Econometrics-1

23.02.2021

Lecture 12

$$\underset{N\times 1}{\underline{y}} = \underset{N\times K}{X}\ \underset{K\times 1}{\underline{\beta}} + \underset{N\times 1}{\underline{u}}$$

A1. $E(\underline{u}|X) = 0 \Rightarrow E(X'\underline{u}) = 0$

A2: $V(u|X) = \sigma^2 I_n \Leftrightarrow E(\underline{u}\,\underline{u}'|X) = \sigma^2 \underset{N\times N}{I_N}$

A3: X is full-rank $\Rightarrow (X'X)$ is non-singular, PD matrix

A4: X is non-stochastic

A5: $\underset{N\times 1}{\underline{u}} \sim N\left(\underset{N\times 1}{\underline{0}}, \underset{N\times N}{\sigma^2 I_N}\right)$

$\underset{K\times 1}{\underline{\beta}}$: need to estimate this coefficient vector

Method of least squares:

$$\min_{\underline{\beta}=\{\beta_1,\beta_2,\ldots\beta_K\}'} \sum_{i=1}^{N} u_i^2 \equiv \min_{\underline{\beta}} \underbrace{S}_{\text{scalar}} = \underset{1\times N}{\underline{u}'}\ \underset{N\times 1}{\underline{u}}$$

$$= \sum_{i=1}^{N} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_K x_{iK})^2$$

$$\min_{\underline{\beta}} \; S = \underline{u}'\underline{u} = (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$$

$$\equiv \min_{\beta} \left(\underline{y}'\underline{y} - 2\beta' X'\underline{y} + \underline{\beta}' X'X\underline{\beta} \right)$$

[FONC]: First-order Necessary condition for optimization.

$$\underline{FOC}: \; \underbrace{\frac{\partial S}{\partial \underline{\beta}}}_{\text{vector}}^{\text{scalar}} \overset{\text{set}}{=} 0$$

[Device: Jacobian]

$$\frac{\partial S}{\partial \underline{\beta}} = -2X'\underline{y} + 2X'X\underline{\beta} \overset{\text{set}}{=} 0$$

$$(X'X)\underline{\beta} = X'\underline{y}$$

$$(X'X)^{-1}(X'X)\underline{\beta} = (X'X)^{-1}X'\underline{y}$$

$$\boxed{\hat{\underline{\beta}}_{OLS} = (X'X)^{-1}X'\underline{y}}$$

(second-order sufficient condition for minimization)

SOSC :

$$\frac{\partial^2 S}{\partial \underline{\beta}\, \partial \underline{\beta}'} > 0$$

$$\frac{\partial^2 S}{\partial \underline{\beta}\, \partial \underline{\beta}'} = \underbrace{2(X'X) > 0}$$

i.e, $X'X$ is PD (as A3).

Statistical properties of the OLS estimation

① $\underline{\hat{\beta}}_{OLS}$ is unbiased

$$\underset{K\times 1}{E(\underline{\hat{\beta}}_{OLS})} = \underset{K\times 1}{\beta}$$

$$E(\hat{\beta}_{OLS}) = \beta$$

Proof:

$$E(\hat{\beta}_{OLS}) = E\left[(X'X)^{-1}X'\underline{y}\right]$$

$$= E\left[(X'X)^{-1}X'(X\beta+\underline{u})\right]$$

$$= E\left[\cancel{(X'X)^{-1}X'X}\,\underline{\beta} + (X'X)^{-1}X'\underline{u}\right]$$

$$= \underline{\beta} + (X'X)^{-1}X'\underbrace{E(\underline{u})}_{=0} \quad \text{(: A1 \& A4)} \rightarrow 0$$

$$= \underline{\beta}$$

② $\underset{K\times 1}{V(\underline{\hat{\beta}}_{OLS})} = \underbrace{\Sigma_\beta}_{\text{dimension} = ?}$

$K \times K.$

$\underset{K\times 1}{V(\underline{\hat{\beta}}_{OLS})} = V\left((X'X)^{-1}X'\underline{y}\right) = V(\underset{K\times N}{A}\ \underset{N\times 1}{\underline{y}})$

Consider x is a scalar r.v.

$y = ax + b$ a, b are constants.

$V(y) = a^2 V(x)$ ✓

One step further... →

$\underset{1\times 1}{y} = \underset{1\times N}{\underline{a}'}\ \underset{N\times 1}{\underline{x}} = [a_1\ a_2 \cdots a_N]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$

$= a_1x_1 + a_2x_2 + \cdots + a_Nx_N = \sum_i a_i x_i$

where $V(\underline{X}) = \Sigma_X$

$$\underset{1\times 1}{V(y)} = E\left[[\underline{a}'\underline{x} - \underline{a}'\underline{\mu}_x] \underbrace{[\underline{a}'\underline{x} - \underline{a}'\underline{\mu}_x]'}_{=(\underline{x}'\underline{a} - \underline{\mu}_x'\underline{a})} \right]$$

$$= E\left[\underline{a}'(\underline{x} - \underline{\mu}_x)(\underline{x}' - \underline{\mu}')\underline{a} \right]$$

$$= \underline{a}'\left\{ E\left[(\underline{x} - \underline{\mu})(\underline{x}' - \underline{\mu}') \right] \right\}\underline{a}$$

$$= \underline{a}' \; E\left[(x - \mu)(x - \mu)' \right] \underline{a}$$

$$= \underline{a}' V(\underline{x}) \underline{a} = \underbrace{\underset{1\times N}{\underline{a}'} \underbrace{\Sigma_x}_{\substack{=V(x) \\ N\times N}} \underset{N\times 1}{\underline{a}}}_{1\times 1}$$

One-step further:

Consider two data points

$$y_1 = a_{11} z_1 + a_{12} z_2 + a_{13} z_3$$

constants / model coefficients

$$y_2 = a_{21} z_1 + a_{22} z_2 + a_{23} z_3$$

random variables

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

$$\underset{2\times 1}{\underline{y}} = \underset{2\times 3}{A} \cdot \underset{3\times 1}{\underline{Z}}$$

Given $E(\underline{Z}) = \underline{\mu}_Z$ $\quad V(\underline{Z}) = \underset{3\times 3}{\Sigma_Z}$

find $V(\underline{y})$

$$V(\underline{y}) = E\left([\underline{y} - E(\underline{y})][\underline{y} - E(\underline{y})]'\right)$$

$$\boxed{V(\underline{y}) = A \cdot V(Z) \cdot A' = A \Sigma_Z A'}$$

Back to our original query

$$V(\hat{\underline{\beta}}_{OLS}) = V\left((X'X)^{-1} X' \underline{y}\right) = V(A\underline{y})$$

$$= A \; V(\underline{y}) A'$$

$$= (X'X)^{-1} X' \; \sigma^2 I_N \underbrace{[(X'X)^{-1} X']'}_{= [X((X'X)^{-1})'] = X(X'X)^{-1}}$$

$$V(\underset{\sim}{\hat{\beta}}_{OLS}) = (X'X)^{-1} X' (\underset{\sim}{\sigma^2} I_N) X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} \underbrace{X' I_N}_{= X'} X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} \cancel{(X'X)} \cancel{(X'X)^{-1}}$$

$$\boxed{\underset{K\times 1}{V(\underset{\sim}{\hat{\beta}}_{OLS})} = \underset{K\times K}{\sigma^2 (X'X)^{-1}}}$$

Aside,

$$\underset{\sim}{\hat{\beta}}_{OLS} = A \cdot \underset{\sim}{y} \qquad A = (X'X)^{-1} X'$$

$$\rightarrow A' = [(X'X)^{-1} X']'$$

$$V(\underset{\sim}{\hat{\beta}}_{OLS}) = A \cdot V(\underset{\sim}{y}) \cdot A'$$

$$A' = [(X')'][(X'X)^{-1}]'$$

$$= X [(X'X)']^{-1}$$

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

(iii) $\hat{\underline{\beta}}_{OLS}$ is an efficient estimator.

Among all the unbiased estimators of $\underline{\beta}$, $\hat{\underline{\beta}}_{OLS}$ has the lowest variance only if $\underline{u} \sim N(\underline{0}, \sigma^2 I_N)$.

That's why $\hat{\underline{\beta}}_{OLS}$ is often termed as

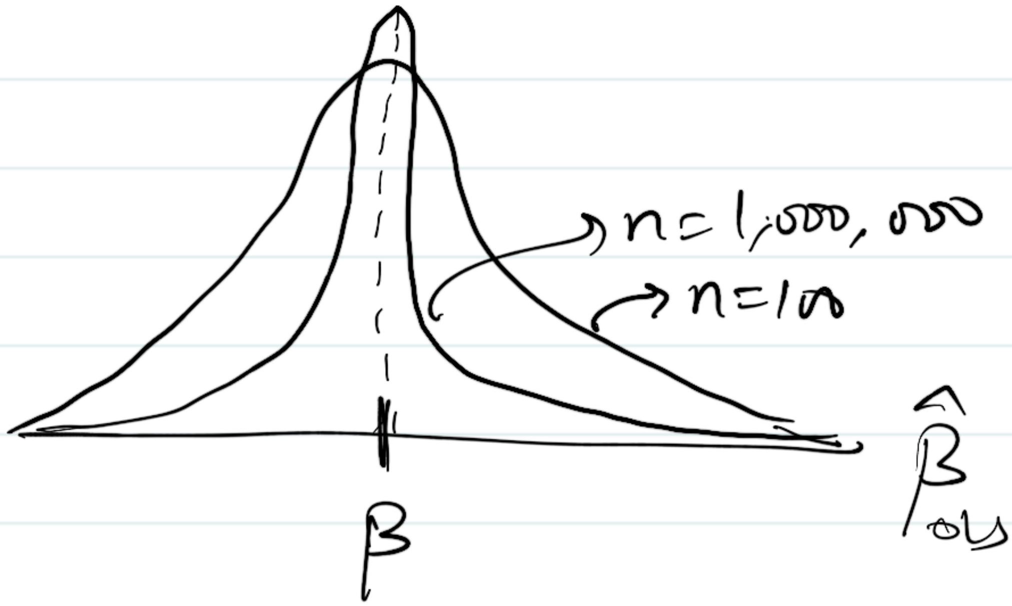
BLUE.

Best Linear Unbiased Estimator.

(iv) Consistency (large sample property)

$\hat{\theta}$ is a consistent estimator of true parameter θ iff (if and only if) for any $\varepsilon > 0$

$$\lim_{N \to \infty} \text{prob}\left(|\hat{\theta} - \theta| < \varepsilon\right) = 1.$$



Consistency is a HIGHLY DESIRABLE PROPERTY, even more desirable than unbiasedness.