Advanced Econometrics I

Jürgen Meinecke

Lecture 5 of 12

Research School of Economics, Australian National University

Roadmap

Ordinary Least Squares Estimation

Weighted and Generalized Least Squares Estimation

Instrumental Variables Estimation

In the last assignment you have learned that the *generalized least squares estimator* $\hat{\beta}^{\mathrm{GLS}}$ is the minimum variance unbiased estimator in the linear regression model under heteroskedasticity

This is a Gauss Markov theorem for the heteroskedastic case

But the derivation assumed knowledge of $\mathrm{E}(ee'|X)$

In real life you don't have that knowledge, and $\hat{\beta}^{\mathrm{GLS}}$ is practically useless, you cannot calculate it

For that reason we call $\hat{\beta}^{\mathrm{GLS}}$ the *infeasible* GLS estimator

There exists a *feasible* variant, but it isn't used much

Let's first revisit the GLS setup

Error variance was

$$
\begin{aligned}
\mathsf{E}(ee'|X) &= \sigma^2 \cdot \Gamma = \sigma^2 \cdot \operatorname{diag}(\gamma_1, \dots, \gamma_N) \\
&= \operatorname{diag}(\sigma_1^2, \dots, \sigma_N^2) =: \Sigma
\end{aligned}
$$

Define $\tilde{Y} := \Gamma^{-1/2} Y$ and $\tilde{X} := \Gamma^{-1/2} X$

The GLS estimator is motivated as the OLS estimator of $\tilde{Y}$ on $\tilde{X}$:

$$
\begin{aligned}
\hat{\beta}^{\mathsf{GLS}} &:= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \\
&= (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}Y \\
&= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \\
&= \left(\sum_{i=1}^{N} X_i X_i'/\sigma_i^2\right)^{-1} \sum_{i=1}^{N} X_i Y_i/\sigma_i^2
\end{aligned}
$$

Each observation is weighed inversely to its error variance

Hence the alternative name *weighted least squares* estimator

How could we turn *infeasible* GLS into a *feasible* estimator?

Idea: use $\hat{\Sigma} = \Sigma + o_p(1)$ in place of Σ

Where does this consistent variance estimator come from?

Easy: OLS will provide a consistent (yet inefficient) estimator of β and therefore also of Σ

This suggests the following two step estimation approach:

(i) run OLS of Y on X, compute $\hat{e}$ and obtain $\hat{\Sigma}$ by imposing some structure on $\mathrm{E}(e_i^2|X_i)$, for example $\mathrm{E}(e_i^2|X_i) = \sigma(X_i)$ where σ is some known function

(ii) compute $\hat{\beta}^{\text{GLS}}_{\text{feas}} := (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y$

Nobody uses this in practice, it's a textbook-only estimator

Feasible GLS does not satisfy the Gauss Markov theorem (because using $\hat{\Sigma}$ instead of Σ adds sampling error)

Roadmap

We have discussed two types of linear models:

linear projection:	$Y_i = X_i'\beta^* + u_i$	$\mathrm{E}(u_i X_i) = 0$
linear regression:	$Y_i = X_i'\beta + e_i$	$\mathrm{E}(e_i \mid X_i) = 0$

If we are only interested in β^*, then OLS is best

The linear regression model can also be estimated via OLS because $\mathrm{E}(e_i|X_i) = 0$ implies $\mathrm{E}(e_i X_i) = 0$

But it is more restrictive: it says that the $\mathrm{E}(Y_i|X_i)$ is linear

It defines a structural model that is very simplistic:
it basically states that the projection *is* the structural relation of interest

In practice, we are often dealing with this model instead:

$$Y_i = X_i'\beta + e_i \qquad \mathrm{E}(e_i X_i) \neq 0$$

It is often called **structural model** to emphasize that β is the coefficient of interest

Correspondingly, β is called **structural parameter**

I find this terminology unfortunate, because there is nothing inherently *structural* about these "models"

They are simple regression equations with the complication that β should not be estimated via OLS
(because $\mathrm{E}(e_i X_i) \neq 0$, and so β isn't the projection coefficient)

When $\mathrm{E}(e_i X_i) \neq 0$ we say that X_i is **endogenous**

Clearly, $\mathrm{E}(e_i | X_i) \neq 0$

The three textbook examples of endogeneity are

- measurement error
- simultaneity, simultaneous equations, and
- omitted variable bias

Let’s have a look

Measurement Error

Let's say the "true" model is

$$Y_i = X_i'\beta + e_i, \qquad \mathsf{E}(e_i|X_i) = 0$$

If you had data on (X_i, Y_i) then OLS would be best

But let's say you only observe $(\tilde{X}_i, Y_i)$ with $\tilde{X}_i = X_i + r_i$ where r_i is a measurement error statistically independent of e_i and X_i

Despite its randomness, this error causes serious problems:

$$Y_i = \tilde{X}_i'\beta + v_i, \qquad \text{where } v_i := e_i - r_i'\beta$$

Can you safely use OLS here? Assuming $\mathsf{E}r_i = 0$,

$$\mathsf{E}(\tilde{X}_i v_i) = \mathsf{E}\big((X_i + r_i)(e_i - r_i'\beta)\big) = -\mathsf{E}(r_i r_i')\beta \neq 0$$

No you cannot! (unless $\beta = 0$ or $\mathsf{E}(r_i r_i') = 0$)

Simultaneity, Simultaneous Equations

Consider the following two equation model

$$Y_{i1} = X_{i1}'\beta_1 + \theta_1 Y_{i2} + e_{i1}$$
$$Y_{i2} = X_{i2}'\beta_2 + \theta_2 Y_{i1} + e_{i2}$$

Let X_{i1} and X_{i2} be well behaved in the sense:
$\mathsf{E}(e_{i1}X_{i1}) = \mathsf{E}(e_{i1}X_{i2}) = \mathsf{E}(e_{i2}X_{i1}) = \mathsf{E}(e_{i2}X_{i2}) = 0$

Further assume $\mathsf{E}(e_{i1}e_{i2}) = 0$ to keep things simple

Using Y_{i1} and Y_{i2} as regressors is problematic:

$$\mathsf{E}(e_{i1}Y_{i2}) = \frac{\theta_2}{1-\theta_1\theta_2}\sigma_1^2 \qquad \mathsf{E}(e_{i2}Y_{i1}) = \frac{\theta_1}{1-\theta_1\theta_2}\sigma_2^2$$

where $\sigma_1^2 := \mathsf{Var}\ (e_{i1})$ and $\sigma_2^2 := \mathsf{Var}\ (e_{i1})$

If $\theta_2 = 0$, then the first equation doesn't have an endogeneity problem and OLS is fine (similarly $\theta_1 = 0$ for second equation)

Omitted Variables Bias

A simple model illustrates the main idea

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + u_i$$

where $\mathrm{E}(u_i|X_{i1}) = 0$, $\mathrm{E}(X_{i2}|X_{i1}) \neq 0$ and you don't observe X_{i2}

You have to omit X_{i2} from the regression

Effectively you are facing the model

$$Y_i = X_{i1}\beta_1 + e_i \qquad e_i := X_{i2}\beta_2 + u_i$$

where $\mathrm{E}(e_i|X_{i1}) \neq 0$

Is this a problem?

Only if $\mathrm{E}(e_i X_{i1}) \neq 0$ which may well be the case

Roadmap

Let's say you have got three scalar rvs X_i, Y_i, Z_i and you have the model

$$Y_i = X_i\beta + e_i \qquad \text{(structural equation)}$$
$$X_i = Z_i\pi + v_i, \qquad \text{(first stage regression)},$$

with $\mathrm{E}(v_i Z_i) = 0$ (so first stage is simply a projection)

Your research interest is β

Should you use OLS to estimate it?

Yes if

- $\mathrm{E}(e_i|X_i) = 0$
  (then you don't really need Z_i at all)
- $\mathrm{E}(e_i|X_i) \neq 0$ but $\mathrm{E}(e_i X_i) = 0$
  (case of *omitted variable non-bias*)

In short: you need $\mathrm{E}(e_i X_i) = 0$ for OLS to make sense

What if $\mathrm{E}(e_i X_i) \neq 0$

Then the existence of Z_i will be helpful as long as $\mathrm{E}(e_i Z_i) = 0$

Notice that $\mathrm{E}(e_i Z_i) = 0$ implies that $\mathrm{E}(e_i v_i) \neq 0$, that is, the error terms of both equations must be correlated

How does Z_i help?

Combine the two equations to get

$$
\begin{aligned}
Y_i &= Z_i \pi \beta + (e_i + v_i \beta) \\
&= Z_i \pi \beta + w_i,
\end{aligned}
$$

where $\mathrm{E}(w_i Z_i) = 0$

Therefore you can consistently estimate $\pi\beta$

Of course you can also consistently estimate π

Simple idea: divide the estimator of $\pi\beta$ by the estimator of π

It follows that you can back out a consistent estimator of β

Alternative motivation: estimate β in two stages

(i) Estimate π via OLS in the first stage regression, create $\hat{X}_i = \hat{\pi} Z_i$

(ii) Regress Y on $\hat{X}_i$ using OLS

The estimator from stage (ii) is numerically identical to the one from the procedure explained on the preceding slide

Why should this make sense? Why can you use $\hat{X}_i$ but not X_i in the structural equation?

Intuition: writing $X_i = \hat{X}_i + \hat{v}_i$ we surmise that

- $\hat{X}_i$ captures the variation of X_i that is *exogenous*
- $\hat{v}_i$ captures the variation of X_i that is *endogneous*

This little example provides a lot of the main ideas about IV estimation already

Unfortunately, however, things get considerably more intricate and complicated once the setup is generalized

It is very important to discuss this extensively in the lecture

IV and 2SLS estimation are pervasive in economics

I'm not sure you can publish a paper only based on OLS

People always cry "endogeneity!" and ask for an instrument

Let's properly understand the pros and cons, and provide best practices

Roadmap

Starting point is the following partition of the linear model

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= X_{i1}'\beta_1 + X_{i2}'\beta_2 + e_i \end{aligned}$$

where $\dim\beta_1 = \dim X_{i1} = K_1 \times 1$

$\dim\beta_2 = \dim X_{i2} = K_2 \times 1$ with $K_1 + K_2 = K$

The two types of regressors are characterized by

$\mathsf{E}(e_i X_{i1}) = 0$ exogenous regressors

$\mathsf{E}(e_i X_{i2}) \neq 0$ endogenous regressors

This immediately tells you that $\beta \neq \beta^*$

Should we use OLS to estimate β?

No, we shouldn't use OLS to estimate β

$\hat{\beta}^{\text{OLS}}$ will consistently estimate β^*

But $\beta \neq \beta^*$

We need something new

Enter the instrumental variable:

Definition (Instrumental Variable (IV))

A $L \times 1$ vector Z_i is called an **instrumental variable (IV)** if

(i)	$\mathrm{E}(Z_i e_i) = 0$	instrument exogeneity
(ii)	$\operatorname{rank} \mathrm{E}(Z_i X_i') = K$	instrument relevance

Notice that X_{i1} does satisfy (i) and will *always* be included in Z_i

A necessary condition for (ii) is $L \geq K$
(at least as many equations as unknowns)

Think of Z_i as partitioned like so:

$$Z_i := \begin{pmatrix} Z_{i1} \\ Z_{i2} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ Z_{i2} \end{pmatrix}$$

Let $\dim Z_{i2} = L_2$; it is clear that $\dim Z_{i1} = K_1$

In other words, the first component of Z_i is always X_{i1} and the second component of Z_i are genuinely *new* instrumental variables that were not included in the model in the first place

The existence of Z_{i2} is crucial to be able to estimate β

Depending on the dimension of Z_{i2} we call the system

$$\dim Z_{i2} = \dim X_{i2} \qquad \text{(exactly identified)}$$

$$\dim Z_{i2} > \dim X_{i2} \qquad \text{(over identified)}$$

$$\dim Z_{i2} < \dim X_{i2} \qquad \text{(under identified)}$$

Roadmap

Now we turn our attention to two **reduced form** regressions:

(i) regressing X_i on Z_i

(ii) regressing Y_i on Z_i

Think of the reduced form as an auxiliary regression that you are using merely as a means to an end

You are not typically interested in the reduced form itself, you are only using it as a tool

The reduced form is usually free of any economic meaning

Reduced form for X_i as dependent variable

Consider the *multivariate* regression model

$$X_i = \pi' Z_i + v_i$$

This notation comprises K regressions with each element of X_i as a dependent variable

Notice that $\dim \pi = L \times K$

Let $\pi = \mathsf{E}(Z_i Z_i')^{-1} \mathsf{E}(Z_i X_i')$, implying $\mathsf{E}(Z_i v_i') = 0$
(this means that the π are the projection coefficients)

Reduced form for Y_i as dependent variable

The reduced form for X_i can be plugged into the original regression:

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= (\pi' Z_i + v_i)'\beta + e_i \\ &= Z_i'\lambda + w_i, \end{aligned}$$

with $\lambda := \pi\beta$ and $w_i := v_i'\beta + e_i$

Notice that $\mathrm{E}(Z_i w_i) = \mathrm{E}(Z_i v_i')\beta + \mathrm{E}(Z_i e_i) = 0$

This means that λ is a projection coefficient, that is, $\lambda = \mathrm{E}(Z_i Z_i')^{-1}\mathrm{E}(Z_i Y_i)$

Collecting results: for the two reduced form coefficients we have

$$\lambda = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i Y_i)$$

$$\pi = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i X_i')$$

The rhs expressions are population moments which are uniquely determined by the distribution that generates the observed data

This implies that λ and π are uniquely determined too

They are *identified*

Great! But wait: we're not interested in λ and π

Instead we want to know about β

Identification of β is not so straightforward, recall:

$$\pi\beta = \lambda$$

Our goal: solve for β

Can't simply divide by π

Let's think about the dimensions

- $\dim \pi = L \times K$
- $\dim \beta = K \times 1$
- $\dim \lambda = L \times 1$

So $\pi\beta = \lambda$ is a system of L equations for K unknowns

Linear algebra tells you that there

- are no solutions or infinitely many solutions if $L < K$
- is hope for unique solution only if $L \geq K$

So let's only consider $L \geq K$

Today we'll focus on the case $L = K$

This case is usually called the *exactly identified* case

(Aside: $L > K$ is called the *over-identified* case)

With $L = K$, to solve $\pi\beta = \lambda$ for β
we need rank $\pi = K$ (full rank) to ensure a unique solution

Notice that π is an upper triangular block matrix (see assignment 5)
for which rank $\pi = K_1 +$ rank $\mathrm{E}(Z_{i2}X'_{i2})$

So it only boils down to whether or not rank $\mathrm{E}(Z_{i2}X'_{i2}) = K_2$

The IV relevance condition makes this happen

Recall the IV relevance condition: rank $\mathsf{E}(Z_i X_i') = K$

This condition implies rank $\mathsf{E}(Z_{i2} X_{i2}') = K_2$

The IV relevance condition therefore ensures that π has full rank, so that we can use matrix inversion to solve $\pi\beta = \lambda$:

$$\begin{aligned}\beta &= \pi^{-1}\lambda \\ &= \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Z_i')\mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i Y_i) \\ &= \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Y_i)\end{aligned}$$

(we have used the fact that $(AB)^{-1} = B^{-1}A^{-1}$)

This solution for β motivates the IV estimator

Roadmap

For the case $L = K$, we've just obtained this solution:

$$\beta = \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Y_i)$$

Applying the analogy principle delivers the estimator

Definition (Instrumental Variable Estimator)

$$\hat{\beta}^{\mathsf{IV}} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i X_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i Y_i\right) = (Z'X)^{-1}Z'Y$$

Aside: when there is only one endogenous variable and one instrumental variable, then the IV estimator is simply

$$\hat{\beta}^{\mathsf{IV}} = \frac{s_{XY}}{s_{XZ}}$$

that is, sample covariance between X_i and Y_i over the sample covariance between X_i and Z_i
(we need this in week 9 when we look at the Wald estimator)

Roadmap

In assignment 5 you are asked to show:

Proposition (Consistency of $\hat{\beta}^{\mathrm{IV}}$)

$\hat{\beta}^{IV} = \beta + o_p(1).$

Proposition (Asymptotic Distribution of $\hat{\beta}^{\mathrm{IV}}$)

$\sqrt{N}(\hat{\beta}^{IV} - \beta) \xrightarrow{d} N\left(0, E(Z_i X_i')^{-1} E(e_i^2 Z_i Z_i') E(X_i Z_i')^{-1}\right).$