The Algebra of Least Squares

Lecture 4

We have an ~~random sample~~ $\cancel{\{(y_i, x_i)\}_{i=1}^{n}}$ joint dist of (y, x)

We want to find out the linear function that

minimizes $S(\beta) = E(y - x\beta)^2$

We already get $\beta^* = E(x'x)^{-1} E xy$.

When we have a sample, it is nature to est.

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

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

where X is $n \times k$ matrix and $y = n \times 1$ vector.

Similarly, we can $\frac{1}{n}\sum_{i=1}^{n} (y_i - x\beta)^2$.

all steps are the same except that we replace $E[\,]$ by $\frac{1}{n}\sum_{i=1}^{n}$

Def: fitted vale $\hat{y} = X\hat{\beta}$.

residual $= y - \hat{y}$

Property of $\hat{e}$.

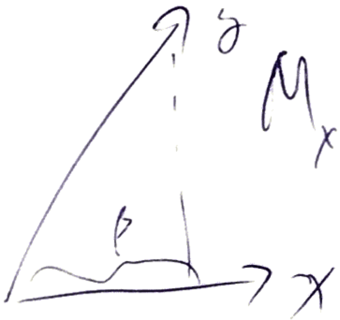
$$\underset{n\times k}{X'}\ \underset{n\times 1}{\hat{e}} = X(Y - X(X'X)^{-1}X'Y)$$

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

$$= XY - \cancel{(X'X)}^{-1}\cancel{(X'X)}^{-1}XY = 0.$$

it implies that $1'e = 0$. if a constant is contained.

$P_x = X'(XX)^{-1}X$. is called a projection matrix.



$M_x = I - P_x$ is an annihilator.

because $X(I - P_x) = 0$.

$P_x P_x = P_x$. $M_x M_x = M_x$.

Whatever matrix if $AA = A$, idempotent

Represent components

$X = [X_1\ X_2]$. $\beta = (\beta_1\ \beta_2)$

$$\hat{Q}^{-1} = \begin{pmatrix} X_1X_1 & X_1X_2 \\ X_2X_1 & X_2X_2 \end{pmatrix}^{-1}$$

$$= \begin{bmatrix} (\hat{Q}_{11} - \hat{Q}_{12}\hat{Q}_{21}\hat{Q}_{22})^{-1} & -(\quad)\hat{Q}_{12}\hat{Q}_{22}^{-1} \\ -(\quad)\hat{Q}_{21}\hat{Q}_{11}^{-1} & (\hat{Q}_{22} - \hat{Q}_{21}\hat{Q}_{11}^{-1}\hat{Q}_{11})^{-1} \end{bmatrix}$$

$$\begin{pmatrix} \hat{Q}_{1y} \\ \hat{Q}_{2y} \end{pmatrix}$$

$$= \hat{Q}_{11\cdot 2}\left(\hat{Q}_{1y} - \hat{Q}_{12}\hat{Q}_{22}\hat{Q}_{2y}\right).$$

$$\hat{Q}_{11\cdot 2} = \frac{1}{n}X_1X_1 - \frac{1}{n}XX'(X_2'X_2)X_2X_1$$

$$= \frac{1}{n}X_1'M_2X_1$$

and. $\hat{Q}_{1y} - \hat{Q}_{11}\hat{Q}_{12}\hat{Q}_{2y} = \cancel{Q}\, X_1 M_2 y$

so $\hat{\beta}_1 = (X_1M_2X_1)^{-1}(X_1'M_2y)$.

Frisch-Waugh-Lovell

FWL Theorem

we can get $\hat{\beta}_1$ by regress. M_2y vs M_2X_1

residual of y on X_2

residual of X_1 on X_2

Lecture 4

Statistical Properties of Least Squares.

Consider a linear regress model

$$y = X\beta + \varepsilon, \quad \text{where } \underline{E(\varepsilon | X) = 0.}$$

an Assumption.

start from $E(\hat{\beta}|X)$

$$E(\hat{\beta}|X) = E((X'X)^{-1}X'y) = E[(X'X)^{-1}(X'(X\beta_0+\varepsilon)]$$

$$= E[E((X'X)^{-1}X'y|X)] + \beta_0$$

$$= E((X'X)^{-1}X'\varepsilon|X] + \beta_0$$

$$= E[(X'X)^{-1}X'E(\varepsilon|X)] + \beta_0 = \beta_0.$$

unbiased than $E(\hat{\beta}) = 0$ of course.

$$\text{var}(\hat{\beta}|X) = E((\hat{\beta}-\beta_0)(\hat{\beta}-\beta_0)'|X)$$

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X]$$

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

if homo. $\text{var}(\hat{\beta}|X) = (X'X)^{-1}(X'X)(X'X)^{-1}\sigma^2$

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

if hsk, $\text{var}(\hat{\beta}|X) = (X'X)^{-1}(\sum X_i X_i' \sigma_i^2)(X'X)^{-1}$

Under HSK,

$$Var(\hat{\beta}|X) = (X'X)^{-1}X'DX(X'X)^{-1}$$

where $D = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$

Gauss-Markov Theorem. OLS is BLUE

under homo.

for any linear estimator $\tilde{\beta} = Ay$ Because $\tilde{\beta}$ is unbiased

$$E(\tilde{\beta}|X) = E(Ay|X) = AX\beta = \beta \quad AX = I_k$$

to make it unbiased. $\text{"} AE(y|X) = AX\beta$.

X is not a square matrix. So $A \neq X^{-1}$

$$Var(\tilde{\beta}|X) = var(A'y|X) = A'DA.$$

$$= \cancel{ADA} \qquad = A'A\sigma^2.$$

To show

for any A such that $A'X = I_k$, we have $A'A - \underbrace{(X'X)^{-1}}_{\text{variance of OLS}}$

is positive semi-definite.

let $C = A - X(X'X)^{-1}$

$$A'A - (X'X)^{-1} = (C + X(X'X)^{-1})'((C + X(X'X)^{-1}) - (X'X)^{-1}$$

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

Note $C'X = (A - X(X'X)^{-1})'X = I - I = 0$

$C'X = 0$ as $A'X = I_k$.

for general case HSK.

BLUE is

$$\hat{\beta} = (X'D^{-1}X)^{-1}X'D^{-1}y \quad \text{weighted.}$$

infeasible.

How to estimate the variance?

propose. homo: $\hat{\sigma}^2 = \frac{1}{n}\sum \hat{e}_i^2$

$$\text{HSK. } \widehat{X'DX} = \frac{1}{n}\sum x_i x_i' \hat{e}_i^2$$

$$\mathrm{Var}(\hat{\beta}) = \frac{1}{n}(X'X)^{-1}\widehat{X'DX}(X'X)^{-1}.$$

Need to have $(X'X)$ invertible.

perfect collinearity. three category introduce three variables, but can only use two of them.

In homo, nearly collinearity.

$$\frac{1}{n}(X'X) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

$$\mathrm{Var}(\hat{\beta}|X) = \frac{\sigma^2}{n}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \frac{\sigma^2}{n(1-\rho^2)}\begin{pmatrix} \boxed{1} & -\rho \\ -\rho & \boxed{1} \end{pmatrix}.$$

$\mathrm{Var}(\hat{\beta}_1|X) = \cdots$