There are three main problems of applying OLS when the model is suffering from :—

(i) Heteroscedasticity or non-constant error variance

(ii) Multicollinearity

(iii) auto-correlation

Heteroscedasticity :— When variance of the error term is not constant for all observation then the problem is known as problem of heteroscedasticity

$$Var(u) = \sigma^2_u \rightarrow \text{Homoscedasticity}$$

$$Var(u) = \sigma^2_u \cdot w_i \rightarrow \text{Heteroscedasticity}$$

$w_i = f(x_i)$

$f'(x_i) > 0$

Consequence of Heteroscedasticity

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

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \qquad x_i = X_i - \bar{X}$$

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

$$= \frac{\sum x_i(\beta x_i + u_i)}{\sum x_i^2}$$

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

$$= \alpha + \beta X_i + u_i - \alpha - \beta\bar{X}$$

$$= \beta(X_i - \bar{X}) + u_i$$

$$= \beta x_i + u_i$$

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

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

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

$$= \beta \text{ as } E(x_i u_i) = 0 \ \forall i.$$

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

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

$$= \alpha - \bar{X}(\hat{\beta} - \beta).$$

$$\therefore E(\hat{\alpha}) = \alpha - \bar{X} E(\hat{\beta} - \beta).$$

$$= \alpha \text{ as } E(\hat{\beta} - \beta) = 0.$$

$\hat{\alpha}$ and $\hat{\beta}$ are still unbiased estimators though heteroscedasticity is present in the model.

Though heteroscedasticity is present, but $E(\hat{\alpha}) = \alpha$, $E(\hat{\beta}) = \beta$ } Estimated parameters are unbiased.

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

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

$$= E\left[\cancel{\beta} + \frac{\sum x_i u_i}{\sum x_i^2} - \cancel{\beta}\right]^2$$

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

$$= \frac{1}{(\sum x_i^2)^2}\left[\sum x_i^2 E(u_i^2) + 0\right].$$

$$= \frac{1}{(\sum x_i^2)^{\cancel{2}}} \cancel{\sum x_i^2} \cdot (\sigma_u^2 \cdot w_i)$$

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

Under homoscedasticity,

$$\text{Var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2}$$

Scanned with CamScanner

as $w_i = f(x_i)$ and is positive then

$$\frac{\sigma^2_u \cdot w_i}{\sum x_i^2} > \frac{\sigma^2_u}{\sum x_i^2}$$

$$\therefore Var(\hat{\beta})_{HE} > Var(\hat{\beta})_{H_0}$$

The minimum variance property of $\hat{\beta}$ is violated due to presence of heteroscedasticity. Now variance increases then $SE(\hat{\beta})$ also increases.

Therefore the parameters are not blue.

- We know that, t statistic for testing the significance of $\hat{\beta}$ and $\hat{\alpha}$ are,

$$t_{\hat{\beta}} = \frac{\hat{\beta}}{SE(\hat{\beta})} \quad \text{and} \quad t_{\hat{\alpha}} = \frac{\hat{\alpha}}{SE(\hat{\alpha})}$$

Due to presence of heteroscedasticity as $Var(\hat{\beta})$ and $Var(\hat{\alpha})$ ↑es. Therefore $SE(\hat{\beta})$ and $SE(\hat{\alpha})$ also ↑es and the result $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$ are underestimated. As a result, the false H_0 may be accepted ie. Type II error is committed.

Therefore total inference analyses is erroneous due to presence of heteroscedasticity and by applying OLS with this problem.

Dection of heteroscedasticity

Goldfeld - Quandt Test

Step-1

We have to order the X obs in decending order

Step-2

We select a center value as c and omit these from total n obs. ie $(n-c)$ obs are these

Step 3

we divide $(n-c)$ obs. by 2

$$\left(\frac{n-c}{2}\right)_1 \text{ and } \left(\frac{n-c}{2}\right)_2 .$$

Step 4

we have to calculate RSS for two group of obs. i.e RSS_1 as $\sum e_1^2$ and RSS_2 as $\sum e_2^2$

Step 5

Now we have done F test with $\left(\frac{n-c}{2} - k\right)$ d.f

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma_u^2$$

(Homoscedasticity)

H_1 : Any one inequality i.e (heteroscedasticity)

The F statistic is,

$$F = \frac{\sum e_1^2}{\sum e_2^2} \sim F\left(\frac{n-c}{2} - k\right)$$

Step 6

If we get, $F_{cal} < F_{tab}$ then H_0 is accepted and there is homoscedasticity

when, $F_{cal} > F_{tab}$ there is heteroscedasticity.

This test is easiest and popular test of detect the presence of heteroscedasticity

Generalised least Square Estimation (GLSE)

$$Y = \alpha + \beta X_i + u_i \quad \text{--- (1)}$$

$$Var(u_i) = \sigma_u^2 \cdot w_i$$

The model is transformed to,

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\alpha}{\sqrt{w_i}} + \frac{\beta X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}}$$

$$\Rightarrow Y_i^* = \alpha^* + \beta^* X_i^* + V_i \quad \text{--- (2)}.$$

$$Var(V_i) = E[V_i - E(V_i)]^2$$

$$= E[V_i^2] = E\left[\left(\frac{u_i}{\sqrt{w_i}}\right)^2\right]$$

PAGE NO.:
DATE:

$$= \frac{1}{w_i} E(u_i^2)$$

$$= \frac{1}{w_i}(\sigma_u^2 \cdot w_i) = \sigma_u^2$$

model (2) is free from heteroscedasticity and we apply OLS in this transformed model then OLS is renamed as generalized least Squares (GLS) estimation process

$$\alpha^* = \frac{\alpha}{\sqrt{w_i}}$$

$$\beta^* = \frac{\beta}{\sqrt{w_i}}$$

Multi collinearity

One of the important assumption of OLS is the two explanatory variables are not exactly perfectly related.

i.e. $r_{x_1 x_2} \neq \pm 1$

There are two types of multi co-linearity —

(i) exact multi collinearity or perfect
(ii) near exact multi collinearity.

In case of exact multicollinearity

$$X_1 = kX_2, \quad k > 0$$

$$Var(X_1) = k^2 \, Var(X_2)$$

$$Cov(X_1, X_2) = Cov(kX_2, X_2)$$

$$= k \cdot Var(X_2)$$

$$Cov(X_1, X_2) = \frac{1}{n}\sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$$

$$= \frac{1}{n}\sum k(X_2 - \bar{X}_2)(X_2 - \bar{X}_2)$$

$$= k \cdot \frac{1}{n}\sum(X_2 - \bar{X})^2$$

$$= k \, Var(X_2)$$

Multicollinearity arises in more than one explanatory variable model. We take three variable model as

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

By applying OLS

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

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

$$\hat{\beta}_2 = \frac{Cov(Y, X_2) \cdot Var(X_1) - Cov(Y, X_1)\, Cov(X_1, X_2)}{Var(X_1)\, Var(X_2) - Cov(X_1, X_2)\, Cov(X_1, X_2)}$$

Now when we apply OLS with the presence of perfect MC then we get,

$$Var(X_1) \cdot Var(X_2) - \{Cov(X_1, X_2)\}^2$$

$$= k^2 [Var(X_2)]^2 - k\, Var(X_2) \cdot k\, Var(X_2)$$

$$= k^2 [Var(X_2)]^2 - k^2 [Var(X_2)]^2$$

$$= 0$$

as denominator of $\hat{\beta}_1$ and $\hat{\beta}_2$ are zero, therefore we are unable to estimate the parameters with exact MC

In case of near exact MC

In case of near exact MC that is $r_{x_1 x_2}$ close to ± 1 then we estimate the parameters but they are not minimum variance i.e BLUE property is lost though they are unbiased

$$Var(\hat{\beta}_1) = \frac{\sigma^2_u}{\Sigma x_1^2 (1 - r_{12}^2)}$$

$$Var(\hat{\beta}_2) = \frac{\sigma^2_u}{\Sigma x_2^2 (1 - r_{12}^2)}$$

and the term, $\frac{1}{(1 - r_{12}^2)}$ is known as Variance Inflationary factor (VIF).

$$\therefore Var(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x_1^2} (VIF), \quad Var(\hat{\beta}_2) = \frac{\sigma^2_u}{\Sigma x_1^2} (VIF)$$

(1) When $r_{12} = 0$ then VIF = 1

(2) When $r_{12} = \pm 1$ then VIF = ∞

$\therefore$ VIF ranges from 1 to ∞

The value of VIF determines the existance of the model with application of OLS. When VIF $\uparrow$ es then Variances and SEs are over estimated and t statistic are underestimated.

Therefore estimation is erroneous and we are unable to get a good fitted model.

Q) 1) Here it is given that $\sum y_i^2 = 636$, $\sum x_i^2 = 664$

$\sum x_i y_i = 492$, $\sum y_i = 621$, $\sum x_i = 461$, $n = 16$

(i) Estimate the parameters in the model $Y_i = \alpha + \beta X_i + u_i$

(ii) Comment on the value of parameters

(iii) Estimate R^2 and comment on the validity of the model.

(iv) Test the hypothesis that $\hat{\beta} = 2$

2)

Output	210.2	210.1	211.5	208.9	207.4	205.2	[illegible]	[illegible]
No of workers (in 1000)	706.2	703.1	701.8	699.1	697.4	695.3	692.7	[illegible]

(i) estimate the linear production fn.

(ii) find average and marginal prod. of labour

(iii) estimate t ratios and test their significance.

3) Estimate the investment fn $I = \alpha x^{\beta} \cdot e^{u}$

the following sample is given

I	9	5.5	8.5	4	3.5	2.5	3	1.5	1.2	1.7	1.5
x	2	3	2	4	3	6	4	6	8	7	9

(i) estimate the investment fn. by OLS.

(ii) Test the significance of coefficient at 1% level of significance

(iii) Construct a 95% confidence interval for β

(iv) find the value of R^2

Scanned with CamScanner

4) A sample of 20 obs. on x and y is too be used for estimating the linear fn $y = \alpha + \beta x + u$. The first 10 obs. yeild the following results:

$\bar{X} = 15.3 \quad \bar{Y} = 160 \quad \sum x_i^2 = 78 \quad \sum y_i^2 = 45600$

$\sum x_i y_i = -7568$ The ten subsequent pairs of value of x and y yeilds:

$\bar{X} = 14.08 \quad \bar{Y} = 166 \quad \sum x_i^2 = 98.16 \quad \sum y_i^2 = 62400$

$\sum x_i y_i = -2308.8$

has the fn. changed over the two decades.

2005

Q. Define Autocorrelation. What are the sources of autocorrelation? (1+3) = 4

Correlation is defined as the measurement of degree of association between two variables. Autocorrelation is defined as the measurement of degree of association between the two values of a same variable. Auto correlation is special type of correlation. It is basically a time series problem. As it is a time series problem, it is also known as serial correlation i.e. the correlation between two series of a same variable.

Here, the two series of disturbance term are correlated. The actual meaning of autocorrelation is,

u_1	u_2
u_2	u_3
u_3	⋮
⋮	
u_{n-1}	u_{n-2}
u_n	u_{n-1}

$$E(u_t \cdot u_{t+s}) \neq 0 \;\; \forall \; t,\; s \neq 0$$

$$\Rightarrow \quad Cov(u_t \cdot u_{t+s}) \neq 0.$$

It means that u_t and u_{t+s} are pairwise autocorrelated i.e. co-variance between u_t and u_{t+s} is not equal to zero. This violates the classical assumption of applying OLS method.

Though autocorrelation is a time series problem but it does not confirm that it is not appear in cross section data. It also comes up in the presence of cross-section data. When we speak of autocorrelation in case of cross section data we measure the degree of association between two disturbance terms of two economic agents.

$$E(u_{it} \cdot u_{jt}) \neq 0.$$

$$\Rightarrow \quad Cov(u_{it} \cdot u_{jt}) \neq 0.$$

At the same point of time, the two disturbance term of the two economic agent is correlated. It is the autocorrelation at the same point of time. This type of autocorrelation is known as contemporaneous auto correlation.

Sources of Autocorrelation

The presence of autocorrelation is very common in emperical research work. Autocorrelation comes up —

(1) <u>Due to mis-specification of a model.</u> The specified model is,

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

But the actual model is:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + u_t$$

We are neglecting a term X_t^2 in the specified model. Then,

$$Cov(u_t \cdot \beta_2 X_t^2 + u_t) = \sigma_u^2$$

$$\text{i.e.} \quad Cov(u_t \cdot \beta_2 X_t^2 + u_t) \neq 0.$$

That means in the specified model autocorrelation is present.

The specified u_t and the actual u_t (i.e. $\beta_2 x_t^2 + u_t$) is correlated, it is not equal to zero. Then we can not estimate the model.

(2) Omission of relevant variables may lead to misspecification.

we specify the model.

$$D_t = \beta_0 + \beta_1 P_t + u_t$$

But the actual model is,

$$D_t = \beta_0 + \beta_1 P_t + \beta_2 Y_t + u_t$$

$$\therefore \quad cov\,(u_t, \beta_2 Y_t + u_t) \neq 0$$

Thus, the omission of certain variables may lead to the appearance of autocorrelation.

(3) Due to some measurement error autocorrelation may comes up. In generally we can write the consumption function as.

$$C_t = \alpha + \beta Y_t + u_t$$

C_t = consumption in period t

Y_t = disposable income in period t

But the actual model is,

$$C_t = \alpha + \beta\,(Y_t + V_t) + u_t$$

$$= \alpha + \beta Y_t + \beta V_t + u_t$$

V_t = measurement errors. Y_t = income in period t

$$cov\,(u_t, \beta V_t + u_t) \neq 0$$

$$\Rightarrow \beta\, cov\,(u_t, V_t) + \sigma_u^2 \neq 0$$

If we introduce proxy variable then measurement error comes up.

(4) In case of cob-web model

$$q_t^d = f\,(p_t, u_t) = a_0 + a_1 p_t + u_{1t} \rightarrow \text{demand function}$$

$$q_t^s = f\,(p_{t-1}, u_t) = b_0 + b_1 p_{t-1} + u_{2t} \rightarrow \text{Supply function}$$

$$q_t^d = q_t^s$$

$$\Rightarrow a_0 + a_1 p_t + u_{1t} = b_0 + b_1 p_{t-1} + u_{2t}$$

$$\Rightarrow p_t = \frac{b_0 - a_0}{a_1} + \frac{b_1}{a_1} p_{t-1} + \frac{u_{2t} - u_{1t}}{a_1}$$

$$\Rightarrow p_t = \frac{b_0 - a_0}{a_1} + \frac{b_1}{a_1} p_{t-1} + V_t$$

$$\Rightarrow p_t = \alpha_0 + \alpha_1 p_{t-1} + V_t.$$

$$p_{t-1} = \alpha_0 + \alpha_1 p_{t-2} + V_{t-1}.$$

In P_{t-1} we can find a error V_{t-1}. Therefore,

$$cov(V_t, V_{t-1}) \neq 0.$$

Thus, if we use cob-web model then there is no gurantee that autocorrelation is not come up.

(5) Due to averaging technique.

Autocorrelation may come up due to averaging technique.

1971
1972
⋮
1999
2000

Then we fit a trend equation.

$$\ln Y_t = \alpha + \beta t$$

$x_1, x_2, x_3, \ldots, x_{20}$ → $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$

Here, $\bar{x}_1, \bar{x}_2, \bar{x}_3$ are also interlinked or correlated. Therefore in case of moving average method autocorrelation also comes up.

When we use averaging technique, moving average technique, interpolation technique, extrapolation technique then the different values of a same variable is correlated i.e. autocorrelation appears.