The outline for Unit 5

UNIT 1. Introduction: The regression model. ✓

UNIT 2. Estimation principles. ✓

UNIT 3: Hypothesis testing principles. ✓

UNIT 4: Heteroscedasticity in the regression model. ✓

UNIT 5: Endogeneity of regressors.

5.1 Errors in variables.
5.2 Simultaneous equation bias.
5.3 Instrumental variables.
5.4 Testing for endogeneity.

Measurement errors

- The dependent variable or the regressors can be measured with error.
- Thinking about the way economic data are reported, measurement error is probably quite prevalent.
- For example, estimates of growth of GDP, inflation, etc. are commonly revised several times. Why should the last revision necessarily be correct?
- Measurement errors in the dependent variable and the regressors have important differences.

Measurement errors - 2

Error of measurement of the dependent variable:

The data generating process is presumed to be

$$\begin{aligned} y^* &= X\beta + \varepsilon \\ y &= y^* + v \\ v_t &\sim iid(0, \sigma_v^2) \end{aligned}$$

where y^* is the unobservable true dependent variable, and y is what is observed.

We assume that ε and v are independent and that $y^* = X\beta + \varepsilon$ satisfies the classical assumptions.

Measurement errors - 3

Given this, we have

$$y + v = X\beta + \varepsilon$$

so

$$\begin{aligned} y &= X\beta + \varepsilon - v \\ &= X\beta + \omega \\ \omega_t &\sim iid(0, \sigma_\varepsilon^2 + \sigma_v^2) \end{aligned}$$

- As long as v is uncorrelated with X, this model satisfies the classical assumptions and can be estimated by OLS. Then, this type of measurement error isn't a problem.

Measurement errors - 4

Error of measurement of the regressors:

The situation isn't so good in this case. The DGP is:

$$\begin{aligned} y_t &= x_t^{*\prime}\beta + \varepsilon_t \\ x_t &= x_t^* + v_t \\ v_t &\sim iid(0, \Sigma_v) \end{aligned}$$

where Σ_v is a $K \times K$ matrix.

Now x^* contains the true, unobserved regressors, and x is what is observed.

Again assume that v is independent of ε, and that the model $y = X^*\beta + \varepsilon$ satisfies the classical assumptions.

Measurement errors - 5

Now we have

$$\begin{aligned} y_t &= (x_t - v_t)' \beta + \varepsilon_t \\ &= x_t'\beta - v_t'\beta + \varepsilon_t \\ &= x_t'\beta + \omega_t \end{aligned}$$

The problem is that now there is a correlation between x_t and ω_t, since

$$\begin{aligned} \mathrm{E}(x_t \omega_t) &= \mathrm{E}\left((x_t^* + v_t)\left(-v_t'\beta + \varepsilon_t\right)\right) \\ &= -\Sigma_v \beta \end{aligned}$$

where

$$\Sigma_v = \mathrm{E}\left(v_t v_t'\right).$$

Measurement errors - 6

Because of this correlation, the OLS estimator is biased and inconsistent.

In matrix notation, write the estimated model as $y = X\beta + \omega$.

We have that

$$\hat{\beta} = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'y}{n}\right)$$

and

$$plim\left(\frac{X'X}{n}\right)^{-1} = plim\frac{(X^{*\prime} + V')(X^* + V)}{n} = (Q_{X^*} + \Sigma_v)^{-1}$$

since X^* and V are independent, and

$$plim\frac{V'V}{n} = \lim \mathrm{E}\frac{1}{n}\sum_{t=1}^{n} v_t v_t'. = \Sigma_v$$

Measurement errors - 7

Likewise,

$$\begin{aligned} plim\left(\frac{X'y}{n}\right) &= plim\frac{\left(X^{*\prime}+V'\right)\left(X^{*}\beta+\varepsilon\right)}{n} \\ &= Q_{X^{*}}\beta \end{aligned}$$

so

$$plim\hat{\beta} = \left(Q_{X^{*}}+\Sigma_{v}\right)^{-1}Q_{X^{*}}\beta$$

So we see that the least squares estimator is $\boxed{\text{inconsistent}}$ when the regressors are measured with error.

- A potential solution to this problem is the instrumental variables (IV) estimator.

Simultaneous equations

- Up until now our model is
$$y = X\beta + \varepsilon$$
where, for purposes of estimation we can treat X as fixed.

- This means that when estimating β we *condition* on X.

- When analyzing dynamic models, we're not interested in conditioning on X, as in the stochastic regressors case.

- Nevertheless, the OLS estimator obtained by treating X as fixed continues to have desirable asymptotic properties even in that case.

Simultaneous equations - Example

An example of a simultaneous equation system is a simple supply-demand system:

$$\begin{aligned} \text{Demand: } q_t &= \alpha_1 + \alpha_2 p_t + \alpha_3 y_t + \varepsilon_{1t} \\ \text{Supply: } q_t &= \beta_1 + \beta_2 p_t + \varepsilon_{2t} \\ \mathrm{E}\left(\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix} \right) &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \\ &\equiv \Sigma, \forall t \end{aligned}$$

The presumption is that q_t and p_t are jointly determined at the same time by the intersection of these equations.

We'll assume that y_t is determined by some unrelated process.

Simultaneous equations - Example - 2

It's easy to see that we have correlation between regressors and errors. Solving for p_t :

$$\begin{aligned}
\alpha_1 + \alpha_2 p_t + \alpha_3 y_t + \varepsilon_{1t} &= \beta_1 + \beta_2 p_t + \varepsilon_{2t} \\
\beta_2 p_t - \alpha_2 p_t &= \alpha_1 - \beta_1 + \alpha_3 y_t + \varepsilon_{1t} - \varepsilon_{2t} \\
p_t &= \frac{\alpha_1 - \beta_1}{\beta_2 - \alpha_2} + \frac{\alpha_3 y_t}{\beta_2 - \alpha_2} + \frac{\varepsilon_{1t} - \varepsilon_{2t}}{\beta_2 - \alpha_2}
\end{aligned}$$

Now consider whether p_t is uncorrelated with ε_{1t} :

$$\mathrm{E}(p_t \varepsilon_{1t}) = \mathrm{E}\left\{ \left(\frac{\alpha_1 - \beta_1}{\beta_2 - \alpha_2} + \frac{\alpha_3 y_t}{\beta_2 - \alpha_2} + \frac{\varepsilon_{1t} - \varepsilon_{2t}}{\beta_2 - \alpha_2} \right) \varepsilon_{1t} \right\} = \frac{\sigma_{11} - \sigma_{12}}{\beta_2 - \alpha_2}$$

Because of this correlation, OLS estimation of the demand equation will be biased and inconsistent. The same applies to the supply equation, for the same reason.

Simultaneous equations - Example - 3

- In this model, q_t and p_t are the *endogenous* varibles (endogs), that are determined within the system.

- y_t is an *exogenous* variable (exogs).

In order to clarify these concepts we need some notation:

▷ Suppose we group together current endogs in the $G \times 1$. vector Y_t.

▷ Group current and lagged exogs, as well as lagged endogs in the $K \times 1$. vector X_t.

▷ Stack the errors of the G equations into the error vector E_t.

Exogeneity

The model, with additional assumptions, can be written as

$$\begin{aligned} Y_t'\Gamma &= X_t'B + E_t' \\ E_t &\sim N(0,\Sigma), \forall t \\ \mathrm{E}(E_t E_s') &= 0, t \neq s \end{aligned}$$

We can stack all n observations and write the model as

$$\begin{aligned} Y\Gamma &= XB + E \\ \mathrm{E}(X'E) &= 0_{(K\times G)} \\ vec(E) &\sim N(0,\Psi) \end{aligned}$$

where Y is $n \times G$, X is $n \times K$, and E is $n \times G$.

Exogeneity - 2

- This system is *complete*, in that there are as many equations as endogs.

- Since there is no autocorrelation of the E_t 's, and since the columns of E are individually homoscedastic, then

$$\Psi = \begin{bmatrix} \sigma_{11}I_n & \sigma_{12}I_n & \cdots & \sigma_{1G}I_n \\ & \sigma_{22}I_n & & \vdots \\ & & \ddots & \vdots \\ & & & \sigma_{GG}I_n \end{bmatrix} = I_n \otimes \Sigma$$

- X may contain lagged endogenous and exogenous variables. These variables are *predetermined.*

- We need to define what is meant by "endogenous" and "exogenous" when classifying the current period variables.

Exogeneity - 3

The model defines a *data generating process*. The model involves two sets of variables, Y_t and X_t, as well as a parameter vector

$$\theta = \left[\begin{array}{ccc} vec(\Gamma)' & vec(B)' & vec^*(\Sigma)' \end{array} \right]'$$

- In general, without additional restrictions, θ is a $G^2+GK+\left(G^2-G\right)/2+G$ dimensional vector. This is the parameter vector that were interested in estimating.

- In principle, there exists a joint density function for Y_t and X_t, which depends on a parameter vector ϕ. Write this density as

$$f_t(Y_t, X_t|\phi, \mathcal{I}_t)$$

  where $\mathcal{I}_t$ is the information set in period t. This includes lagged Y_t's and lagged X_t's.

Exogeneity - 4

- This density can be factored into the density of Y_t conditional on X_t times the marginal density of X_t :

$$f_t(Y_t, X_t|\phi, \mathcal{I}_t) = f_t(Y_t|X_t, \phi, \mathcal{I}_t) f_t(X_t|\phi, \mathcal{I}_t)$$

- If not all parameters in ϕ affect both factors: ϕ_1 indicates elements of ϕ that enter into the conditional density and ϕ_2 for parameters that enter into the marginal. We have

$$f_t(Y_t, X_t|\phi, \mathcal{I}_t) = f_t(Y_t|X_t, \phi_1, \mathcal{I}_t) f_t(X_t|\phi_2, \mathcal{I}_t)$$

- Recall that the model is

$$\begin{aligned} Y_t'\Gamma &= X_t'B + E_t' \\ E_t &\sim N(0, \Sigma), \forall t \\ \mathrm{E}(E_t E_s') &= 0, t \neq s \end{aligned}$$

Exogeneity - 5

Normality and lack of correlation over time imply that the observations are independent, so we can write the log-likelihood function as the sum of likelihood contributions of each observation:

$$\begin{aligned}
\ln L(Y|\theta, \mathcal{I}_t) &= \sum_{t=1}^{n} \ln f_t(Y_t, X_t|\phi, \mathcal{I}_t) \\
&= \sum_{t=1}^{n} \ln\left(f_t(Y_t|X_t, \phi_1, \mathcal{I}_t) f_t(X_t|\phi_2, \mathcal{I}_t)\right) \\
&= \sum_{t=1}^{n} \ln f_t(Y_t|X_t, \phi_1, \mathcal{I}_t) + \sum_{t=1}^{n} \ln f_t(X_t|\phi_2, \mathcal{I}_t)
\end{aligned}$$

Definition: X_t is *weakly exogenous* for θ (the original parameter vector) if there is a mapping from ϕ to θ that is invariant to ϕ_2. More formally, for an arbitrary (ϕ_1, ϕ_2), $\theta(\phi) = \theta(\phi_1)$.

Exogeneity - 6

Supposing that X_t is weakly exogenous, then the MLE of ϕ_1 using the joint density is the same as the MLE using only the conditional density

$$\ln L(Y|X, \theta, \mathcal{I}_t) = \sum_{t=1}^{n} \ln f_t(Y_t|X_t, \phi_1, \mathcal{I}_t)$$

since the conditional likelihood doesn't depend on ϕ_2.

- With weak exogeneity, knowledge of the DGP of X_t is irrelevant for inference on ϕ_1, and knowledge of ϕ_1 is sufficient to recover the parameter of interest, θ. Since the DGP of X_t is irrelevant, we can treat X_t as fixed in inference.

- By the invariance property of MLE, the MLE of θ is $\theta(\hat{\phi}_1)$,and this mapping is assumed to exist in the definition of weak exogeneity.

- Of course, we'll need to figure out just what this mapping is to recover $\hat{\theta}$ from $\hat{\phi}_1$. This is the famous *identification problem.*

Exogeneity - 7

- With lack of weak exogeneity, the joint and conditional likelihood functions maximize in different places. For this reason, we can't treat X_t as fixed in inference. The joint MLE is valid, but the conditional MLE is not.

- In resume, we require the variables in X_t to be weakly exogenous if we are to be able to treat them as fixed in estimation.

- Lagged Y_t satisfy the definition, since they are in the conditioning information set, e.g., $Y_{t-1} \in \mathcal{I}_t$.

- Lagged Y_t aren't exogenous in the normal usage of the word, since their values *are* determined within the model, just earlier on.

- *Weakly exogenous* variables include *exogenous* (in the normal sense) variables as well as all *predetermined* variables.

Instrumental Variables Estimation

Let's consider the general problem of a linear regression model with correlation between regressors and the error term:

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &\sim iid(0, I_n\sigma^2) \\ \mathrm{E}(X'\varepsilon) &\neq 0. \end{aligned}$$

Consider some matrix W which is formed of variables uncorrelated with ε. This matrix defines a projection matrix

$$P_W = W(W'W)^{-1}W'$$

so that anything that is projected onto the space spanned by W will be uncorrelated with ε.

Instrumental Variables Estimation - 2

Transforming the model with this projection matrix we get

$$P_W y = P_W X\beta + P_W \varepsilon$$

or

$$y^* = X^*\beta + \varepsilon^*$$

Now we have that ε^* and X^* are uncorrelated, since this is simply

$$\begin{aligned} \mathrm{E}(X^{*\prime}\varepsilon^*) &= \mathrm{E}(X' P_W' P_W \varepsilon) \\ &= \mathrm{E}(X' P_W \varepsilon) \end{aligned}$$

and

$$P_W X = W(W'W)^{-1}W'X$$

is the fitted value from a regression of X on W. This is a linear combination of the columns of W, so it must be uncorrelated with ε.

Instrumental Variables Estimation - 3

- This implies that applying OLS to the model

$$y^* = X^*\beta + \varepsilon^*$$

will lead to a consistent estimator, given a few more assumptions.

- This is the *generalized instrumental variables estimator*.
- W is known as the matrix of instruments.
- The IV estimator is

$$\hat{\beta}_{IV} = (X'P_W X)^{-1} X'P_W y$$

Instrumental Variables Estimation - 4

$$
\begin{aligned}
\hat{\beta}_{IV} &= (X'P_W X)^{-1} X'P_W (X\beta + \varepsilon) \\
&= \beta + (X'P_W X)^{-1} X'P_W \varepsilon
\end{aligned}
$$

so

$$
\begin{aligned}
\hat{\beta}_{IV} - \beta &= (X'P_W X)^{-1} X'P_W \varepsilon \\
&= \left(X'W(W'W)^{-1}W'X\right)^{-1} X'W(W'W)^{-1}W'\varepsilon
\end{aligned}
$$

Now we can introduce factors of n to get

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

Instrumental Variables Estimation - 5

Assuming that each of the terms with a n in the denominator satisfies a LLN, so that

- $\frac{W'W}{n} \xrightarrow{p} Q_{WW}$, a finite pd matrix
- $\frac{X'W}{n} \xrightarrow{p} Q_{XW}$, a finite matrix with rank K ($=$ cols(X))
- $\frac{W'\varepsilon}{n} \xrightarrow{p} 0$

then the plim of the rhs is zero, since we assume that W and ε are uncorrelated, e.g., $\mathrm{E}(W_t'\varepsilon_t) = 0.$

Given these assumptions the IV estimator is consistent

$$\hat{\beta}_{IV} \xrightarrow{p} \beta.$$

Instrumental Variables Estimation - 6

Furthermore, scaling by $\sqrt{n},$ we have

$$\sqrt{n}\left(\hat{\beta}_{IV}-\beta\right)=\left(\left(\frac{X'W}{n}\right)\left(\frac{W'W}{n}\right)^{-1}\left(\frac{W'X}{n}\right)\right)^{-1}\left(\frac{X'W}{n}\right)\left(\frac{W'W}{n}\right)^{-1}\left(\frac{W'\varepsilon}{\sqrt{n}}\right)$$

Assuming that the far right term satisfies a CLT, so that

- $\frac{W'\varepsilon}{\sqrt{n}} \xrightarrow{d} N(0, Q_{WW}\sigma^2)$

then we get

$$\sqrt{n}\left(\hat{\beta}_{IV}-\beta\right) \xrightarrow{d} N\left(0, (Q_{XW}Q_{WW}^{-1}Q'_{XW})^{-1}\sigma^2\right)$$

We need estimators for Q_{XW}, Q_{WW} and σ^2.

Instrumental Variables Estimation - 7

The estimators for Q_{XW} and Q_{WW} are the obvious ones.

An estimator for σ^2 is

$$\widehat{\sigma^2_{IV}} = \frac{1}{n}\left(y - X\hat{\beta}_{IV}\right)'\left(y - X\hat{\beta}_{IV}\right).$$

This estimator is consistent following the proof of consistency of the OLS estimator of σ^2, when the classical assumptions hold.

The formula used to estimate the variance of $\hat{\beta}_{IV}$ is

$$\hat{V}(\hat{\beta}_{IV}) = \left((X'W)\,(W'W)^{-1}\,(W'X)\right)^{-1}\widehat{\sigma^2_{IV}}$$

Instrumental Variables Estimation - 8

The IV estimator is:

1. Consistent

2. Asymptotically normally distributed

3. Biased in general, since even though $\mathrm{E}(X'P_W\varepsilon) = 0$, $\mathrm{E}(X'P_WX)^{-1}X'P_W\varepsilon$ may not be zero, since $(X'P_WX)^{-1}$ and $X'P_W\varepsilon$ are not independent.

An important point is that the asymptotic distribution of $\hat{\beta}_{IV}$ depends upon Q_{XW} and Q_{WW}, and these depend upon the choice of W. *The choice of instruments influences the efficiency of the estimator*.

2SLS Estimation

When we have no information regarding cross-equation restrictions or the structure of the error covariance matrix, one can estimate the parameters of a single equation of the system without regard to the other equations:

In the first stage, each column of Y_1 is regressed on *all* the weakly exogenous variables in the system. The fitted values are

$$\begin{aligned}\hat{Y}_1 &= X(X'X)^{-1}X'Y_1 \\ &= P_X Y_1 = X\hat{\Pi}_1\end{aligned}$$

The second stage substitutes $\hat{Y}_1$ in place of Y_1, and estimates by OLS.

The original model is

$$y = Y_1\gamma_1 + X_1\beta_1 + \varepsilon = Z\delta + \varepsilon$$

and the second stage model is $y = \hat{Y}_1\gamma_1 + X_1\beta_1 + \varepsilon$.

2SLS Estimation - 2

Since X_1 is in the space spanned by X, $P_X X_1 = X_1$, so we can write the second stage model as

$$y = P_X Y_1 \gamma_1 + P_X X_1 \beta_1 + \varepsilon \equiv P_X Z \delta + \varepsilon$$

The OLS estimator applied to this model is

$$\hat{\delta} = (Z' P_X Z)^{-1} Z' P_X y$$

which is exactly what we get if we estimate using IV, with the reduced form predictions of the endogs used as instruments. Note that if we define

$$\hat{Z} = P_X Z = \begin{bmatrix} \hat{Y}_1 & X_1 \end{bmatrix}$$

so that $\hat{Z}$ are the instruments for Z, then we can write

$$\hat{\delta} = (\hat{Z}' Z)^{-1} \hat{Z}' y$$

2SLS Estimation - 3

The 2SLS varcov estimator is

$$\hat{V}(\hat{\delta}) = \left(Z'\hat{Z}\right)^{-1} \hat{\sigma}^2_{IV}$$

which can also be written as $\hat{V}(\hat{\delta}) = \left(\hat{Z}'\hat{Z}\right)^{-1} \hat{\sigma}^2_{IV}$.

Properties of 2SLS:

1. Consistent
2. Asymptotically normal
3. Biased when the mean exists (the existence of moments is a technical issue we won't go into here).
4. Asymptotically inefficient, except in special circumstances.

The Hausman's test

Consider the simple linear regression model $y_t = x_t'\beta + \epsilon_t$.

We assume that the functional form and the choice of regressors is correct, but that the some of the regressors may be correlated with the error term.

For example, this will be a problem if

- if some regressors are endogenous,
- or some regressors are measured with error,
- or lagged values of the dependent variable are used as regressors and ϵ_t is autocorrelated.

The Hausman's test - 2

The idea behind the Hausman's test is the following: A pair of consistent estimators converge to the same probability limit, while if one is consistent and the other is not they converge to different limits.

If we accept that one is consistent (*e.g.*, the IV estimator), but we are doubting if the other is consistent (*e.g.*, the OLS estimator), we might try to check if the difference between the estimators is significantly different from zero.

Under the null hypothesis that they both consistent and CLT, we have

$$[I_K \quad -I_K] \begin{bmatrix} \sqrt{n}\left(\tilde{\beta} - \beta_0\right) \\ \sqrt{n}\left(\hat{\beta} - \beta_0\right) \end{bmatrix} = \sqrt{n}\left(\tilde{\beta} - \hat{\beta}\right),$$

will be asymptotically normally distributed as

$$\sqrt{n}\left(\tilde{\beta} - \hat{\beta}\right) \xrightarrow{d} N\left(0, V_\infty(\tilde{\beta}) - V_\infty(\hat{\beta})\right).$$

The Hausman's test - 2

So,

$$n\left(\tilde{\beta}-\hat{\beta}\right)'\left(V_\infty(\tilde{\beta})-V_\infty(\hat{\beta})\right)^{-1}\left(\tilde{\beta}-\hat{\beta}\right)\overset{d}{\rightarrow}\chi^2(\rho),$$

where ρ is the rank of the difference of the asymptotic variances.

A statistic that has the same asymptotic distribution is

$$\left(\tilde{\beta}-\hat{\beta}\right)'\left(\hat{V}(\tilde{\beta})-\hat{V}(\hat{\beta})\right)^{-1}\left(\tilde{\beta}-\hat{\beta}\right)\overset{d}{\rightarrow}\chi^2(\rho).$$

This is the Hausman test statistic, in its original form.

The reason that this test has power under the alternative hypothesis is that in that case the "OLS" estimator will not be consistent, and will converge to β_A, say, where $\beta_A \neq \beta_0$. Then the mean of the asymptotic distribution of vector $\sqrt{n}\left(\tilde{\beta}-\hat{\beta}\right)$ will be $\beta_0-\beta_A$, a non-zero vector, so the test statistic will eventually reject, regardless of how small a significance level is used.