18.2.20

Econometrics

Economics + Statistics + Mathematics

We use statistics and mathematics to deduce the economic theory.

Classical Linear Regression Model (CLRM)

$$Y_i = \underline{a} + \underline{b} X_i + u_i$$

Y_i → dependent Variable ⇒ Explained Variable

X_i → Independent Variable ⇒ Explanatory Variable

u_i → Random disturbance term

a and b are known as parameters of the model

$i = 1, 2, \ldots n$

$n \rightarrow$ no. of samples.

$x_i = (X_i - \bar{X})$ $\qquad xy = (X_i - \bar{X})(Y_i - \bar{Y})$

$y_i = (Y_i - \bar{Y})$

We estimate this CLRM by applying ordinary least Square (OLS) method.

To apply OLS we have to make some important assumptions →

(1) Linearity holds in case of parameters as well as in case of variables i.e ~~both~~ power of (X & Y) and (a & b) must be equal to 1.

$$Y_i = a + bX_i + U_i \quad \text{(correct)}$$

Incorrect model
$$\begin{cases} Y_i = a + \sqrt{b}\,X_i + U_i & (\times) \\ Y_i = a + b\sqrt{X_i} + U_i & (\times) \end{cases}$$

(2) Expected value of u_i for all given x_i must be equal to zero.

$$E(u_i / x_i) = 0 \quad \forall \; i = 1, 2 \ldots n$$

(3) Variance of error term is same for all obs

i.e.

$$Var(u_1) = Var(u_2) = \ldots = Var(u_n) = \sigma_u^2 \quad \forall \; i = 1, 2, \ldots n$$

$$Var(u_i) = E[u_i - E(u_i)/X_i]^2$$

$$= E[u_i^2]/X_i \quad \text{as } E(u_i) = 0$$

$$= \sigma_u^2$$

This assumption is known as assumption of homoscedasticity

(4) The explainatory variable is non-stochastic i.e non random

$$E(X_i) = X_i \quad \forall \; i = 1, 2 - n$$

(5) The explainatory variables are not perfectly related i.e. there is no perfect multicollinearity

$$Y_i = a + b_1 x_{1i} + b_2 x_{2i} + u_i$$

$$r_{x_1 x_2} \neq \pm 1 \quad \text{i.e } x_1 \neq K X_2$$

(6) There is no autocorrelation

i.e u_i's are not correlated.

$$cov(u_i, u_j) = 0$$

$$= E[u_i u_j / x_i x_j] = 0 \quad \forall \; i \neq j$$

(7) The explainatory variable X_i and random disturbance term are uncorrelated.

$$cov(X_i, u_i) = 0 \quad \forall \quad i = 1, 2 \ldots n$$

(8) The dependent variable Y_i and u_i are uncorrelated

$$cov(Y_i, u_i) = 0 \quad \forall \quad i = 1, 2 \ldots n$$

(9) The no. of ~~various~~ parameters ~~is~~ must be less than the no. of obs.

$$\boxed{k \leq n}$$

where k is the no. of parameters.

In 2 - variable case, $k = 2$

$$\boxed{2 \leq n}$$

(10) The model must be correctly specified, ~~there~~ otherwise, there will be specification bias.

Estimation

~~ere~~ The estimated or fitted line is,

$$\hat{Y} = \hat{a} + \hat{b} X_i$$

The error term is defined as,

$$e_i = (Y_i - \hat{Y})$$

$$= (Y_i - \hat{a} - \hat{b} X_i)$$

$$\Rightarrow \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b} X_i)^2$$

Now, our aim is to minimise the square of the error term, the FOC of minimisation requires,

$$\frac{\partial \Sigma e_i^2}{\partial \hat{a}} = 2 \sum_{i=1}^{n} \left[Y_i - \hat{a} - \hat{b} X_i \right] (-1) = 0$$

$$\Rightarrow \cancel{2} \sum_{i=1}^{n} \left[Y_i - \hat{a} - \hat{b} X_i \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i - n\hat{a} - \hat{b} \sum_{i=1}^{n} X_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i - \hat{b} \sum_{i=1}^{n} X_i = n\hat{a}$$

$$\Rightarrow \hat{a} = \frac{\sum Y_i}{n} - \hat{b}\frac{\sum X_i}{n}$$

$$\boxed{\hat{a} = \bar{Y} - \hat{b}\bar{X}}$$

$$\frac{\partial \sum e_i^2}{\partial \hat{b}} = 2\sum_{i=1}^{n}\left(Y_i - \hat{a} - \hat{b}X_i\right)\left(-X_i\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left[Y_i X_i - \frac{\left(\sum Y_i - \hat{b}\sum X_i\right)}{n} X_i - \hat{b} X_i^2 \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left[Y_i X_i - \frac{\left(\sum Y_i\right) X_i}{n} + \hat{b}\frac{\left(\sum X_i\right) X_i}{n} - \hat{b} X_i^2 \right] = 0$$

$$\Rightarrow \sum Y_i X_i - \frac{\left(\sum Y_i\right)\left(\sum X_i\right)}{n} + \hat{b}\frac{\left(\sum X_i\right)\left(\sum X_i\right)}{n} - \hat{b}\sum X_i^2 = 0$$

$$\Rightarrow \hat{b}\left[\frac{\sum X_i^2}{n} - \left(\frac{\sum X_i}{n}\right)^2\right] = \left[\frac{\sum X_i Y_i}{n} - \frac{\left(\sum X_i\right)}{n}\frac{\left(\sum Y_i\right)}{n}\right]$$

$\Rightarrow \quad \hat{b} \; var(x) = cov(x, y)$

$$\Rightarrow \quad \boxed{\hat{b} = \frac{cov(x,y)}{Var(x)}}$$

These are estimated values of $\hat{a}$ and $\hat{b}$.

21/02/20

Variance of $\hat{b}$

$$Var(\hat{b}) = E\left[\hat{b} - E(\hat{b})\right]^2$$

Now, as $\hat{b}$ is unbiased estimator of b.
Therefore $\Rightarrow$

$$\boxed{E(\hat{b}) = b} = E\left[\hat{b} - b\right]^2$$

$$\hat{b} = \frac{cov(x,y)}{Var(x)} = \frac{1/n \; \Sigma(x_i - \bar{x})(y_i - \bar{y})}{1/n \; \Sigma(x_i - \bar{x})^2}$$

$$= \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i \, (\cancel{a} + b x_i + u_i - \cancel{a} - b\bar{x})}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i \, \{b(x_i - \bar{x}) + u_i\}}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i \, \{b x_i + u_i\}}{\Sigma x_i^2}$$

$$= \frac{b \, \Sigma x_i^2 + \Sigma x_i u_i}{\Sigma x_i^2}$$

$$\Rightarrow \hat{b} = b + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$\Rightarrow E(\hat{b}) = E(b) + \frac{\sum x_i \, E(u_i)}{\sum x_i^2}$$

$$\boxed{E(\hat{b}) = b} \quad \text{as } E(u_i) = 0$$

$\therefore$ $\hat{b}$ is unbiased estimator.

$$E[\hat{b} - b]^2 = \text{Var}(\hat{b})$$

$$\Rightarrow \text{Var}(\hat{b}) = E\left[\frac{\sum x_i u_i}{\sum x_i^2}\right]^2$$

$$= \frac{1}{(\sum x_i^2)^2} E\left[\sum x_i^2 u_i^2 + 2 \sum_{\substack{i<j \\ i=j}}\sum x_i u_i x_j u_j\right]$$

$$= \frac{1}{(\sum x_i^2)^2}\left[\sum x_i^2 E(u_i^2) + 2 \sum_{i<j}\sum x_i x_j E(u_i u_j)\right]$$

$$= \frac{\cancel{\sum x_i^2} \, \sigma^2_u}{(\sum x_i^2)^{\cancel{2}}} + 0 \quad \text{as } E(u_i u_j) = 0$$

by the assumption

$$= \sigma^2_u / \sum x_i^2$$

$$\boxed{\begin{aligned} \sum x_i^2 &= \sum (x_i - \bar{x})^2 \\ \text{Var}(x) &= \frac{1}{n}\sum (x_i - \bar{x})^2 \end{aligned}}$$

$$= \frac{\sigma^2_u}{n \, \text{var}(x)}$$

$$\boxed{\text{Var}(\hat{b}) = \frac{\sigma^2_u}{n \, \text{var}(x)}}$$

Now, $\hat{a} = \bar{Y} - \hat{b}\bar{X} = a + b\bar{X} - \hat{b}\bar{X}$

as mean of $u = 0$

$$\therefore \boxed{\hat{a} = a - \bar{x}(\hat{b} - b)}$$

$$E(\hat{a}) = E(a) - \bar{x}\, E(\hat{b} - b)$$

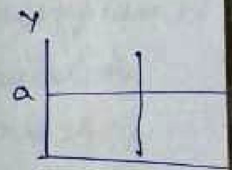
$$= E(a) - \bar{x}\, E\left[\frac{\Sigma x_i u_i}{\Sigma x_i^2}\right]$$

$$= a - \bar{x}\,\frac{\Sigma x_i\, E(u_i)}{\Sigma x_i^2}$$

$\boxed{E(\hat{a}) = a}$ as $E(u_i) = 0$

with intercept $\leftarrow$ $\boxed{Y = a + bx_i + u_i}$

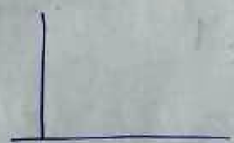
when $b = 0$, $\boxed{Y_i = a}$



$\boxed{Y = bx_i + u_i}$

$\downarrow$

without intercept



$Cov(\hat{a}, \hat{b}) \longrightarrow$ Relation b/w $\hat{a}$ & $\hat{b}$

$$= E\left[\{\hat{b} - E(\hat{b})\}\{\hat{a} - E(\hat{a})\}\right]$$

$$= E\left[(\hat{b} - b)(\hat{a} - a)\right]$$ as $\hat{b}$ and $\hat{a}$ are unbiased estimator

$$= E\left[(\hat{b} - b)\{-(\hat{b} - b)\cdot\bar{X}\}\right]$$

$\boxed{y, Y}$

$$= -\bar{X}\, E\left[\hat{b} - b\right]^2$$

$$= -\bar{X}\, Var(\hat{b}) = -\bar{X}\,\frac{\sigma_u^2}{\Sigma x_i^2}$$

as $Var(\hat{b}) = \sigma_u^2 / \Sigma x_i^2$

$$\underline{Var(\hat{a})} = E\left[\hat{a} - E(a)\right]^2$$

$$= E\left[\hat{a} - a\right]^2$$

$$= E\left[-\bar{x}(\hat{b} - b)\right]^2$$

$$= \bar{x}^2 \, E(\hat{b} - b)^2$$

$$= \bar{x}^2 \frac{\sigma_u^2}{\Sigma x_i^2} = \boxed{\frac{\bar{x}^2 \sigma_u^2}{n \, var(x)}}$$

<u>Goodness of fit</u>

$$r^2 = \left\{ \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y} \right\}^2$$

$$= \frac{Cov^2(x, y)}{Var(x) \, Var(y)}$$

$0 \leq r^2 \leq 1$ $\quad \because r^2 \to 1$ then the model is good fitted.

Relation b/w TSS, RSS and ESS

Total sum of square (TSS) is defined as total variation explained by explanatory variable ~~is known as~~ ESS. and residual sum of square known as TSS.

$$TSS = ESS + RSS$$

TSS ↓ Total sum of square.

ESS ↓ Explained by explanatory variable

RSS ↓ unexplained variation which is from u_i

$$ESS = (1-r^2)\sum x_i^2$$

$$= (1-r^2)\sum(x_i-\bar{x})^2$$

$$= (1-r^2)\, S_{yy}.$$

$$RSS = \sum \hat{u}_i^2$$

= sum of square explained by error term

The unbiased estimate of error variance.

$$\hat{u}_i^2 = \frac{RSS}{n-2}$$

Now, goodness of fit

$$\Rightarrow \frac{ESS}{TSS}.$$

As ESS $\rightarrow$ TSS then

$r^2 \rightarrow 1$

As ESS $\uparrow$es then r^2 $\uparrow$es

The model is goodfitted.

Testing of hypothesis

$$\hat{a}, \hat{b} \sim t_{(n-2)}$$

$\hat{a}$ and $\hat{b}$ follows 't' distribution

The hypothesis are $\Rightarrow$

$H_0 : \hat{a} = 0$

$H_1 : \hat{a} \gtrless 0.$

When H_0 is true then

$\hat{a}$ is not significant

Then in case of $\hat{b}$ then,

$H_0 : \hat{b} = 0$

$H_1 : \hat{b} \gtrless 0.$

we test the significancy of $\hat{a}$ and $\hat{b}$.

The test statistic is,

$$t_{\hat{a}} = \frac{\hat{a} - \hat{a}_{H_0}}{SE(\hat{a})}$$

$$= \frac{\hat{a}}{SE(\hat{a})}$$

and $$t_{\hat{b}} = \frac{\hat{b} - \hat{b}_{H_0}}{SE(\hat{b})}$$

$$= \hat{b} / SE(\hat{b})$$

we test the significancy of parameters.

There is a relation b/w r^2 and $t_{\hat{b}}$.

$$r^2 = \frac{t_{\hat{b}}^2}{t_{\hat{b}}^2 + n - 2}$$

Now, we are concentrating about $\hat{a}$ and $\hat{b}$

$\hat{a}$ and $\hat{b}$ are best linear unbiased estimator. (BLUE)

when $\hat{a}$ and $\hat{b}$ are BLUE the theorem is Gauss. Markov Theorem

95% confidence limit of $\hat{a}$ and $\hat{b}$.

$$\hat{a} = \hat{a}_{H_0} \pm t_{0.025}\, SE(\hat{a})$$

$$\hat{b} = \hat{b}_{H_0} \pm t_{0.025}\, SE(\hat{b})$$

$\hat{a} = \hat{a}_{H_0} \pm t_{0.005} \, SE(\hat{a})$

$\hat{b} = \hat{b}_{H_0} \pm t_{0.005} \, SE(\hat{b})$

Log-linear Model

$$Y = \alpha X^{\beta} \cdot u_i$$

$$\Rightarrow \log Y = \log\alpha + \beta \log X + \log u_i$$

$$\Rightarrow \boxed{Y^* = \alpha^* + \beta X^* + u_i^*}$$

$Y^* = \log Y \qquad X^* = \log X$

$\alpha^* = \log\alpha \qquad u_i^* = \log u_i$

$\Rightarrow \alpha = \text{anti log of } \alpha^*$

Semi-log model

$$\log Y = \alpha + X\log\beta + u_i$$

$$\Rightarrow \boxed{Y^* = \alpha + \beta^* X + u_i}$$

CLRM without intercept

$$Y_i = bX_i + u_i$$

Now, there is no intercept

i.e $a = 0$.

⊛ Formula

Let us defined,

$$S_{xx} = \Sigma(x_i - \bar{x})^2$$

$$S_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y})$$

$$S_{yy} = \Sigma(y_i - \bar{y})^2$$

1) $\hat{b} = \dfrac{S_{xy}}{S_{xx}}$; $\hat{a} = \bar{y} - \hat{b}\bar{x}$

2) $RSS = S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}$

$$= S_{yy} - \hat{b} S_{xy}$$

$$= S_{yy}(1 - r^2)$$

3) $r^2 = \dfrac{ESS}{TSS} = \hat{b}\dfrac{S_{xy}}{S_{yy}}$

4) $E(\hat{a}) = a$, $E(\hat{b}) = b$

$Var(\hat{a}) = \sigma^2\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right)$

$Var(\hat{b}) = \dfrac{\sigma^2}{S_{xx}}$

$cov(\hat{a}, \hat{b}) = \sigma^2\left(\dfrac{-\bar{x}}{S_{xx}}\right)$

$\hat{\sigma}^2 = \dfrac{RSS}{n-2}$, $\hat{\sigma}^2$ is an unbiased estimator of σ^2

5) $r^2 = \dfrac{t_{\hat{b}}^2}{t_{\hat{b}}^2 + (n-2)}$

6) $t_{\hat{a}} = \dfrac{\hat{a}}{SE(\hat{a})}$, when $H_0 : \hat{a} = 0$

$t_{\hat{a}} = \dfrac{\hat{a} - \hat{a}_{H_0}}{SE(\hat{a})}$, when $H_0 : \hat{a} = \hat{a}_0$

Similarly we can write in case of $\hat{b}$.

7) $ESS = \hat{b}\, S_{xy}$, $TSS = S_{yy}$, $ESS = r^2 S_{yy}$

8) $\dfrac{RSS}{\hat{\sigma}^2}$ follows χ^2 distribution with d.f $(n-2)$

3- Variable CLRM

The three variable model is defined as $\longrightarrow$

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

where X_1 and X_2 are two explainatory variables.

β_1 and β_2 are two slope parameters.

α is the intercept parameter

Now when all the assumptions of applying OLS are satisfied then we can able to estimate the parameters.

The aim of the researcher is to minimise the square of the error term

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

where $\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$

$$= \sum_{i=1}^{n} \left[Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right]^2$$

F.O.C of minimisation requires,

$$\frac{\partial \Sigma e_i^2}{\partial \hat{\alpha}} = 2 \Sigma \left(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)(-1) = 0$$

$$\Rightarrow \Sigma Y_i - n\hat{\alpha} - \hat{\beta}_1 \Sigma X_{1i} - \hat{\beta}_2 \Sigma X_{2i} = 0$$

$$\Rightarrow \hat{\alpha} = \frac{\Sigma Y_i}{n} - \hat{\beta}_1 \frac{\Sigma X_{1i}}{n} - \hat{\beta}_2 \frac{\Sigma X_{2i}}{n}$$

$$\Rightarrow \boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2}$$

$$\frac{\partial \Sigma e_i^2}{\partial \hat{\beta}_1} = 2 \Sigma \left(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)(-X_{1i}) = 0$$

$$\Rightarrow \Sigma \left[Y_i X_{1i} - \hat{\alpha} X_{1i} - \hat{\beta}_1 X_{1i}^2 - \hat{\beta}_2 X_{2i} X_{1i} \right] = 0$$

$$\Rightarrow \Sigma \left[Y_i X_{1i} - \left\{ \frac{\Sigma Y_i - \hat{\beta}_1 \Sigma X_{1i} - \hat{\beta}_2 \Sigma X_{2i}}{n} \right\} X_{1i} - \hat{\beta}_1 X_{1i}^2 - \hat{\beta}_2 X_{2i} X_{1i} \right] = 0$$

$$\Rightarrow \sum \left[Y_i X_{1i} - \frac{(\Sigma Y_i) X_{1i} + \hat{\beta}_1 X_{1i} \Sigma X_{1i} + \hat{\beta}_2 X_{1i} \Sigma X_{2i}}{n} \right.$$

$$\left. - \hat{\beta}_1 X_{1i}^2 - \hat{\beta}_2 X_{2i} X_{1i} \right] = 0$$

$$\Rightarrow \frac{\Sigma Y_i X_{1i}}{n} - \frac{(\Sigma X_{1i})(\Sigma Y_i)}{n} + \hat{\beta}_1 \frac{(\Sigma X_{1i})(\Sigma X_{1i})}{n}$$

$$- \hat{\beta}_1 \Sigma X_{1i}^2 + \hat{\beta}_2 \frac{(\Sigma X_{1i})(\Sigma X_{2i})}{n} - \hat{\beta}_2 \Sigma X_{2i} X_{1i} = 0$$

$$\Rightarrow \operatorname{cov}(Y, X_1) - \hat{\beta}_2 \left[\frac{\Sigma X_{2i} X_{1i}}{n} - \frac{(\Sigma X_{1i})}{n} \frac{(\Sigma X_{2i})}{n} \right] + \hat{\beta}_1 \left(\frac{\Sigma X_{1i}}{n}\right)^2 - \hat{\beta}_1 \frac{\Sigma X_{1i}^2}{n} = 0$$

$$\Rightarrow \operatorname{cov}(Y, X_1) - \hat{\beta}_2 \operatorname{cov}(X_2, X_1) = \hat{\beta}_1 \left[\frac{\Sigma X_{1i}^2}{n} - \left(\frac{\Sigma X_{1i}}{n}\right)^2 \right]$$

$$\Rightarrow \operatorname{cov}(Y, X_1) - \hat{\beta}_2 \operatorname{cov}(X_2, X_1) = \hat{\beta}_1 \operatorname{var}(X_1)$$

$$\Rightarrow \hat{\beta}_1 = \frac{cov(Y, X_1) - \hat{\beta}_2 \, cov(X_2, X_1)}{Var(X_1)}$$

Similarly, we can get,

$$\frac{\partial \Sigma e_i^2}{\partial \hat{\beta}_2} = 2 \sum \left(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) (-X_{2i}) = 0 .$$

$$\therefore \hat{\beta}_2 = \frac{cov(Y, X_2) - \hat{\beta}_1 \, cov(X_1, X_2)}{Var(X_2)}$$

$$\therefore \hat{\beta}_1 = \frac{cov(Y, X_1) - \left\{ \dfrac{cov(Y, X_2) - \hat{\beta}_1 cov(X_2, X_1)}{Var(X_2)} \right\} cov(X_1, X_2)}{Var(X_1)}$$

$$= \frac{cov(Y, X_1) \, Var(X_2) - cov(Y, X_2) \, cov(X_1, X_2) + \hat{\beta}_1 \{cov(X_1, X_2)\}^2}{Var(X_1) \, Var(X_2)}$$

$$\Rightarrow \hat{\beta}_1 \left[Var(X_1) \, Var(X_2) - cov(X_1, X_2) \, cov(X_1, X_2) \right] = cov(Y, X_1) \, Var(X_2) - cov(Y, X_2) \, cov(X_1, X_2)$$

$$\Rightarrow \hat{\beta}_1 = \frac{cov(Y, X_1)\, Var(X_2) - cov(X_1, X_2)\, cov(Y, X_2)}{Var(X_1)\, Var(X_2) - cov(X_1, X_2)\, cov(X_1, X_2)}$$

$$\therefore \hat{\beta}_2 = \frac{cov(Y, X_2)\, Var(X_2) - cov(X_1, X_2)\, cov(Y, X_1)}{Var(X_1)\, Var(X_2) - \{cov(X_1, X_2)\}^2}$$

The problem of heteroscedasticity

When error variance is not constant then there is the problem of heteroscedasticity.

$$\sigma_u^2 = \sigma_u^2 . X_i^2$$

where $X_i^2 > 0$ i.e. $f(X_i) = X_i^2$

Now error varience is variable i.e. different for different obs. When there is heteroscedasticity and we apply OLS then there arises some problem.

Consequences of hetero scedasticity

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta} = \frac{cov(X, Y)}{Var(X)}$$

$$= \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i (\alpha + \beta X_i + u_i - \alpha - \beta \bar{X})}{\Sigma x_i^2}$$

$\hat{u}_i = e_i$

u_i = random disturbance term.

21/02/2020

Variance of $\hat{b}$

$$Var(\hat{b}) = E[\hat{b} - E(\hat{b})]^2$$

Now, as $\hat{b}$ is unbiased estimator of b therefore $\Rightarrow$

$$\boxed{E(\hat{b}) = b}$$

$$= E[\hat{b} - b]^2$$

$$\hat{b} = \frac{cov(x, y)}{Var(x)}$$

$$= \frac{1/n \sum (x_i - \bar{x})(y_i - \bar{y})}{1/n \sum (x_i - \bar{x})^2}$$

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \frac{\sum x_i (\cancel{a} + b x_i + u_i - \cancel{a} - b\bar{x})}{\sum x_i^2}$$

$$= \frac{\sum x_i \{b (x_i - \bar{x}) + u_i\}}{\sum x_i^2}$$

$$= \frac{\sum x_i \{b x_i + u_i\}}{\sum x_i^2}$$

$$= \frac{b \sum x_i^2 + \sum x_i u_i}{\sum x_i^2}$$

$$\Rightarrow \hat{b} = b + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$E(\hat{b}) = E(b) + \frac{\sum x_i E(u_i)}{\sum x_i^2}$$

$$\boxed{E(\hat{b}) = b} \quad \text{as} \quad E(u_i) = 0.$$

$\therefore$ $\hat{b}$ is unbiased estimator

$$E[\hat{b} - b]^2 = Var(\hat{b})$$

$$\Rightarrow Var(\hat{b}) = E\left[\frac{\sum x_i u_i}{\sum x_i^2}\right]^2$$

$$= \frac{1}{(\sum x_i^2)^2} E\left[\sum x_i^2 u_i^2 + 2 \sum_{\substack{i<j \\ i \neq j}} \sum x_i u_i \, x_j u_j\right]$$

$$= \frac{1}{(\sum x_i^2)^2}\left[\sum x_i^2 E(u_i^2) + 2 \sum_{i<j} \sum x_i x_j \, E(u_i u_j)\right]$$

$$= \frac{\cancel{\sum x_i^2} \, \sigma_u^2}{(\sum x_i^2)^{\cancel{2}}} + 0 \quad \text{as} \quad E(u_i u_j) = 0$$

by the assumption

$$= \frac{\sigma_u^2}{\sum x_i^2}$$

$$= \frac{\sigma_u^2}{n \cdot Var(x)}$$

$$\sum x_i^2 = \sum (x_i - \bar{x})^2$$

$$Var(x) = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$\boxed{Var(\hat{b}) = \frac{\sigma_u^2}{n \cdot Var(x)}}$$

Now, $\hat{a} = \bar{Y} - \hat{b}\bar{X}$

$$= a + b\bar{X} + \cancel{\bar{u}} - \hat{b}\bar{X}$$

as mean of $u = 0$

$$\Rightarrow \boxed{\hat{a} = a - (\hat{b} - b) \cdot \bar{X}}$$

$$E(\hat{a}) = E(a) - \bar{X}\, E(\hat{b} - b)$$

$$= E(a) - \bar{X}\, E\left[\frac{\sum x_i u_i}{\sum x_i^2}\right]$$

$$= a - \bar{X} \frac{\sum x_i E(u_i)}{\sum x_i^2}$$

as $E(u_i) = 0$

$$\boxed{E(\hat{a}) = a}$$

$\cancel{E(\hat{a})}$

$$\text{Var}(\hat{a}) = E\left[\hat{a} - E(\hat{a})\right]^2$$

$$= E\left[\hat{a} - a\right]^2$$

$$= E\left[-\bar{x}\,(\hat{b} - b)\right]^2$$

$$= \bar{x}^2\, E(\hat{b} - b)^2$$

$$= \bar{x}^2 \frac{\sigma_u^2}{\Sigma x_i^2}$$

$$= \frac{\bar{x}^2\, \sigma_u^2}{n \cdot \text{Var}(x)}$$

Goodness of fit

$$r^2 = \left\{\frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}\right\}^2$$

$$= \frac{\text{cov}^2(x,y)}{\text{Var}(x)\,\text{Var}(y)}$$

$0 \le r^2 \le 1$ $\because$ $r^2 \to 1$ then the model is good fitted.

12/07/2020

The problem of heteroscedasticity

When error variance is not constant then there is the problem of heteroscedasticity.

$$\sigma_{u_i}^2 = \sigma_u^2 \cdot x_i^2$$

where $x_i^2 > 0$. i.e. $f(x_i) = x_i^2$

Now error variance is variable i.e. different for different obs.

When there is heteroscedasticity and we apply OLS then there arises some problem.

Consequences of heteroscedasticity

$\hat{u}_i = e_i$

u_i = random disturbance term

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta} = \frac{\operatorname{cov}(X, Y)}{\operatorname{Var}(X)}$$

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \frac{\sum x_i (\alpha + \beta X_i + u_i - \alpha - \beta \bar{X}_i)}{\sum x_i^2}$$

$$= \frac{\Sigma x_i \left(\beta x_i + u_i \right)}{\Sigma x_i^2}$$

$$= \beta + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$$

$$\therefore \quad \hat{\beta} = \beta + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$$

$$\Rightarrow \quad E\left(\hat{\beta}\right) = \beta \quad \text{as} \quad E(u_i) = 0$$

As well as,

$$E(\hat{\alpha}) = \alpha.$$

$\hat{\alpha}$ and $\hat{\beta}$ are unbiased.

Now,

$$Var\left(\hat{\beta}\right) = E\left[\hat{\beta} - E(\hat{\beta})\right]^2$$

$$= E\left[\hat{\beta} - \beta\right]^2$$

$$= E\left[\frac{\Sigma x_i u_i}{\Sigma x_i^2}\right]^2$$

$$= \frac{1}{(\Sigma x_i^2)^2} E\left[\Sigma x_i u_i\right]^2$$

$$= \frac{1}{(\Sigma x_i)^2} E\left[\Sigma x_i^2 u_i^2 + 2 \sum_i \sum_j x_i u_i x_j u_j\right]$$

$$= \frac{1}{(\Sigma x_i^2)^2} \left[\sigma \Sigma x_i^2 \, E(u_i) + 2 \sum_{i<j}\sum x_i x_j \, E(u_i u_j) \right]$$

$$= \frac{\cancel{\Sigma x_i^2}}{(\Sigma x_i^2)^{\cancel{2}}} E(u_i) \quad \because E(u_i u_j) = 0$$

$$= \frac{\sigma_u^2 \cdot x_i^2}{\Sigma x_i^2}$$

$$Var(u_i)_{\text{homoscedasticity}} = \frac{\sigma_u}{\Sigma x_i^2}$$

$$Var(u_i)_{\text{heteroscedasticity}} = \frac{\sigma_u^2 \cdot x_i^2}{\Sigma x_i^2}$$

$\therefore$ $Var(u_i)$ homoscedasticity $<$

$Var(u_i)$ heteroscedasticity

as $x_i^2 > 0$.

$\therefore$ Minimum variance property

is lost.

i.e. $\hat{\alpha}$ and $\hat{\beta}$ are not

remain BLUE

When Var$(\hat{\beta})$ ↑es then SE$(\hat{\beta})$ ↑es

then $t_{\hat{\beta}}$ ↓es i.e. rejected null

hypothesis i.e will be accepted.

There is error of 2nd type.

$$t_{\hat{\beta}} = \left(\frac{\hat{\beta}}{SE(\hat{\beta})\uparrow}\right)\downarrow$$