}_{IV}$. Therefore, we rely on the asymptotic properties of the estimator.

GIV estimator, large sample properties, consistency

$\hat{\boldsymbol{\beta}}_{GIV}$ is consistent. The proof is very similar to that of the $\hat{\boldsymbol{\beta}}_{IV}$.

GIV estimator, large sample properties, asy. efficiency

Asymptotic variance of $\hat{\boldsymbol{\beta}}_{GIV}$ is equal to or smaller than that of $\hat{\boldsymbol{\beta}}_{IV}$. That is, $\hat{\boldsymbol{\beta}}_{GIV}$ is at least as efficient as the $\hat{\boldsymbol{\beta}}_{IV}$. We do not prove this.

GIV estimator, large sample properties, asy. normality

Derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}_{GIV}$ is very similar to that of $\hat{\boldsymbol{\beta}}_{IV}$.

GIV estimator, large sample properties, asy. normality

$$\hat{\boldsymbol{\beta}}_{GIV} \overset{a}{\sim} N\left[\boldsymbol{\beta}, \sigma^2 \frac{1}{n}\left[\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right]\left(\mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right]\right)^{-1} \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right]^{-1}\right].$$

GIV estimator, large sample properties, asy. normality

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GIV}\right] = \hat{\sigma}^2 \left[\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}.$$

GIV estimator, note one

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

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

GIV estimator, note two

For future reference, note that

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

GIV estimator, note three

Est. Asy. Var $\left[\hat{\boldsymbol{\beta}}_{GIV}\right]$ takes an alternative form. Using $\hat{\boldsymbol{X}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}$,

$$
\begin{aligned}
\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GIV}\right] &= \hat{\sigma}^2\left[\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1} \\
&= \hat{\sigma}^2\left[\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1} \\
&= \hat{\sigma}^2\left[\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\right)'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1} \\
&= \hat{\sigma}^2\left[\left(\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1} \\
&= \hat{\sigma}^2\left[\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right]^{-1}
\end{aligned}
$$

Does this look familiar?

GIV estimator, note four

$\boldsymbol{z}_i$ is the $L \times 1$ vector of instruments. $\boldsymbol{x}_i$ is the $K \times 1$ vector of regressors.

GIV estimator, note four

Suppose $L = K$. The number of instruments is equal to the number of endogenous variables. $\boldsymbol{Z}'\boldsymbol{X}$ is a $K \times K$ square matrix. It has full rank. Square matrices are nonsingular and invertible if they have full rank. Hence, $\boldsymbol{Z}'\boldsymbol{X}$ is invertible.

GIV estimator, note four

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

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

GIV estimator, note four

Suppose $L > K$. The number of instruments is larger than the number of endogenous variables. $\boldsymbol{Z}'\boldsymbol{X}$ is $L \times K$ with rank $K < L$. $\boldsymbol{Z}'\boldsymbol{X}$ is not invertible. Then, $\hat{\boldsymbol{\beta}}_{GIV} \neq \hat{\boldsymbol{\beta}}_{IV}$.

GIV estimator, example

```
. reg lwage educ age age2 black
```

Source	SS	df	MS
Model	88.0908302	4	22.0227076
Residual	340.908673	2,215	.153909108
Total	428.999503	2,219	.193330105

Number of obs	=	2,220
F(4, 2215)	=	143.09
Prob > F	=	0.0000
R-squared	=	0.2053
Adj R-squared	=	0.2039
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

GIV 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

GIV 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
```