LESSON 7 CLASSICAL LINEAR REGRESSION (Part 3)

The model

$y = X\beta + u \rightarrow (1)$

$Eu = 0 \rightarrow (2)$

$Euu' = \sigma^2 I_n \rightarrow (3)$

Estimated β

$$\hat{\beta} = \underbrace{(X'X)^{-1} X'}_{A} y = Ay$$

This part is non-stochastic

$$\underset{(K\times 1)}{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}} = \underset{(K\times n)}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ - & - & - & - \\ a_{K1} & a_{K2} & \cdots & a_{Kn} \end{bmatrix}} \underset{(n\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}$$

values of a_{ij} s are function of X_{ij} values.

These are therefore non-stochastic and constant.

$\hat{\beta}_1 = a_{11}Y_1 + a_{12}Y_2 + \dots + a_{1n}Y_n$

similarly

$\hat{\beta}_2 = a_{21}Y_1 + a_{22}Y_2 + \dots + a_{2n}Y_n$

In general,

$\hat{\beta}_i = a_{i1}Y_1 + a_{i2}Y_2 + \dots + a_{in}Y_n$

So $\hat{\beta}_i$ are linear combinations of Y_t values ← LINEAR

Are they also unbiased?

Let us see

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) \quad | \text{ using equation } (1)$$

$$= \underbrace{(X'X)^{-1}X'X}_{I_K}\beta + \underbrace{(X'X)^{-1}X'}_{A(k\times n)}u$$

$$= \underset{(k\times 1)}{\beta} + \underset{(k\times n)}{A}\underset{(n\times 1)}{u}$$

$$E(\hat{\beta}/X_{it} \text{ values}) = E(\beta + Au)$$
$$= E(\beta) + E(Au)$$
$$= \beta + A\,E(U)$$
$$= \beta$$

$\because \beta$ have only constant elements and A has only non-stochastic elements

$[\because E(U) = 0$ by equation (3)$]$

$\therefore E\hat{\beta}_1 = \beta_1,\ E\hat{\beta}_2 = \beta_2 \dots E\hat{\beta}_K = \beta_K$ i.e. OLS estimators are unbiased too

Are the OLSEs efficient?

To answer this we have to get the expression of variances of $\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_k$ and see if these variations are the smallest possible.

On our way to do that we ~~must~~ need to learn the idea of a variance - covariance matrix.

If x is a column vector of m random variables and $Ex = \mu$, then

$E(x-\mu)(x-\mu)'$ is called the variance covariance matrix of x. Why? Because the elements on the diagonal of this matrix will give variances of variables in x and the elements off the diagonal will give covariances between pairs of these variables.

Let us illustrate with a small (2×1) vector -

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{(2\times1)}$ $\mu = \begin{bmatrix} Ex_1 \\ Ex_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

Now, $E(x-\mu)(x-\mu)'$

$$= E \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}_{(2\times1)} \begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}_{(1\times2)}$$

$$= E \begin{bmatrix} (x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)^2 \end{bmatrix}_{(2\times2)}$$

$$= \begin{bmatrix} E(x_1-\mu_1)^2 & E(x_1-\mu_1)(x_2-\mu_2) \\ E(x_2-\mu_2)(x_1-\mu_1) & E(x_2-\mu_2)^2 \end{bmatrix}_{(2\times2)}$$

$$= \begin{bmatrix} Var X_1 & Cov X_1X_2 \\ Cov X_2X_1 & Var X_2 \end{bmatrix}$$ by definition

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$ by symbols

Now let us try to get the variance-covariance matrix of $\hat{\beta}$ (we denote by $V\hat{\beta}$)

$$V(\hat{\beta}) = E[\{\hat{\beta} - E(\hat{\beta})\}\{\hat{\beta} - E(\hat{\beta})\}']$$

$$V(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

[Earlier we found that $\hat{\beta} = \beta + (X'X)^{-1}X'u$

$\Rightarrow \hat{\beta} - \beta = (X'X)^{-1}X'u$]

Thus, $V\hat{\beta} = E\{(X'X)^{-1}X'u\} \times \{(X'X)^{-1}X'u\}'$

$= E\{(X'X)^{-1}X'u\}\{u'X''(X'X)^{-1'}\}$ | $\because (AB)' = B'A'$

$= E\,(X'X)^{-1}X'uu'X(X'X)^{-1'}$

$= (X'X)^{-1}X'\,E uu'\,X(X'X)^{-1'}$

[$\because$ X has only non-stochastic element, we can treat them as constant and take the E operation across $(X'X)^{-1}X'$ part]

$V\hat{\beta} = (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1'}$ | By equation (3)

$= \sigma^2 (X'X)^{-1}X' I_n X (X'X)^{-1'}$ | $\because \sigma^2$ is constant

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

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

| since $X'X$ and $(X'X)^{-1}$ is symmetric $\therefore (X'X)^{-1'} = (X'X)^{-1}$

Summing up, we have

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

We shall resume from here to check if OLSE of β, i.e. $\hat{\beta} = (X'X)^{-1}X'y$, which is linear and unbiased is also efficient.

Lesson 7 CLASSICAL LEANER REGRESSION (Part 4)

We have shown that OLSE of β is linear (in Y_t s) and unbiased [i.e. $E(\hat{\beta}/X) = \beta$]. Next question is — Is $\hat{\beta}$ also efficient ??

Background

The model

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{u} \longrightarrow (1)$$

$$\rho X = k \longrightarrow (2)$$

$$EU = \underset{\text{(null)}}{0} \longrightarrow (3)$$

$$Euu' = \sigma^2 I_n \longrightarrow (4)$$

[As of now we do not need 5th assumption]

Estimators

$$\hat{\beta} = \underbrace{(X'X)^{-1}X'}_{\substack{A \\ k\times n}}y$$

$$= Ay$$

$$E\hat{\beta} = \beta$$

$$V\hat{\beta} = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$

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

Proof: (that $\hat{\beta}$ is also efficient).

Let us arbitrarily design linear, unbiased estimator b of β. Let the matrix c contain the difference of coefficient of y in b from those in $\hat{\beta}$.

i.e.

elements of c are arbitrary constants

$\cancel{b+(A}$ $b = (A+c)y$

$$E(b/x) = E\{(A+c)y\}$$
$$= E\{(A+c)(x\beta+u)\}$$
$$= E(A+c)x\beta + E(A+c)u$$
$$= (A+c)x\beta + (A+c)Eu$$

[$\because$ A and c has constant only and x is non-stochastic]

$[\because Eu = 0$ by equation (2)

$$= (A+c)x\beta$$

$$= \underbrace{Ax}_{I_K}\beta + cx\beta = \beta + cx\beta = (I+cx)\beta$$

Given that b is unbiased [i.e. $E(b/x) = \beta$].

$\underbrace{cx}_{\text{null}} = 0$ $\quad \therefore \underbrace{x'c'}_{\text{null}} = 0$

$$Vb = E\{(b-Eb)(b-Eb)'\}$$
$$= E\{(A+c)u\}\{(A+c)u\}'$$

$b = (A+c)x\beta + (A+c)u$

$E(b/x) = (A+c)x\beta + 0$

$(b - Eb) = (A+c)u$

$$= E(A+c)uu'(A+c)'$$
$$= (A+c)Euu'(A'+c')$$
$$= \sigma^2(A+c)(A'+c')$$

by (5)th equation $Euu' = \sigma^2 I_n$

$$= \sigma^2 AA' + \sigma^2 Ac' + \sigma^2 cA' + \sigma^2 cc'$$

$Ac' = (x'x)^{-1} \underbrace{x'c'}_{0} = 0$ by unbiasedness of b

AA' ~~are~~ we have seen earlier $= (x'x)^{-1}$

similarly $cA' = 0$

Thus $\underset{(k\times k)}{V(b)} = \sigma^2 (X'X)^{-1} + \sigma^2 CC'$ | V denotes var-cov

$= V\hat{\beta} + \sigma^2 CC'$

Let us pick the elements in 1st row and 1st column —

$$\sigma^2 b_1 = \sigma^2 \hat{\beta}_1 + \sigma^2 (c_{11}\ c_{12} \cdots c_{1n}) \begin{pmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1n} \end{pmatrix}$$

$$= \sigma^2 \hat{\beta}_1 + \sigma^2 (c_{11}^2 + c_{12}^2 + \cdots + c_{1n}^2)$$

$$= \sigma^2 \hat{\beta}_1 + \sigma^2 \sum_{j=1}^{n} c_{1j}^2$$

In general

$$\sigma^2 b_i = \sigma^2 \hat{\beta}_i + \sigma^2 \sum_{j=1}^{n} c_{ij}^2$$

$\because \sum c_{ij}^2 > 0$ [being sum of squares]

$\therefore \sigma^2 b_i \geq \sigma^2 \hat{\beta}_i$

No linear unbiased estimator can have lower variance than the OLS estimators of β_i s.

Now let us check when equality will hold.

i.e. $\sum c_{ij}^2 = (c_{i1}^2 + c_{i2}^2 + \cdots c_{in}^2 = 0$

For this to hold each c_{ij} must be individually zero i.e. $c_{ij} = 0 \quad \forall\, i$ and j

i.e. $C \rightarrow$ null

In that case $b = (A+C)y$

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

$\Rightarrow$ OLS estimators alone have the smallest variance in the class of linear unbiased estimators.