Advanced Econometrics I

Jürgen Meinecke Lecture 5 of 12

Research School of Economics, Australian National University

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 $\hat{\beta}^{\text{GLS}}$ 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 E(ee'|X)

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

For that reason we call $\hat{\beta}^{\text{GLS}}$ 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

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

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} : $\hat{\beta}^{\text{GLS}} := (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$ $= (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}Y$ $= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$ $= \left(\sum_{i=1}^{N} X_iX'_i/\sigma_i^2\right)^{-1}\sum_{i=1}^{N} X_iY_i/\sigma_i^2$

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 $E(e_i^2|X_i)$, for example $E(e_i^2|X_i) = \sigma(X_i)$ where σ is some known function
- (ii) compute $\hat{\beta}_{\text{feas}}^{\text{GLS}} := (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y$

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)

Ordinary Least Squares Estimation

Instrumental Variables Estimation Motivation Main Idea in a Nutshell General Setup Identification Definition of IV Estimator Large Sample Properties of IV Esti

We have discussed two types of linear models:

linear projection: $Y_i = X'_i \beta^* + u_i$ $E(u_i X_i) = 0$ linear regression: $Y_i = X'_i \beta + e_i$ $E(e_i | X_i) = 0$

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

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

But it is more restrictive: it says that the $E(Y_i|X_i)$ 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: $Y_i = X_i'\beta + e_i \qquad \mathbb{E}(e_iX_i) \neq 0$

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 $E(e_iX_i) \neq 0$, and so β isn't the projection coefficient) When $E(e_iX_i) \neq 0$ we say that X_i is **endogenous**

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

The three textbook examples of endogeneity are

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

Let's have a look

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

Can you safely use OLS here? Assuming $Er_i = 0$, $E(\tilde{X}_i v_i) = E((X_i + r_i)(e_i - r'_i\beta)) = -E(r_i r'_i)\beta \neq 0$

No you cannot! (unless $\beta = 0$ or $E(r_i r'_i) = 0$)

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: $E(e_{i1}X_{i1}) = E(e_{i1}X_{i2}) = E(e_{i2}X_{i1}) = E(e_{i2}X_{i2}) = 0$ Further assume $E(e_{i1}e_{i2}) = 0$ to keep things simple Using Y_{i1} and Y_{i2} as regressors is problematic: $E(e_{i1}Y_{i2}) = E(e_{i1}(X'_{i2}\beta_2 + \theta_2Y_{i1} + e_{i2})) = \theta_2 E(e_{i1}Y_{i1})$ $= \theta_2 \mathbb{E} \left(e_{i1} (X'_{i1} \beta_1 + \theta_1 Y_{i2} + e_{i1}) \right)$ $= \theta_1 \theta_2 \mathbb{E} \left(e_{i1} Y_{i2} \right) + \sigma_1^2$ $=\frac{\theta_2}{1-\theta_1\theta_2}\sigma_1^2\neq 0$

where $\sigma_1^2 := \text{Var}(e_{i1})$

Therefore, in the two equation model

$$\begin{split} Y_{i1} &= X'_{i1}\beta_1 + \theta_1 Y_{i2} + e_{i1} \\ Y_{i2} &= X'_{i2}\beta_2 + \theta_2 Y_{i1} + e_{i2} \end{split}$$

the errors are not uncorrelated with all regressors

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

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)

A simple model illustrates the main idea

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

where $E(u_i|X_{i1}) = 0$, $E(X_{i2}|X_{i1}) \neq 0$ 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 $E(e_i|X_{i1}) \neq 0$

Is this a problem?

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

Ordinary Least Squares Estimation

Instrumental Variables Estimation

Motivation

Main Idea in a Nutshell

General Setup

Identification

Definition of IV Estimator

Large Sample Properties of IV Estimator

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

$$\begin{array}{ll} Y_i = X_i \beta + e_i & (\text{structural equation}) \\ X_i = Z_i \pi + v_i, & \mathsf{E}(v_i Z_i) = 0 & (\text{first stage regression}) \end{array}$$

Notice: first stage is simply a projection

Your research interest is β

Should you use OLS to estimate it? Yes if

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

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

What if $E(e_i X_i) \neq 0$

Then the existence of Z_i will be helpful as long as $E(e_iZ_i) = 0$

Notice that $E(e_iZ_i) = 0$ implies that $E(e_iv_i) \neq 0$, that is, the error terms of both equations must be correlated

How does Z_i help?

Combine the two equations to get

$$Y_i = Z_i \pi \beta + (e_i + v_i \beta)$$
$$= Z_i \pi \beta + w_i,$$

where $E(w_i Z_i) = 0$

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 eta

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 see that

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

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

Ordinary Least Squares Estimation

Instrumental Variables Estimation

Motivation

Main Idea in a Nutshell

General Setup

Identification

Definition of IV Estimator

Large Sample Properties of IV Estimator

Starting point is the following partition of the linear model $Y_i = X'_i\beta + e_i$ $= X'_{i1}\beta_1 + X'_{i2}\beta_2 + e_i$

where dim β_1 = dim $X_{i1} = K_1 \times 1$ dim β_2 = dim $X_{i2} = K_2 \times 1$ with $K_1 + K_2 = K$

The two types of regressors are characterized by $E(e_iX_{i1}) = 0$ (exogenous regressors) $E(e_iX_{i2}) \neq 0$ (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 eta

```
\hat{\beta}^{OLS} will consistently estimate \beta^*, 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) $E(Z_ie_i) = 0$ (ii)rank $E(Z_iX'_i) = K$ instrument relevance

Notice that X_{i1} does satisfy (i) and will *always* be included in Z_i Intuition for (ii): nonzero correlation between X_i and Z_i

A necessary condition for (ii) is $L \ge K$ (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

 $\begin{array}{ll} \mbox{Depending on the dimension of } Z_{i2} \mbox{ we call the system} \\ \mbox{dim } Z_{i2} = \mbox{dim } X_{i2} \mbox{ (exactly identified)} \\ \mbox{dim } Z_{i2} > \mbox{dim } X_{i2} \mbox{ (over identified)} \\ \mbox{dim } Z_{i2} < \mbox{dim } X_{i2} \mbox{ (under identified)} \end{array}$

Ordinary Least Squares Estimation

Instrumental Variables Estimation

Motivation

Main Idea in a Nutshell

General Setup

Identification

Definition of IV Estimator

Large Sample Properties of IV Estimator

Now we turn our attention to two reduced form regressions:

- (i) regressing X_i on Z_i (this is the first stage regression)
- (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 $\pi = E(Z_i Z'_i)^{-1} E(Z_i X'_i)$, implying $E(Z_i v'_i) = 0$ (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:

$$\begin{aligned} & \swarrow_i = X'_i \beta + e_i \\ & = (\pi' Z_i + v_i)' \beta + e_i \\ & = Z'_i \lambda + w_i, \end{aligned}$$

γ

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

Notice that $E(Z_i w_i) = E(Z_i v'_i)\beta + E(Z_i e_i) = 0$

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

Collecting results: for the two reduced form coefficients we have
$$\begin{split} \lambda &= \mathsf{E}(Z_i Z_i')^{-1} \mathsf{E}(Z_i Y_i) \\ \pi &= \mathsf{E}(Z_i Z_i')^{-1} \mathsf{E}(Z_i X_i') \end{split}$$

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

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

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 $L \ge K$

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

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 $\pi = K_1 + \text{rank } \mathbb{E}(Z_{i2}X'_{i2})$

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

The IV relevance condition makes this happen

Recall the IV relevance condition: rank $E(Z_iX'_i) = K$

This condition implies rank $E(Z_{i2}X'_{i2}) = K_2$

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

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

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

This solution for β motivates the IV estimator

Ordinary Least Squares Estimation

Instrumental Variables Estimation

Motivation

Main Idea in a Nutshell

General Setup

Identification

Definition of IV Estimator

Large Sample Properties of IV Estimator

For the case L = K, we've just obtained this solution: $\beta = E(Z_i X'_i)^{-1} E(Z_i Y_i)$

Applying the analogy principle delivers the estimator

Definition (Instrumental Variable Estimator)

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

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

$$\hat{\beta}^{\rm IV} = \frac{s_{ZY}}{s_{XZ}}$$

that is, sample covariance between Z_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)

Ordinary Least Squares Estimation

Instrumental Variables Estimation

Motivation

Main Idea in a Nutshell

General Setup

Identification

Definition of IV Estimator

Large Sample Properties of IV Estimator

In assignment 5 you are asked to show:

Proposition (Consistency of $\hat{\beta}^{|V|}$)

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

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

 $\sqrt{N}(\hat{\beta}^{|V}-\beta) \xrightarrow{d} N\left(0, E(Z_iX_i')^{-1}E(e_i^2Z_iZ_i')E(X_iZ_i')^{-1}\right).$