# Problems of Econometrics

There are three main problems due to their importance in terms of their probabilities and effect on the results

1. **Multicollinearity Problem.**
2. **Autocorrelation Problem.**
3. **Hetroscedasticity Problem.**

# Multicollinearity Problem

You get the problem of Multicollinearity when correlate two or more explanatory variables with very strong linear relationship so that it becomes difficult to separate the effect of each variable on the dependent variable.

**Multicollinearity** term is composed of three sections:-

**Multi** → multiple

**Co** → common correlation

**Linearity** → linear

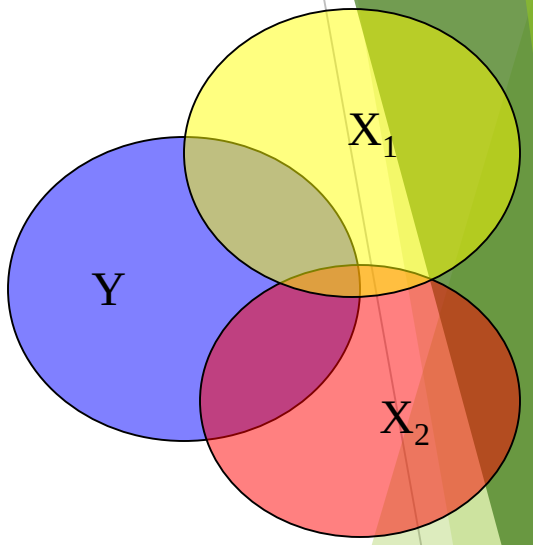
The assumption of multicollinearity is there is no perfect or semi-perfect linear relationship between any of the explanatory variables. As well as it should be the number of parameters to be estimated less than the sample size.

$$\text{Rank}(X) = k + 1 < n$$

# Illustrating Multicollinearity

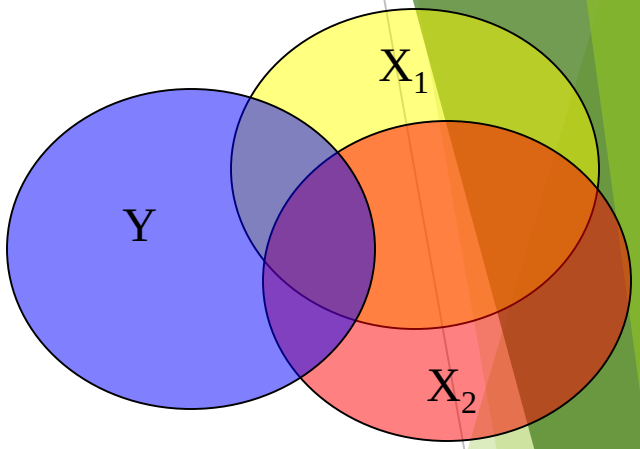
• When correlation among X's is low, OLS has lots of information to estimate $\beta$

• This gives us confidence in our estimates of $\beta$



•When correlation among X's is high, OLS has very little information to estimate $\beta$.

•This makes us relatively uncertain about our estimate of $\beta$.



## Reasons or Sources of Multicollinearity Problem

**1.** Could change some of the explanatory variables together because of lack of data collection from a broad base, or because of the nature of the variables. When the value of one of the explanatory variables depends on the value of one or more of the explanatory variables, i.e., one of explanatory variable can be written as linear combination for the other variables.

**Ex1:**// perfect collinearity

| | | | | | |
|---|---|---|---|---|---|
| $X_1$ : | 10 | 15 | 18 | 24 | 9 |
| $X_2$ : | 50 | 75 | 90 | 120 | 45 |

Then $X_1$ & $X_2$ are perfectly correlated because $X_2 = 5X_1$.

**Ex2:**// Family income ($X_1$)= husband's income ($X_2$) + wife's income ($X_3$)

$$X_1 = X_2 + X_3$$

Then $X_1$ , $X_2$ & $X_3$ are perfectly correlated.

**2.** Possible to share all explanatory variables at a particular time trend (general trend) [which the variables move together], or that one of the explanatory variables may be its value lag time (lag-variable) from the other that goes in the direction of another time. In this case arises the problem of multicollinearity between the explanatory variables.

**For example//**

Let; $X_t$ : variable of time series

$X_{t-1}$ : lag variable

Then $X_t$ and $X_{t-1}$ are correlated.

- Lag – variable: means that variable ($X_{t-1}$) lags than the other variable ($X_t$) at a particular time interval.

**3.** Misspecification of the model :

For example: when you add polynomial terms to the model, and especially if the range of the variable X is small, in this case the problem multicollinearity may occur .

**Ex:**// $Y_i = \beta_o + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + U_i$

**4.** when the number of explanatory variables be greater than the number of observations (sample size), [ $k > n$].

# Consequences of Multicollinearity

- In case perfect correlation ($r_{Xi,\,Xj} = \pm 1$)

1- (X'X) singular matrix $\rightarrow$ $|X'X| = 0$

$\Rightarrow$ $(X'X)^{-1} \rightarrow \infty$

2- Cannot estimation the values of parameters by $\hat{\beta} = (X'X)^{-1} X'Y \rightarrow \infty$

3- Cannot estimation the Standard errors for estimators $V\text{-}Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} \rightarrow \infty$



4- Cannot procedure the statistical tests [ $t$, $F$ ].

- In case semi-perfect correlation $r_{X_i X_j} \neq \pm 1 \rightarrow 1$

1- $|X'X| \neq 0 \rightarrow 0$ (Very small)

2- The value of estimated parameters $(\hat{\beta})$ for model be inaccurate and very large.

3- Standard errors for estimators will be very large $\sigma^2 (X'X)^{-1} \rightarrow \infty$

4- The confidence intervals become large and wrong.

# 5- Wrong conclusions

a. Wrong explanation for effects some variables on dependent variable result for t-test, (i.e., insignificant but in real significant).

b. The result of *t*-test insignificant while the value of $R^2$ very large and the result of F-test significant.

c. Wrong signs for some estimators.

**6- The estimates of parameters not be stable or not very robust,** i.e., estimates become very sensitive (speed impact) to addition some data or simple changes in data or analysis a part of sample may lead to large changes in the values of that estimates.

# Methods for Detecting Multicollinearity

**1- The determinant of matrix (X'X) method**

**a. If |X'X| = 0 then we have perfect multicollinearity.**

**b. If |X'X| → 0 then we have higher multicollinearity.**

**2- The result of *t*-test insignificant while the value of $R^2$ very large and the result of *F*-test significant.**

**3- Scatter-plot for explanatory variables and correlation matrix for variables.**

# 4- Variance Inflation Factor (VIF)

$$VIF_j = \frac{1}{1 - R_j^2} \quad , \quad j = 1,\ldots,k\,(No.\ of\ \text{var}iables)$$

Where compute $R_j^2$ by regress each explanatory variable ($X_j$) on the rest of explanatory variables.

a. If VIF > 10 , then very high multicollinearity

b. If $VIF \leq 10 \rightarrow 1$ ,then no multicollinearity

## 5- Clien Method

Compute multiple correlation coefficient which equals to square root of multiple determination coefficient $R^2_{Y.X_1,\ldots,X_k}$ for general linear model, and compare it with simple correlation coefficients between explanatory variables.

If $R_{Y.X_1,\ldots,X_k}$ for model is greater than simple correlation coefficients that means no multicollinearity, and vice versa.

**Ex:** Test multicollinearity problem using Clien method , if and

$$R^2_{Y.X_1,\ldots,X_k} = 0.932$$

**Sol.**

$$R = \begin{pmatrix} 1 & .956 & .919 \\ & 1 & .901 \\ & & 1 \end{pmatrix}$$

$$R_{Y.X_1,\ldots,X_k} = \sqrt{R^2_{Y.X_1,\ldots,X_k}} = \sqrt{0.932} = 0.966$$

Since multiple correlation coefficient is greater than simple correlation coefficients that means no multicollinearity problem.

# 6- Farrar-Glaubber Test

**1. Chi-Square Test ($\chi^2$- test):**
To detect the presence of the multi collinearity problem in the function included for several explanatory variables.

**2. F- Test:** To determine correlated variables linearly.

**3. *t*-Test:** To determine the variables that cause the multi collinearity.

# First: Chi-Square Test ($\chi^2$- test)

- The hypotheses of test

$H_o$: The explanatory variables $X_j$'s are orthogonal (uncorrelated : Independent)

$H_1$: The explanatory variables $X_j$'s are not orthogonal (correlated : dependent)

- Calculate the value of test statistic (Calculated value for $\chi^2$ : Cal $\chi^2$), as follows

$$Cal\ \chi^2 = \left[\frac{k}{3} - n + \frac{11}{6}\right] \cdot \ln|R|$$

where; $n$: sample size.

$k$: No. of explanatory variables.

ln |R|: The natural logarithm for the value of determination of simple correlation coefficients matrix between explanatory variables.

$$|R| = \begin{vmatrix} 1 & r_{12} & r_{13} & .. & r_{1k} \\ r_{21} & 1 & r_{23} & .. & r_{2k} \\ .. & .. & .. & .. & .. \\ r_{k1} & r_{k2} & r_{k3} & .. & 1 \end{vmatrix}$$

$r_{X_i, X_j}$ : simple correlation coefficient between $X_i$, $X_j$

$$r_{ij} = r_{X_i, X_j} = \frac{\sum x_i x_j}{\sqrt{\sum x_i^2}\sqrt{\sum x_j^2}} \quad , \quad -1 \le r_{ij} \le +1$$

$$\sum x_i x_j = \sum X_i X_j - n \overline{X}_i \overline{X}_j$$

$$\sum x_i^2 = \sum X_i^2 - n \overline{X}_i^2 \quad , \quad \sum x_j^2 = \sum X_j^2 - n \overline{X}_j^2$$

- Under particular significant level ($\alpha$), we find the tabulated value (Tab $\chi^2$), as follows

$$Tab\ \chi^2_{(\alpha,\ \frac{k(k-1)}{2})},\ where\ \frac{k(k-1)}{2}\ is\ \text{degrees of freedom}$$

- We compare the value of (Cal $\chi^2$) with (Tab $\chi^2$)
- if Cal $\chi^2 \geq$ Tab $\chi^2$ , we rejected $H_o$
- if Cal $\chi^2 <$ Tab $\chi^2$ , we not rejected $H_o$

# Second: F- Test

After testing the presence of multicollinearity problem according to $\chi^2$ – test requires determine which variable from the explanatory variables correlated linearly, where lead to occur this problem, we perform such diagnosis by using F-test, as follows

- The hypotheses of test

$$H_o : R^2_{j.\,2,3,\ldots,k} = 0$$

$$H_1 : R^2_{j.\,2,3,\ldots,k} > 0$$

- we calculate the general form for test statistic (calculated value) for this test and for each variable, as follow;

$$Cal\ F_j = \frac{(R^2_{j.2,3,\ldots,k}) / (k-1)}{(1 - R^2_{j.2,3,\ldots,k}) / (n-k)}$$

Where;

$R_{j.2,3,\ldots,k}$: Multiple correlation coefficient between $X_j$ and the rest of studied explanatory variables.

➢ We compare the value of Cal $F_j$ with Tab $F$ with degrees of freedom $(k-1)$, $(n-k)$ and under particular significant level $(\alpha)$.

- if Cal $F_j \geq$ Tab $F$ , we rejected $H_o$

This means that the variable $X_j$ correlated linearly with the rest explanatory variables.

- if Cal $Fj$ < Tab $F$ , we not rejected Ho

This means that the variable Xj not correlated linearly with the rest explanatory variables.

❖Apply this test for each explanatory variable to determine all explanatory variables that correlated or not correlated with the rest of explanatory variables.

# Third: *t* – Test

After determination the correlated variables by F – test, we use *t* – test for all possible pairs of r.v's .

- The hypotheses test

$$H_o : r_{ij.1,2,3,\ldots,k} = 0$$

$$H_1 : r_{ij.1,2,3,\ldots,k} \neq 0$$

- we calculate the test statistic value (calculated value) for this test $(Cal\ t_{ij.\,1,2,3,\ldots,k})$ with respect to two explanatory variables $X_i$ , $X_j$ , as follows

$$Cal\ t_{ij.\,1,2,3,\ldots,k} = \frac{(r_{ij.\,1,2,3,\ldots,k})\sqrt{n-k}}{\sqrt{1-r^2_{ij.\,1,2,3,\ldots,k}}}$$

Where;

$r_{ij.\,1,2,\ldots,k}$: The partial correlation coefficient between $X_i$ , $X_j$ with hold the rest of explanatory variables constant.

**Ex:** If we have three explanatory variables $X_i$, $X_j$, $X_k$

- The multiple correlation coefficient between $X_i$ and $X_j$ , $X_k$ is

$$R_{i.jk} = \sqrt{\frac{r_{ij}^2 + r_{ik}^2 - 2 r_{ij} r_{ik} r_{jk}}{1 - r_{jk}^2}}$$

- The partial correlation coefficient between $X_i$ , $X_j$ with hold $X_k$ constant is

$$r_{ij.k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{1 - r_{ik}^2}\ \sqrt{1 - r_{jk}^2}}$$

➢We compare the value of $(Cal\ t_{i\,j.\,1,2,3,\ldots,k})$ with (Tab.$t$), with degree of freedom $(n-k)$ and under particular significant level $(\alpha)$, (Tab $t_{((\alpha/2),\,(n-k))}$).

• If Cal. $t_{i\,j.1,2,\ldots,k} \geq$ Tab $t_{((\alpha/2),\,(n-k))}$, we reject $H_o$, This means that The partial correlation coefficient between $X_i$ , $X_j$ with hold the rest of explanatory variables constant is significant, which both are responsible for the problem multicollinearity in model.

If Cal. $t_{i\,j.1,2,..,k}$ < Tab $t_{((\alpha/2),\,(n-k))}$, we not rejected $H_o$ , This means that The partial correlation coefficient between $X_i$ , $X_j$ with hold the rest of explanatory variables constant is not significant, which both are not responsible for the problem multicollinearity in model.

**Example:** From the following data for three explanatory variables, test multicollinearity problem between explanatory variables ($X_1$, $X_2$, $X_3$) under significant level ($\alpha = 0.05$), and using **Farrar – Glaubber test.**

$$n = 23 \;,\quad \sum X_1 = 89.9 \;,\quad \sum X_2 = 416115 \;,\quad \sum X_3 = 366000$$

$$\sum X_1^2 = 395.21, \quad \sum X_2^2 = 9{,}530{,}780{,}489 \;,\quad \sum X_3^2 = 8{,}221{,}026{,}400$$

$$\sum X_1 X_2 = 14391509 \;,\; \sum X_1 X_3 = 12131426 \;,\quad \sum X_2 X_3 = 8{,}625{,}226{,}000$$

$$Tab\, t_{(0.025,\,20)} = 2.086 \quad,\quad Tab\, \chi^2_{(0.05,\,3)} = 7.816 \;,\; Tab\, F_{(0.05\,,\,2\,,20)} = 3.49$$

# Sol: Farrar-Glaubber Test

## 1- Chi-Square Test ($\chi^2$- test)

$$r_{ij} = r_{X_i, X_j} = \frac{\sum x_i x_j}{\sqrt{\sum x_i^2}\sqrt{\sum x_j^2}} \quad , \quad -1 \le r_{ij} \le +1$$

$$\sum x_i x_j = \sum X_i X_j - n\overline{X}_i \overline{X}_j$$

$$\sum x_i^2 = \sum X_i^2 - n\overline{X}_i^2 \quad , \quad \sum x_j^2 = \sum X_j^2 - n\overline{X}_j^2$$

# Sol: Farrar-Glaubber Test

$H_o$: The explanatory variables $X_j$'s are orthogonal
(uncorrelated : Independent)
$H_1$: The explanatory variables $X_j$'s are not orthogonal
(correlated : dependent)

➢ Calculate the value of test statistic (Calculated value for $\chi^2$ : Cal $\chi^2$), as follows

$$Cal\ \chi^2 = \left[\frac{k}{3} - n + \frac{11}{6}\right] \cdot \ln|R|$$

where; $n$: sample size.
$k$: No. of explanatory variables.

$$r_{xi,xj} = \frac{\sum xi, xj}{\sqrt{\sum xi^2}\sqrt{\sum xj^2}} \quad , r_{x1,x2} = \frac{\sum x1, x2}{\sqrt{\sum x1^2}\sqrt{\sum x2^2}}$$

$$r_{x1,x3} = \frac{\sum x1, x3}{\sqrt{\sum x1^2}\sqrt{\sum x3^2}} \quad r_{x2,x3} = \frac{\sum x2, x3}{\sqrt{\sum x2^2}\sqrt{\sum x3^2}}$$

$$R = \begin{pmatrix} 1 & -0.63 & -0.67 \\ -0.63 & 1 & 0.92 \\ -0.67 & 0.92 & 1 \end{pmatrix}$$

$H_0$: $X_1, X_2, X_3$ *are othogonal (Uncorrelated)*
$H_1$: $X_1, X_2, X_3$ *are not othogonal (correlated)*

$$Cal\ \chi^2 = \left[\frac{k}{3} - n + \frac{11}{6}\right] \cdot \ln|R|$$

$$|R| = \begin{vmatrix} 1 & -0.63 & -0.67 \\ -0.63 & 1 & 0.92 \\ -0.67 & 0.92 & 1 \end{vmatrix} = 0.084464$$

$$Cal\ \chi^2 = \left[\frac{3}{3} - 23 + \frac{11}{6}\right] \cdot (-2.4714298) \quad = 49.84$$

$Cal\ \chi^2 > Tab\ \chi^2$ , werejected $H_0$ then the explanatory variables $X_1, X_2, X_3$ are not orthogonal (correlated) , we have mullticollinearity problem among the explanatory variables

2-F-test

- For X1

$$H_0 : R^2_{1.23} = 0$$

$$H_1 : R^2_{1.23} > 0$$

$$Cal\ F_j = \frac{(R^2_{j.2,3,\ldots,k}) \,/\, (k-1)}{(1 - R^2_{j.2,3,\ldots,k}) \,/\, (n-k)}$$

$$R^2{}_{1.23} = \frac{(r_{12})^2 + (r_{13})^2 - 2(r_{12} r_{13} r_{23})}{1 - (r_{23})^2}$$

$$R^2{}_{1.23} = \frac{(-0.63)^2 + (-0.67)^2 - 2(-0.63) * (-0.67) * (0.92)}{1 - (0.92)^2} = 0.45$$

$$Cal\ F1 = \frac{(R^2_{1.2,3}) \,/\, (k-1)}{(1 - R^2_{1.2,3}) \,/\, (n-k)}$$

$$Cal\ F1 = \frac{(0.45) \,/\, (2)}{(1 - 0.45) \,/\, (20)} = 8.182$$

Cal F > Tab F we rejected Ho then

This means that the variable $X_1$ correlated linearly with the rest explanatory variables.

B// for $(X_2)$

$H_0 : R^2_{2.13} = 0$

$H_1 : R^2_{2.13} > 0$

$$R^2_{2.13} = \frac{r^2_{21} + r^2_{23} - 2r_{12}r_{23}r_{13}}{1 - r^2_{13}} = \frac{(-0.63)^2 + (0.92)^2 - 2(-0.63)(-0.67)(0.92)}{1 - (-0.67)^2} = .8473$$

$$\text{Cal } F_{X_2} = \frac{R^2_{2.13}/(k-1)}{(1 - R^2_{2.13})/(n-k)} = \frac{0.8473/2}{(1 - 0.8473)/20} = \frac{0.4236}{0.00763} = 55.5$$

$\text{Cal } F_{X_2} = 55.5$

$\text{Tab } F_{(0.05,2,20)} = 3.49$

$55.5 > 3.49$ , $\text{Cal } F_{X_2} > \text{Tab } F$ we reject $H_0$

This means that the variable $(X_2)$ correlated linearly with other explanatory variables $(X_1, X_3)$ and therefore it is the source of the problem of multicollinearity.

C// for $(X_3)$

$H_0 : R^2_{3.12} = 0$

$H_1 : R^2_{3.12} > 0$

$$R^2_{3.12} = \frac{r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2_{12}} = \frac{(-0.67)^2 + (0.92)^2 - 2(-0.63)(-0.67)(0.92)}{1 - (-0.63)^2} = 0.8605$$

$$\text{Cal } F_{X_3} = \frac{R^2_{3.12}/(k-1)}{(1 - R^2_{3.12})/(n-k)} = \frac{0.8605/2}{(1 - 0.8605)/20} = \frac{0.4302}{0.00697} = 61.72$$

$\text{Cal } F_{X_3} = 61.72$

$\text{Tab } F_{(0.05,2,20)} = 3.49$

$61.72 > 3.49$ , $\text{Cal } F_{X_3} > \text{Tab } F$ we reject $H_0$

This means that the variable $(X_3)$ correlated linearly with other explanatory variables $(X_1, X_2)$ and therefore it is the source of the problem of multicollinearity too.

3- $t-test$: A // for $(X_1, X_2)$:

$H_0 : r_{12.3} = 0$

$H_1 : r_{12.3} \neq 0$

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{(-0.63)-(-0.67)(0.92)}{\sqrt{(1-(-0.67)^2)}\sqrt{(1-(0.92)^2)}} = \frac{-0.0136}{(0.7424)(0.3919)} = -0.047$$

$$\text{Cal } t_{12} = \frac{r_{12.3}\sqrt{n-k}}{\sqrt{1-r_{12.3}^2}} = \frac{-0.047\sqrt{20}}{\sqrt{1-(-0.047)^2}} = \frac{-0.2102}{0.998895} = -0.21043$$

$$\text{Tab } t_{(0.025,20)} = \pm 2.086 \quad , \quad -0.21043 < -2.086 \Rightarrow \text{Cal } t_{12} < \text{Tab}t \text{ we not reject } H_0$$

This means that the partial correlation between two explanatory variables $(X_1, X_2)$ with hold $(X_3)$ constant not significant, which both are responsible from the problem multicollinearity in model.

B // for $(X_1, X_3)$:

$$H_0 : r_{13.2} = 0$$

$$H_1 : r_{13.2} \neq 0$$

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} = \frac{(-0.67) - (-0.63)(0.92)}{\sqrt{(1 - (-0.63)^2)}\sqrt{(1 - (0.92)^2)}} = \frac{-0.0904}{(0.7766)(0.3919)} = -0.297$$

$$\text{Cal } t_{13} = \frac{r_{13.2}\sqrt{n-k}}{\sqrt{1 - r_{13.2}^2}} = \frac{-0.297\sqrt{20}}{\sqrt{1 - (-0.297)^2}} = \frac{-1.3282}{0.9549} = -1.39093$$

$$\text{Tab } t_{(0.025,20)} = \pm 2.086 \quad , \quad -1.39093 < -2.086 \Rightarrow \text{Cal } t_{12} < \text{Tab } t \text{ we not reject } H_0$$

This means that the partial correlation between two explanatory variables $(X_1, X_3)$ with hold $(X_2)$ constant not significant, which both are not responsible from the problem multicollinearity in model too.

C // for $(X_2, X_3)$:

$$H_0 : r_{23.1} = 0$$

$$H_1 : r_{23.1} \neq 0$$

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}} = \frac{(0.92) - (-0.63)(-0.67)}{\sqrt{(1 - (-0.63)^2)}\sqrt{(1 - (-0.67)^2)}} = \frac{0.4979}{(0.7766)(0.7424)} = 0.8636$$

$$\text{Cal } t_{23} = \frac{r_{23.1}\sqrt{n-k}}{\sqrt{1 - r_{23.1}^2}} = \frac{0.8636\sqrt{20}}{\sqrt{1 - (0.8636)^2}} = \frac{3.8621}{0.5042} = 7.6599$$

$$\text{Tab } t_{(0.025, 20)} = \pm 2.086 \quad , \quad 7.6599 > 2.086 \Rightarrow \text{Cal } t_{12} > \text{Tab } t \text{ we reject } H_0$$

This means that the partial correlation between two explanatory variables $(X_2, X_3)$ with hold $(X_1)$ constant significant, which both are responsible from the problem multicollinearity in model.

Example H.W// In the following data we are measuring the quantity y for several values of $X_1$ , $X_2$ and $X_3$ (explonatory variables) we will use the following tables of values:

| Y | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|
| 0.19 | 0.5 | 0.4 | 0.3 |
| 0.28 | 0.8 | 0.6 | 0.2 |
| 0.30 | 0.9 | 0.7 | 1.1 |
| 0.25 | 1.1 | 1.2 | 2.1 |
| 0.29 | 1.3 | 1.4 | 0.8 |
| 0.28 | 1.4 | 1.7 | 0.4 |

$$\sum_{i=1}^{6} y_i = 1.59 \ , \ \sum_{i=1}^{6} X_1 = 6 \ , \ \sum_{i=1}^{6} X_2 = 6 \ , \ \sum_{i=1}^{6} X_3 = 4.9 \ , \ \sum_{i=1}^{6} X_1^2 = 6.56 , \ \sum_{i=1}^{6} X_2^2 = 7.3 \ , \ \sum_{i=1}^{6} X_3^2 = 6.55$$

$$\sum_{i=1}^{6} X_1 Y_i = 1.633 \ , \ \sum_{i=1}^{6} X_2 Y_i = 1.636 \ , \ \sum_{i=1}^{6} X_3 Y_i = 1.312 \ , \ \sum_{i=1}^{6} Y_i^2 = 0.4295 \ , \ \sum_{i=1}^{6} X_1 X_2 = 6.83 \ ,$$

$$\sum_{i=1}^{6} X_1 X_3 = 5.21 \ , \ \sum_{i=1}^{6} X_2 X_3 = 5.33$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 = 0.066 + 0.462 X_1 - 0.257 X_2 - 0.008 X_3$$

$$\text{Tab } \chi^2_{(0.05,3)} = 7.815 \quad , \quad \text{Tab } F_{(0.05,2,3)} = 9.55 \quad , \quad \text{Tab } t_{(0.025,3)} = \pm 3.182$$

From these information above test multicollinearity problem between explanatory variables $(X_1, X_2, X_3)$ under significant level $(\alpha = 0.05)$ and using **Farrar – Glaubber test.**

# Remedial Tools for Multicollinearity

1- Collecting Additional Data (increase of sample size to get more information), or collecting new data.

2- Using prior given information for parameters and combine it in model.

**Ex:** Let we know that $\beta_2 = 0.1\beta_1$ (prior information)

Then; $Y_i = \beta_o + \beta_1 X_1 + \beta_2 X_2 + e_i$

$= \beta_o + \beta_1 X_1 + 0.1\beta_1 X_2 + e_i$

$= \beta_o + \beta_1 X + e_i$ , *where* $X = X_1 + 0.1X_2$

First we estimate the value of $\beta_1$ and then we estimate the value of $\beta_2$ from previous relation.

3- Transformation of Functional Relation.

**Ex:** suppose we have time series data

$$Y_t = \beta_o + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t$$

Then; $Y_{t-1} = \beta_o + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + e_{t-1}$

Now; $Y_t - Y_{t-1} = \beta_1 (X_{1t} - X_{1t-1}) + \beta_2 (X_{2t} - X_{2t-1}) + (e_t - e_{t-1})$

$$Y_t^* = \beta_1 Z_t + \beta_2 T_t + V_t$$

In this case the differences of variables not correlated with each other, i.e., $Z_t$ and $T_t$ are uncorrelated. But in this case may create another problem that is the new error term ($V_t$) May be not satisfy their assumptions, it is the errors should be uncorrelated.

4- Review of specification of the Model.

5-Omit a Variable. Multicollinearity can be reduced by removing one of the highly correlated variables.

6- Centering the Data.
Use this procedure to reduction the trace of Multicollinearity on results.

$$Centering\ the\ Data\ = x_i = X_i - \overline{X}$$

7- Another Methods.

- Ridge Regression. This technique introduces a small amount of bias into the coefficients to reduce their variance.
- Principal Components Regression.
- Factor Analysis.

# Autocorrelation Problem

This problem occur in case presence the relation (correlation) among random terms (residuals), i.e.,

$$Cov(u_i, u_j) = E(u_i, u_j) \neq 0 \quad , \quad \forall\, i \neq j$$

This means the random variable(error term) which occur through particular time ($u_t$) correlated with random variable which preceding it ($u_{t-1}$)or which next it ($u_{t+1}$).

Meaning that the Pearson's $r$ between the residuals from OLS and the same residuals lagged on period is non-zero.

Simplest type of autocorrelation is call 1st order autocorrelation, According to it the random error for each time depend on the random error for preceding time for it linearly, can express by the following form;

$u_t = \rho u_{t-1} + e_t$ coefficient between each random error ($u_t$) in time $t$ and random error which preceding it $u_{t-1}$ or which (First order autocorrelation)

The residuals are related to their preceding values.

Where;

$\rho$ : Simple autocorrelation (ut)($u_{t+1}$), $[-1 \leq \rho \leq 1]$.

$e_t$ : error term for linear model ($u_t = \rho u_{t-1} + e_t$) (non autocorrelated white noise), satisfies all previous assumptions of the random term

$$e_t \sim N(0, \sigma_e^2)$$

$$Cov(e_i, e_j) = 0 \text{ , for all } i \neq j$$

$$Cov(e_i, X_i) = 0$$

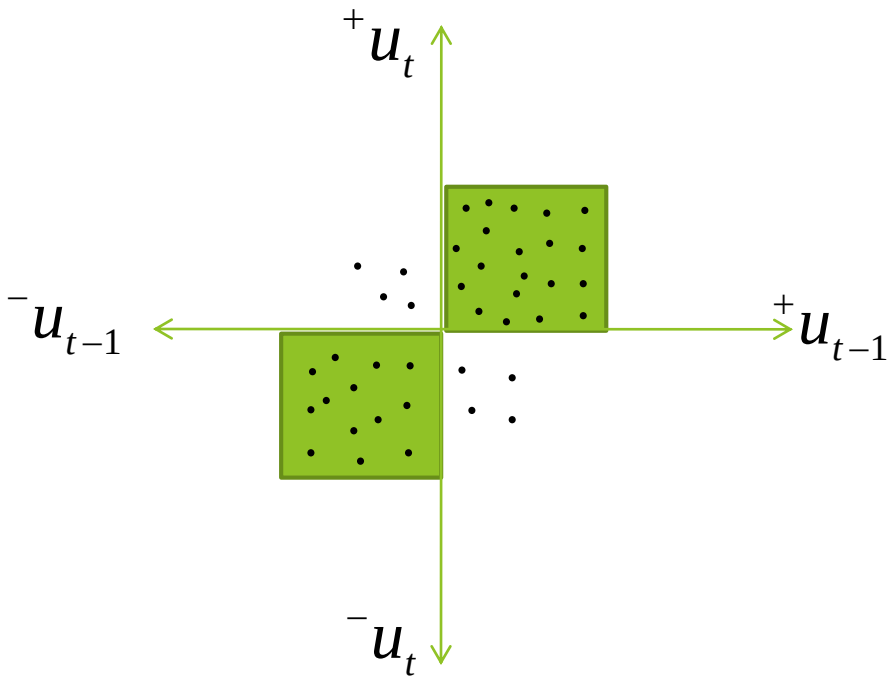
Autocorrelation is correlation between $(u_i, u_j)$ returns to the relation not between two or more different variables but between sequential values for the same variable.

$$Y_t = \beta_o + \beta_1 X_t + u_t$$
$$u_t = \rho u_{t-1} + e_t$$

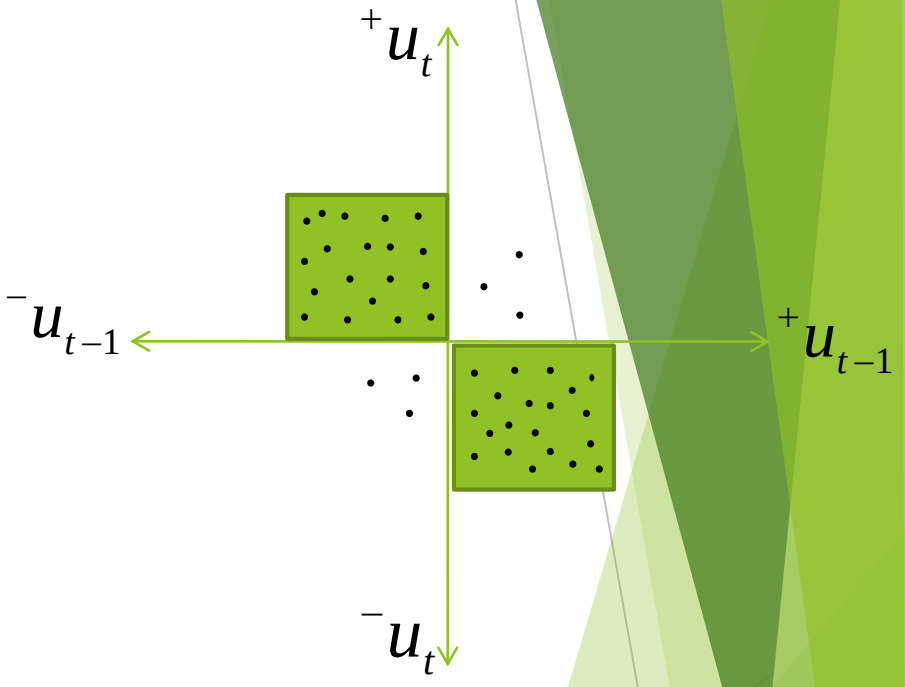
- Autocorrelation problem arise in most studies and researches which depend on time series data.

# Determination the type of Autocorrelation Graphically

## First Method;



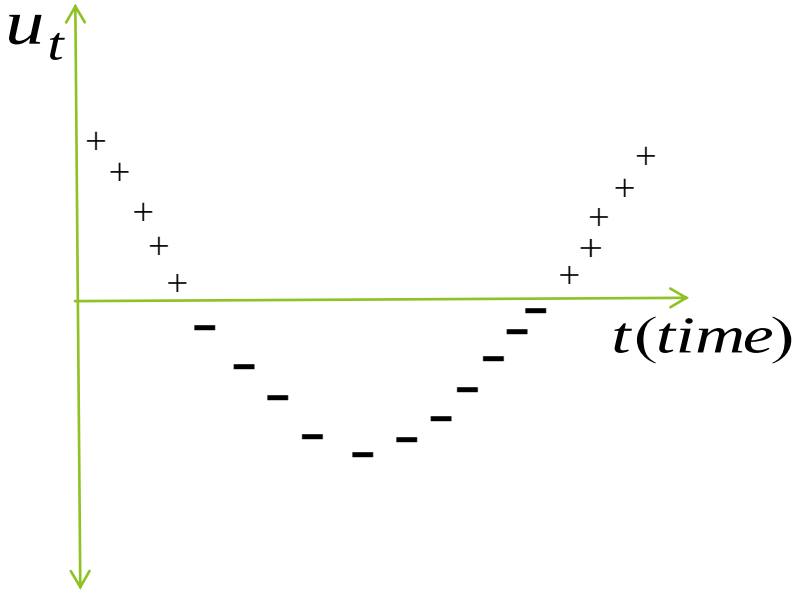

“Positive Autocorrelation”

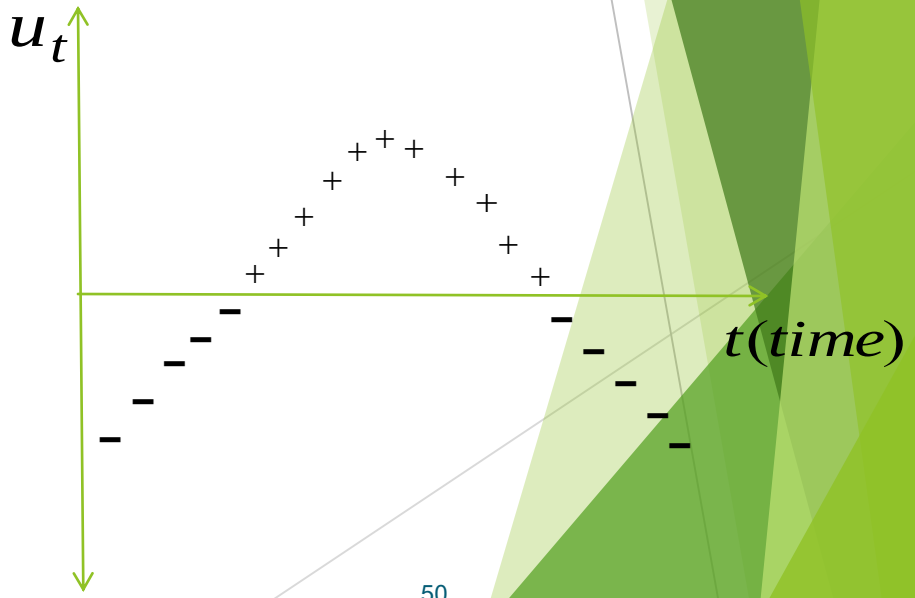



“Negative Autocorrelation”

**Second Method;** By this method we perform regress for residual values as function of time ($t$), if the values of sequential residuals arise for us uniform form this indicate there is Autocorrelation, as follows;

**1.** Positive Autocorrelation: This type occur when the number of sequential residuals have the same sign. i.e., the set of errors are positive follow it negative set then positive other and so on, as the following graph;

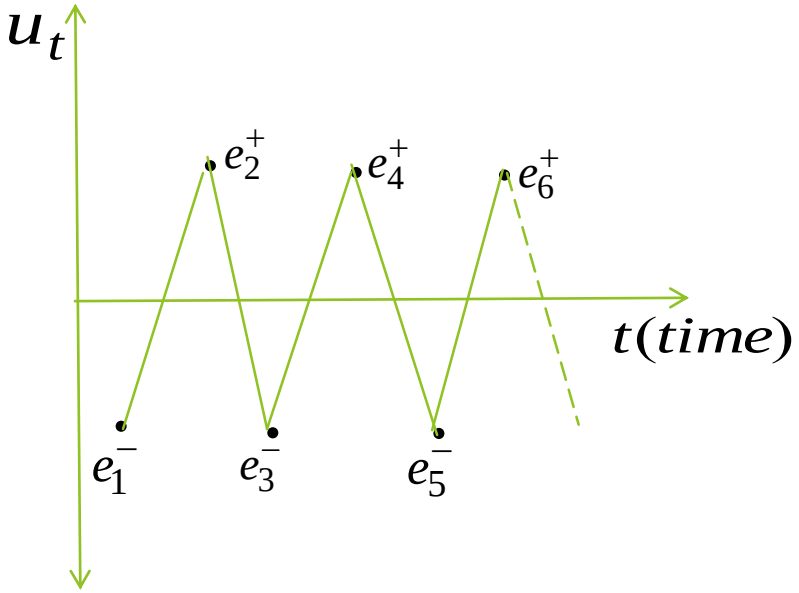


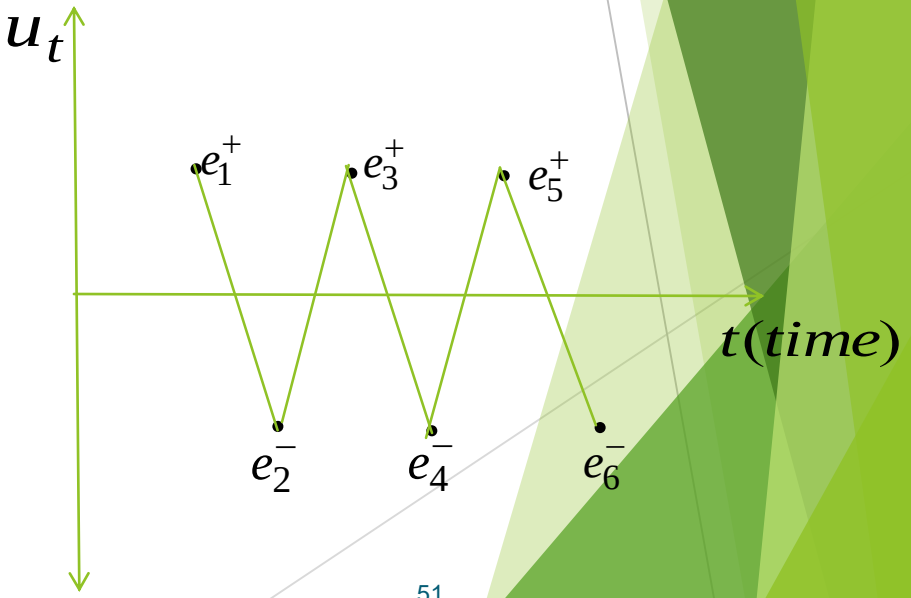






(( *Positive Autocorrelation*))

**2.** Negative Autocorrelation: This type occur when successional errors be alternating in sign. i.e., negative random error follow it positive random error and so on, as the following graph;






(( *Negative Autocorrelation*))

# Reasons of Autocorrelation Problem

**1.** mis-specification the mathematical form of the model (Wrong functional form).

**Ex:** if the real relation is second order but we depend on the model of first order, in this case the error term in model will contain consequences $X^2$ and this leads to get Autocorrelation in this term.

**2.** Inaccuracy of the information or data (mis-specification of a random variable $u_t$)

**Ex:** in case(wars, volcanoes, floods, quakes,..etc) then their consequences expand on next times.

**3.** Deleting some of explanatory variables from the model.

**4.** Autocorrelation bias

This case arises in crises or disturbances that are located in one of the regions to affect the economic budget in the neighboring regions

**5.** Modifications or Transformations of the data

- Interpolation of missing data with depend on the others observations values,
- differencing

# Consequences of Autocorrelation

1. Coefficient of estimates are unbiased, but the estimates are not BLUE.
2. The estimated variance of error term be biased (underestimated).Hence hypothesis tests (*t*, F) are suspect (Inaccuracy).
3. Low accuracy of the estimated parameters by OLS method.
4. Inaccuracy of the future forecasts.

# The expectation of correlated random variable

$u_t = \rho\, u_{t-1} + e_t$ (the autoregressive form)

$e_t \sim N(0, \sigma_e^2)$ , $\mathrm{Cov}(e_i, e_j) = 0, \ \forall \ i \neq j$ , $\mathrm{Cov}(e_i, X_i) = 0$

$$
\begin{aligned}
u_t &= \rho\, u_{t-1} + e_t \\
&= \rho\,(\rho\, u_{t-2} + e_{t-1}) + e_t \\
&= \rho^2 u_{t-2} + \rho\, e_{t-1} + e_t \\
&= \rho^2 (\rho\, u_{t-3} + e_{t-2}) + \rho\, e_{t-1} + e_t \\
&= \rho^3 u_{t-3} + \rho^2 e_{t-2} + \rho\, e_{t-1} + e_t \\
&= \text{... (continue to substitute)} \\
&= e_t + \rho\, e_{t-1} + \rho^2 e_{t-2} + \rho^3 e_{t-3} + \ldots = \sum_{i=0}^{\infty} \rho^i e_{t-i} \\
&= \text{(the moving average form)}
\end{aligned}
$$

$$\mathrm{E}(u_t) = \mathrm{E}(e_t) + \rho\, \mathrm{E}(e_{t-1}) + \rho^2\, \mathrm{E}(e_{t-2}) + \rho^3\, \mathrm{E}(e_{t-3}) + \ldots$$

$\because \quad \mathrm{E}(e_t) = \mathrm{E}(e_{t-1}) = \mathrm{E}(e_{t-2}) = \ldots = 0$

$\therefore \quad \mathrm{E}(u_t) = 0$

# The variance of correlated random variable

$$u_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \rho^3 e_{t-3} + \ldots$$

$$v(u_t) = E[u_t - E(u_t)]^2 \quad , \quad E(u_t) = 0$$

$$= E(u_t)^2 = E[e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \rho^3 e_{t-3} + \ldots]^2$$

$$= E[e_t^2 + \rho^2 e_{t-1}^2 + \rho^4 e_{t-2}^2 + \rho^6 e_{t-3}^2 + \ldots + 2\rho e_t e_{t-1} +$$

$$+ 2\rho^2 e_t e_{t-2} + 2\rho^3 e_t e_{t-3} + \ldots\ldots]$$

$$= E(e_t^2) + \rho^2 E(e_{t-1}^2) + \rho^4 E(e_{t-2}^2) + \rho^6 E(e_{t-3}^2) + \ldots\ldots$$

$$\because \quad E(e_t^2) = E(e_{t-1}^2) = E(e_{t-2}^2) = \ldots\ldots = \sigma_e^2$$

$$and \quad E(e_t e_{t-1}) = E(e_t e_{t-2}) = E(e_{t-1} e_{t-2}) = \ldots\ldots = 0$$

$$\therefore \quad v(u_t) = \sigma_e^2 + \rho^2 \sigma_e^2 + \rho^4 \sigma_e^2 + \ldots\ldots = \sigma_e^2 (1 + \rho^2 + \rho^4 + \ldots\ldots)$$

$$\therefore v(u_t) = \frac{1}{1 - \rho^2} \sigma_e^2 = \sigma_u^2$$

# The covariance of correlated random variable

$$\operatorname{cov}(u_t, u_{t-1}) = E[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]$$

$$\because E(u_t) = E(u_{t-1}) = 0$$

$$\therefore \operatorname{cov}(u_t, u_{t-1}) = E[u_t][u_{t-1}]$$

$$= E\left(e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \ldots\right)\left(e_{t-1} + \rho e_{t-2} + \rho^2 e_{t-3} + \ldots\right)$$

$$= E\left(e_t + \rho\left(e_{t-1} + \rho e_{t-2} + \rho^2 e_{t-3} + \ldots\right)\left(e_{t-1} + \rho e_{t-2} + \rho^2 e_{t-3} + \ldots\right)\right)$$

$$= E\left(e_t\left(e_{t-1} + \rho e_{t-2} + \rho^2 e_{t-3} + \ldots\right) + \rho\left(e_{t-1} + \rho e_{t-2} + \rho^2 e_{t-3} + \ldots\right)^2\right)$$

$$= E e_t e_{t-1} + \rho E e_t e_{t-2} + \rho^2 E e_t e_{t-3} + \ldots + \rho E\left(e_{t-1} + \rho e_{t-2} + \ldots\right)^2$$

$\because E(e_{t-1} + \rho e_{t-2} + ...)^2 = v(u_{t-1})$ , $\because v(u_t) = v(u_{t-1}) = \frac{1}{1-\rho^2}\sigma_e^2$

*and* $E(e_t e_{t-1}) = E(e_t e_{t-2}) = E(e_t e_{t-3}) = ..... = 0$

$\therefore \text{cov}(u_t, u_{t-1}) = 0 + \rho\sigma_u^2 = \rho\sigma_u^2 = \rho\frac{1}{1-\rho^2}\sigma_e^2$

$\therefore \text{cov}(u_t, u_{t-2}) = \rho^2\sigma_u^2$

*and so on*

$\therefore \text{cov}(u_t, u_{t-i}) = \rho^i\sigma_u^2$ , $\sigma_u^2 = \frac{1}{1-\rho^2}\sigma_e^2$

$\rho$ : *correlation coefficient between random variables*

$\rho = \frac{\text{cov}(u_t, u_{t-1})}{\sqrt{v(u_t)}\sqrt{v(u_{t-1})}} = \frac{\text{cov}(u_t, u_{t-1})}{v(u_t)} = \frac{\text{cov}(u_t, u_{t-1})}{\sigma_u^2}$ *Autocorrelation coefficien*

# The variance – covariance matrix of correlated random variables

$$E(u_t) = 0 \quad , \quad v(u_t) = \sigma_u^2 = \frac{1}{1-\rho^2}\,\sigma_e^2 \quad , \quad \operatorname{cov}(u_t\,,\,u_{t-i}) = \rho^i\,\sigma_u^2$$

$$\text{var}-\text{cov}(u_t) = E\,\underline{u}\,\underline{u}' = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u^2 & \rho^2\sigma_u^2 & .... & \rho^{n-1}\sigma_u^2 \\ & \sigma_u^2 & \rho\sigma_u^2 & .... & \rho^{n-2}\sigma_u^2 \\ & & \sigma_u^2 & .... & \rho^{n-3}\sigma_u^2 \\ & & & .... & .... \\ & & & & \rho^{n-n}\sigma_u^2 \end{pmatrix}$$

$$= \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{n-1} \\ & 1 & \rho & ... & \rho^{n-2} \\ & & 1 & ... & \rho^{n-3} \\ & & & ... & ... \\ & & & & 1 \end{pmatrix} = \sigma_u^2\ \Omega$$

$\Omega$ : *Autocorrelations matrix*

$$\therefore\ \text{var}-\text{cov}(u_t) = E\,\underline{u}\,\underline{u}' = \sigma_u^2\ \Omega$$

# The estimated parameters by OLS method are unbiased in the case of Autocorrelation problem

$$\underline{\hat{\beta}}_{OLS} = (X'X)^{-1} X'\underline{Y} \qquad , \quad \underline{Y} = X\,\underline{\beta} + \underline{U}$$

$$= (X'X)^{-1} X' (X\,\underline{\beta} + \underline{U}) = (X'X)^{-1} X' X\,\underline{\beta} + (X'X)^{-1} X'\,\underline{U}$$

$$= \underline{\beta} + (X'X)^{-1} X'\,\underline{U}$$

$$E(\underline{\hat{\beta}}_{OLS}) = \underline{\beta} + (X'X)^{-1} X'\,E(\underline{U}) \quad , \; E(\underline{U}) = \underline{0}$$

$$\therefore \; E(\underline{\hat{\beta}}_{OLS}) = \underline{\beta} \qquad \Rightarrow \; \underline{\hat{\beta}}_{OLS} \textit{ is unbiased est. for } \underline{\beta}$$

# The Variance – Covariance Form for estimated parameters by OLS method in the case of Autocorrelation problem

$$\text{var}-\text{cov}(\underline{\hat{\beta}}) = E\left(\underline{\hat{\beta}} - E(\underline{\hat{\beta}})\right)\left(\underline{\hat{\beta}} - E(\underline{\hat{\beta}})\right)'$$

$$\because \underline{\hat{\beta}} = \underline{\beta} + (X'X)^{-1}X'\underline{U} \quad \Rightarrow \quad \underline{\hat{\beta}} - \underline{\beta} = (X'X)^{-1}X'\underline{U} \quad , \quad E(\underline{\hat{\beta}}) = \underline{\beta}$$

$$\therefore \text{var}-\text{cov}(\underline{\hat{\beta}}) = E\left((X'X)^{-1}X'\underline{U}\right)\left((X'X)^{-1}X'\underline{U}\right)'$$

$$= E\left((X'X)^{-1}X'\underline{U}\,\underline{U}'X(X'X)^{-1}\right)$$

$$= (X'X)^{-1}X'E(\underline{U}\,\underline{U}')X(X'X)^{-1} \quad , \quad E(\underline{U}\,\underline{U}') = \sigma_u^2\,\Omega$$

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ & 1 & \rho & \dots & \rho^{n-2} \\ & & \dots & \dots & \dots \\ & & & & 1 \end{pmatrix}$$

$$\text{var}-\text{cov}(\underline{\hat{\beta}}) = \sigma_u^2\,(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

We conclude from above form that estimated parameters are not best estimators because the variance of these parameters is not minimum because contain error which represent with autocorrelations matrix.

# Durbin-Watson Test (D.W)

**1.** Determination the hypotheses of D.W test

$H_o : \rho = 0$ ; *there is no Autocorrelation between random errors*

$H_1 : \rho \neq 0$ ; *there is Autocorrelation between random errors from the first order*

**2.** We calculate test statistic value for D.W

$$D.W = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \qquad (( \text{Calculated value}))$$

$$= \frac{\sum_{t=2}^{n} \hat{u}_t^2 - 2\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1} + \sum_{t=2}^{n} \hat{u}_{t-1}^2}{\sum_{t=1}^{n} \hat{u}_t^2}$$

$$when\ n \to \infty\ then\ \sum_{t=2}^{n} \hat{u}_t^2 \cong \sum_{t=1}^{n} \hat{u}_t^2 \cong \sum_{t=2}^{n} \hat{u}_t^2 - 1$$

$$\therefore D.W = \frac{2\sum_{t=2}^{n} \hat{u}_t^2 - 2\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{n} \hat{u}_t^2} = 2 - \frac{2\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{n} \hat{u}_t^2} = 2 - \frac{2\,\mathrm{cov}(\hat{u}_t, \hat{u}_{t-1})}{\mathrm{var}(u_t)}$$

$$= 2 - 2\hat{\rho} = 2(1 - \hat{\rho})$$

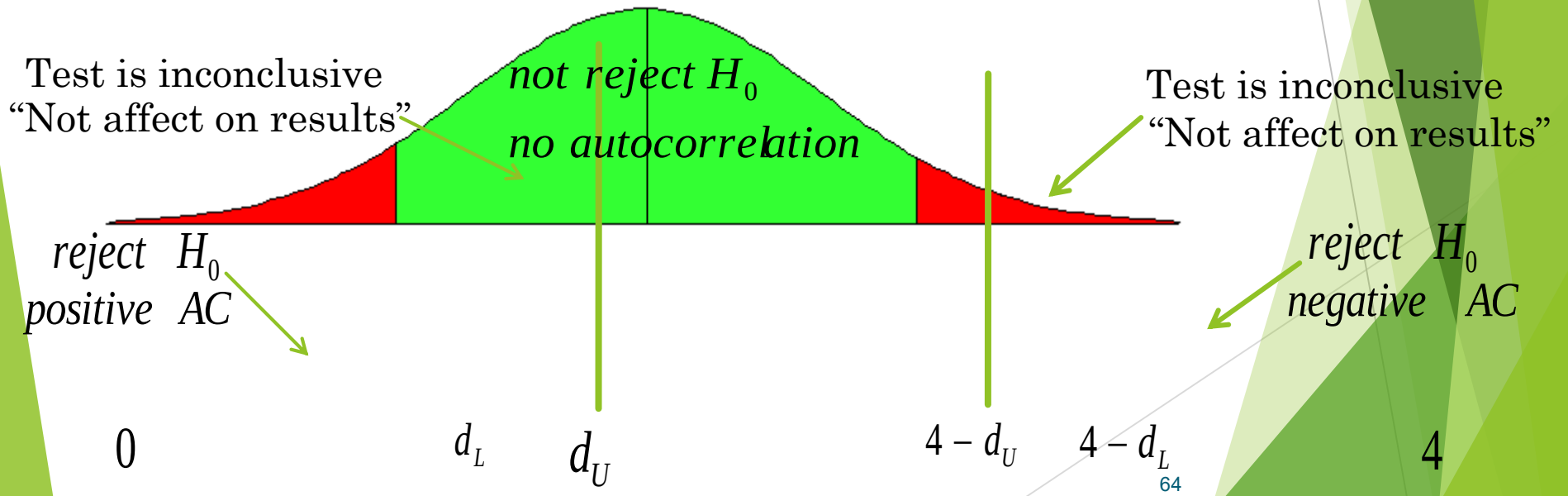
$if;\quad \hat{\rho} = 0 \Rightarrow D.W = 2 \Rightarrow$ indicate to no Autocorrelation

$\hat{\rho} = +1 \Rightarrow D.W = 0 \Rightarrow$ indicate to positive Autocorrelation

$\hat{\rho} = -1 \Rightarrow D.W = 4 \Rightarrow$ indicate to negative Autocorrelation

$$\therefore \quad 0 \le D.W \le 4$$

**3.** We compare calculated value with tabulated value ($d_L$,$d_U$), with taking in account number of observations($n$) and number of parameters (p) under particular significant level ($\alpha$).



**Ex:** From the following data

$X_t$: 6.3 , 6 , 5.9 , 3 , 5 , 6.3 , 5.6 , 3.6 , 2.5 , 2.9 , 2.2 , 3.9 , 4.5 , 4.3 , 4

$Y_t$: 2.76 , 4.76 , 8.75 , 7.78 , 6.18 , 9.5 , 5.14 4.76 , 16.7 , 27.68 , 26.64 , 13.71 , 12.32 , 15.73 , 13.59

1) Estimate simple linear model.

2) Test the problem of Autocorrelation between errors, if you know the tabulated value for (D.W) under significant level 5% and degrees of freedom (1,15) are: $d_L = 1.08$ , $d_U = 1.38$

$$n = 15 \quad , \quad \Sigma X_t = 66 \quad , \Sigma Y_t = 176$$

$$\Sigma X_t^2 = 317.96 \quad , \quad \Sigma X_t Y_t = 669.121$$

**Sol: 1)** $\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y} = \begin{pmatrix} 15 & 66 \\ 66 & 317.96 \end{pmatrix}^{-1} \begin{pmatrix} 176 \\ 669.121 \end{pmatrix} = \begin{pmatrix} 28.541 \\ -3.82 \end{pmatrix}$

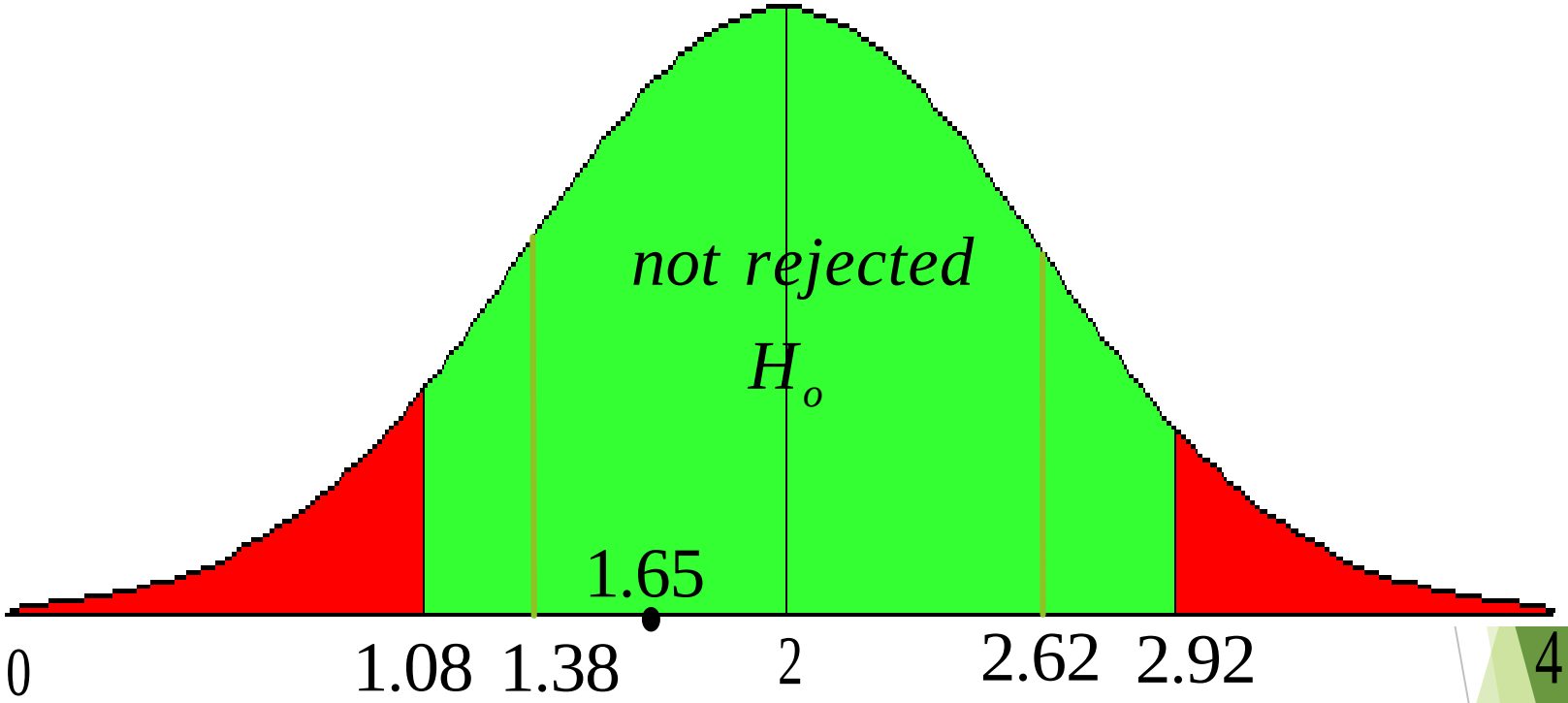
$\hat{Y}_i = 28.541 - 3.82 X_i \ , \ u_i = Y_i - \hat{Y}_i$

**2)** $H_o: \rho = 0$

$H_1: \rho \neq 0$

$$D.W = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} = \frac{667.9924}{404.283363} = 1.65229$$

| | $X_t$ | $Y_t$ | $\hat{Y}_i$ | $\hat{u}_t$ | $\hat{u}_t^2$ | $\hat{u}_{t-1}$ | $\hat{u}_t - \hat{u}_{t-1}$ | $(\hat{u}_t - \hat{u}_{t-1})^2$ |
|---|---|---|---|---|---|---|---|---|
| | … | … | … | … | … | … | … | … |
| | … | … | … | … | … | … | … | … |
| **Sum** | | | | | **404.2834** | | | **667.9924** |




We note that the calculated value for D.W fall in not rejected region $H_o$, this means we not rejected $H_o$ and this indicate there is no Autocorrelation problem between errors.

Ex// Test the Auto correlation by using Durbin-Watson statistic

| $e_i$ | $e_{i-1}$ | $(e_i - e_{i-1})$ | $(e_i - e_{i-1})^2$ | $e_i^2$ |
|---|---|---|---|---|
| **-1.108** | **-** | **-** | **-** | **1.227664** |
| **-2.72** | **-1.108** | **-3.828** | **14.65358** | **7.3984** |
| **2.044** | **-2.72** | **4.764** | **22.6957** | **4.177936** |
| **3** | **2.044** | **0.956** | **0.913936** | **9** |
| **-2.944** | **3** | **-5.944** | **35.33114** | **8.667136** |
| **2.072** | **-2.944** | **5.016** | **25.16026** | **4.293184** |
| **-1.892** | **2.072** | **-3.964** | **15.7133** | **3.579664** |
| | | | **114.4679** | **38.34398** |

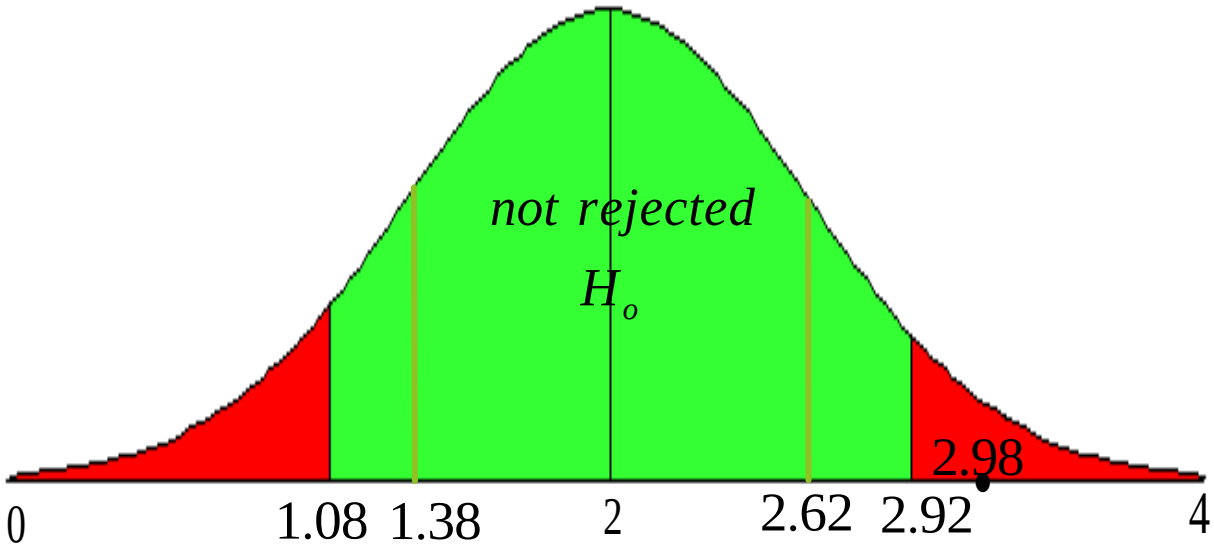
**Sol: 1)**

$H_o: \rho = 0$

$H_1: \rho \neq 0$

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} ei^2} = 2.985$$

$d_L$=1.08 $d_u$=1.38




We note that the calculated value for D.W falls in rejected region $H_o$, which means we rejected $H_o$ and this indicates there is an Autocorrelation problem between errors.

# Remedial Methods of Autocorrelation

1. The Generalized Least Squares (GLS) Method.
2. First difference method(Cochran-Orcutt Method).
3. Iterative Method.

# The Generalized Least Squares Method

This method remedies the problem of Autocorrelation between random variables in standard models as well as its remedy the problem of Heteroscedasticity,as follows; when we have Autocorrelation between errors, then;

$$E\,\underline{u}\,\underline{u}' = \sigma_u^2\,\Omega$$

Then Ω is a square and symmetric matrix of order $(n \times n)$

and has inverse. When the error term(random variable) follows Markov form the first order;

$$u_t = \rho\, u_{t\text{-}1} + e_t$$

Then Ω matrix takes the following form:-

$$\Omega = \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \dots & \hat{\rho}^{n-1} \\ & 1 & \hat{\rho} & \dots & \hat{\rho}^{n-2} \\ & & 1 & \dots & \hat{\rho}^{n-3} \\ & & & \dots & \dots \\ & & & & 1 \end{pmatrix}_{n \times n}$$

$$\Omega^{-1} = \frac{1}{1-\hat{\rho}^2} \begin{pmatrix} 1 & -\hat{\rho} & 0 & 0 & \dots & 0 \\ -\hat{\rho} & (1+\hat{\rho}^2) & -\hat{\rho} & 0 & \dots & 0 \\ 0 & -\hat{\rho} & (1+\hat{\rho}^2) & -\hat{\rho} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & (1+\hat{\rho}^2) & -\hat{\rho} \\ 0 & 0 & 0 & \dots & -\hat{\rho} & 1 \end{pmatrix}$$

The method of **GLS** collected the **OLS** method manner to make it take into account the relationship correlation among random variables, and thus, the estimators and variances of this method will be as follows compared with the **OLS** method.

| | OLS Method | GLS Method |
|---|---|---|
| 1. | $E(\underline{u}\,\underline{u}') = \sigma_u^2\, I$ | $E(\underline{u}\,\underline{u}') = \sigma_u^2\, \Omega$ |
| 2. | $\underline{\hat{\beta}}_{OLS} = (X'X)^{-1}X'\underline{Y}$ | $\underline{\hat{\beta}}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\underline{Y}$ |
| 3. | $\hat{\sigma}_u^2 = \frac{\underline{Y}'\underline{Y} - \underline{\hat{\beta}}'_{OLS}\, X'\underline{Y}}{n-k-1}$ | $\hat{\sigma}_u^2 = \frac{\underline{Y}'\Omega^{-1}\underline{Y} - \underline{\hat{\beta}}'_{GLS}\, X\Omega^{-1}\underline{Y}}{n-k-1}$ |
| 4. | $\text{var}-\text{cov}(\underline{\hat{\beta}}) = \sigma_u^2(X'X)^{-1}$ | $\text{var}-\text{cov}(\underline{\hat{\beta}}) = \sigma_u^2(X'\Omega^{-1}X)^{-1}$ |

The **GLS** method needs to prior information about the parameter $\rho$ and how enter it to the matrix $\Omega$ because the elements of $\Omega$ are unknown , and we can estimate it as follows:-

1. Iterative Method to get $\hat{\rho}$ .
2. Durbin-Watson Method: we apply OLS method, and calculate $\hat{\rho}$ using D.W. statistic, as follows;

$$D.W = 2 - 2\,\hat{\rho} = 2(1 - \hat{\rho})$$

$$1 - \hat{\rho} = \frac{D.W}{2} \quad \Rightarrow \quad \therefore \quad \hat{\rho} = 1 - \frac{D.W}{2}$$

3. Using Theil – Nagar Method.

By this method we estimate $\hat{\rho}$ by the following form;

$$\hat{\rho} = \frac{n^2\left(1 - \frac{D.W}{2}\right) + k^2}{n^2 - k^2}$$

Where:

$n$: the sample size (No. of observations).

$k$: the No. of estimated parameters (with $\beta_o$)

**Ex:** Random sample of size (5) observations

$Y_t$: 1 3 2 1 0

$X_t$: 2 5 4 3 1

1) Estimate coefficients of model using:

a) OLS method. , $Y_t = \beta_o + \beta_1 X_t + u_t$

b) GLS method.

If you know $u_t \sim N(0, 0.3\Omega)$

$u_t = \rho\, u_{t-1} + e_t$ , from first order

And estimated value for $\rho = -0.7$

2) Estimate var – cov. Matrix for estimated coefficients by GLS method.

# Sol: Simple Regression – Y: a) OLS Method

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix}^{-1} \begin{pmatrix} 7 \\ 29 \end{pmatrix}, \quad |X'X| = 275 - 225 = 50$$

| Parameter | Estimate | Standard Error | T Statistic | P-Value |
|---|---|---|---|---|
| CONSTANT | -0.7 | 0.331662 | -2.11058 | 0.1253 |
| X | 0.7 | 0.1 | 7.0 | 0.0060 |

**Analysis of Variance**

| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| Model | 4.9 | 1 | 4.9 | 49.00 | 0.0060 |
| Residual | 0.3 | 3 | 0.1 | | |
| Total (Corr.) | 5.2 | 4 | | | |

$\hat{Y}_t = -0.7 + 0.7X_t$

R-squared = 94.2308 percent

R-squared (adjusted for d.f.) = 92.3077 percent

Standard Error of Est. = 0.316228

Mean absolute error = 0.2

Durbin-Watson statistic = 1.16667

## b) GLS Method

$$\underline{\hat{\beta}}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \underline{Y}$$

$$\Omega^{-1} = \frac{1}{0.51} \begin{pmatrix} 1 & .7 & 0 & 0 & 0 \\ .7 & 1.49 & .7 & 0 & 0 \\ 0 & .7 & 1.49 & .7 & 0 \\ 0 & 0 & .7 & 1.49 & .7 \\ 0 & 0 & 0 & .7 & 1 \end{pmatrix}$$

$$\hat{\underline{B}} = \frac{0.51}{0.51}\left[\begin{pmatrix}1&1&1&1&1\\2&5&4&3&1\end{pmatrix}\begin{pmatrix}1&.7&0&0&0\\.7&1.49&.7&0&0\\0&.7&1.49&.7&0\\0&0&.7&1.49&.7\\0&0&0&.7&1\end{pmatrix}\begin{pmatrix}1&2\\1&5\\1&4\\1&3\\1&1\end{pmatrix}\right]^{-1} \times \begin{pmatrix}1&1&1&1&1\\2&5&4&3&1\end{pmatrix}\begin{pmatrix}1&.7&0&0&0\\.7&1.49&.7&0&0\\0&.7&1.49&.7&0\\0&0&.7&1.49&.7\\0&0&0&.7&1\end{pmatrix}\begin{pmatrix}1\\3\\2\\1\\0\end{pmatrix}$$

$$= \left[\begin{pmatrix}1.7&2.89&2.89&2.89&1.7\\5.5&11.65&11.56&7.97&3.1\end{pmatrix}\begin{pmatrix}1&2\\1&5\\1&4\\1&3\\1&1\end{pmatrix}\right]^{-1}\begin{pmatrix}1.7&2.89&2.89&2.89&1.7\\5.5&11.65&11.56&7.97&3.1\end{pmatrix}\begin{pmatrix}1\\3\\2\\1\\0\end{pmatrix}$$

$$= \begin{pmatrix}12.07&39.78\\39.78&142.5\end{pmatrix}^{-1}\begin{pmatrix}19.04\\71.54\end{pmatrix} = \frac{1}{1375.266}\begin{pmatrix}142.5&-39.78\\-39.78&12.07\end{pmatrix}\begin{pmatrix}19.04\\71.54\end{pmatrix}$$

$$= \begin{pmatrix}1.0362&-0.2893\\-0.2893&0.0878\end{pmatrix}\begin{pmatrix}19.04\\71.54\end{pmatrix} = \begin{pmatrix}-0.9673\\0.7731\end{pmatrix}, \quad \hat{Y}_t = -0.9673 + 0.7731X_t$$

2.

$$\hat{\sigma}_u^2 = \frac{\underline{Y}' \Omega^{-1} \underline{Y} - \underline{\hat{\beta}}'_{GLS} X' \Omega^{-1} \underline{Y}}{n - k - 1} = 0.3$$

$$\text{var} - \text{cov}(\underline{\hat{\beta}}) = \sigma_u^2 (X' \Omega^{-1} X)^{-1}$$

$$= 0.3 \begin{pmatrix} 1.0362 & -0.2893 \\ -0.2893 & 0.0878 \end{pmatrix} = \begin{pmatrix} 0.31086 & -08679 \\ -08679 & 0.02634 \end{pmatrix}$$

$$= \begin{pmatrix} v(\hat{\beta}_o) & \text{cov}(\hat{\beta}_o, \hat{\beta}_1) \\ \text{cov}(\hat{\beta}_o, \hat{\beta}_1) & v(\hat{\beta}_1) \end{pmatrix}$$

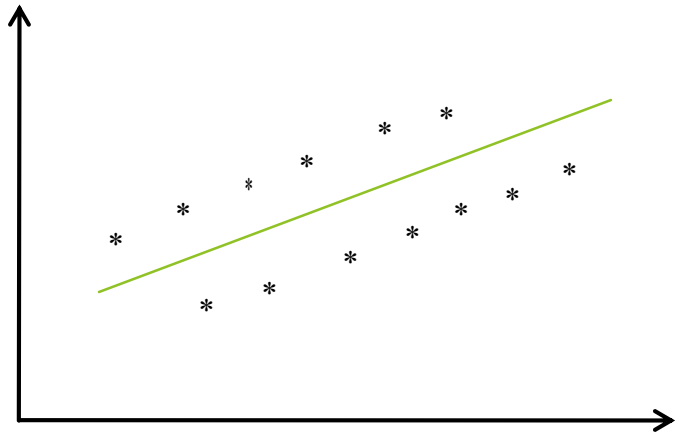
# Heteroscedasticity Problem

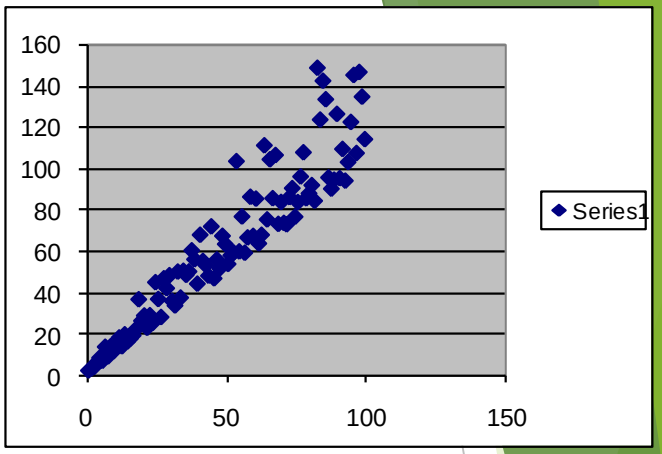
**Definition:** Heteroscedasticity is a problem when the error terms do not have a constant variance, $E(u_i^2) = \sigma_i^2$

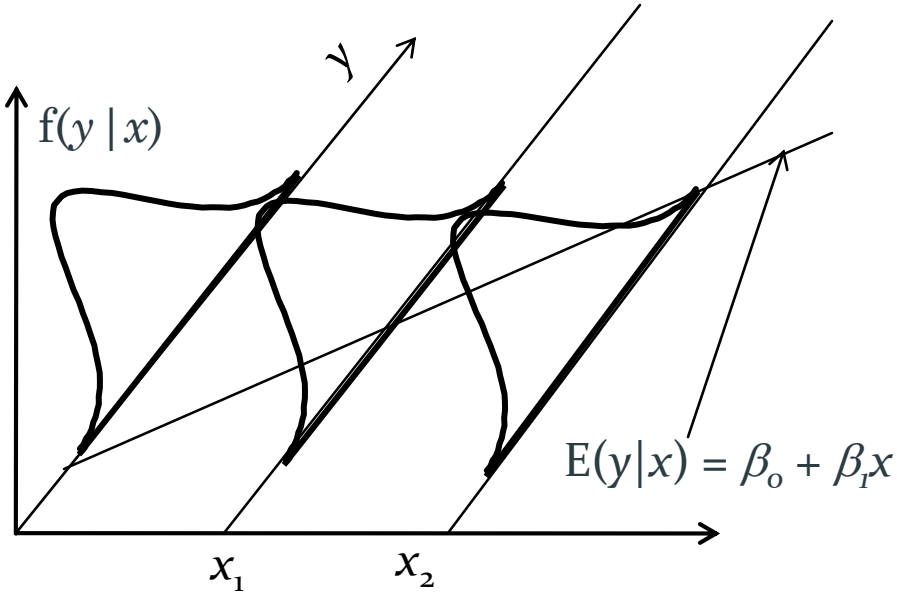
i.e., $\text{var}(u_i) \neq \sigma_u^2 \quad , \quad i = 1, 2, \ldots, n$

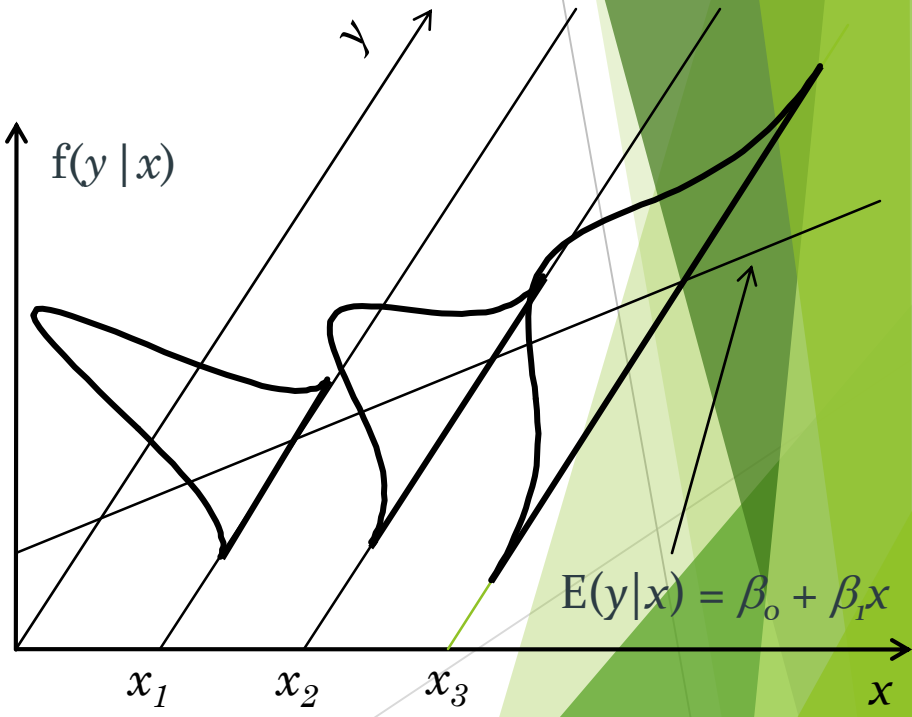
$or; \ \sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_n^2 \neq \sigma_u^2$

- i.e. the probability distribution of random variable ($u_i$) not constant for all explanatory variable values, this means, there is systematic relation between random variable and explanatory variable, i.e., $\text{cov}(u_i, X_i) \neq 0$
- That is, they may have a larger variance when values of some $X_i$ (or the $Y_i$'s themselves) are large (or small).






**Homoscedastic Case**

**Heteroscedastic Case**

# Reasons of Heteroscedasticity Problem

- It may be caused by:

  1- Model misspecification.

  2- Estimation parameters of model based on cross sectional data sets.

  3- Outliers in data.

# Consequences of Heteroscedasticity

1. The estimated coefficients using OLS method are not satisfy in it BLUE property (not minimum variance).
2. The t – test and F – test results may be misleading.
3. The estimated coefficients are inaccurate consequently that the using of estimated model become not logical and then leads to an inaccurate results

# Tests for Detecting of Heteroscedasticity

## 1- Informal Methods

**a-** Graph the data and look for patterns!

**b-** Plot the residuals against each of the X's variables.

# 2- Goldfeld-Quandt Test

This test assumes normal distribution and there is no autocorrelation between errors ($u_i$'s)

The steps of this test as follows;

**1.** Determination the hypotheses of test:

$H_o : \sigma_1^2 = \sigma_2^2 = .... = \sigma_n^2 = \sigma_u^2$ *or Homogeneity of errors* var*iances*

$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq .... \neq \sigma_n^2 \neq \sigma_u^2$ *or Heterogeneity of errors* var*iances*

2. Order the data in ascending order according to the values of Xi

**3.** We chose a particular number named **c** from the middle of observations and delete it from analysis.

**c**: No. of deleted middle observations from analysis.

Can be **(c = *n* / 3 or less than it)**

In general;

If $n = 30$ then c = 8 , $n = 40$ then c = 12

$n = 60$ then c = 16

The remain observations [ $n$ – c] divide into two sets of equal size (two samples), [ the size of each partial sample equal to $n_1 = n_2 = (n - c) / 2$ observation].

The first partial sample contains small partial values for X and denoted $X_{1i}$ and the second contains large partial values for X and denoted $X_{2i}$ .

**4.** Apply separate regressions for each partial sample (on both upper and lower samples) to find the estimation for coefficients of linear relation between Y & X, and find $(S_1^2, S_2^2)$ .

$$S_1^2 = \frac{\sum_{i=1}^{n_1} e_{1i}^2}{n_1 - k - 1} \quad , \quad S_2^2 = \frac{\sum_{i=1}^{n_2} e_{2i}^2}{n_2 - k - 1}$$

where; $k$: No. of explanatory variables in model.

5. Compute calculated value of F (Cal F), as follows;

$$Cal\ F = \frac{S_2^2}{S_1^2} \ , \ Tab\,F(\alpha, v_1, v_2)$$

$$v_1 = v_2 = n_1 - k - 1 \ (\text{degrees of freedom})$$

**6.** We compare between (Cal F) and (Tab F)

If Cal F $\geq$ Tab F , we rejected $H_o$

This means there is heterogeneity problem between errors variances.

Or this indicate hetrogeneity of error variance.

If Cal F < Tab F , we not rejected Ho

This means there is no heterogeneity problem between errors variances.

Or this indicate homogeneity of error variance.

**Ex:** From the following data , test if there is heterogeneity problem between errors variances or not, using **Goldfeld – Quandt** test, if you know,

$n$ = 10 , Tab F(0.05, 2, 2) = 19

| $X_i$ | $Y_i$ |
|---|---|
| 39 | 65 |
| 43 | 74 |
| 21 | 52 |
| 64 | 82 |
| 57 | 92 |
| 47 | 74 |
| 28 | 73 |
| 75 | 98 |
| 34 | 56 |
| 52 | 75 |

$$H_o : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{11}^2 = \sigma_u^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_{11}^2 \neq \sigma_u^2$$

$n = 10$ if $c = 2$ deleted observations

$n - c = 10 - 2 = 8$ remained observations

$n_1 = n_2 = \frac{n - c}{2} = \frac{8}{2} = 4$ *No. of* observations *in each sample*

# Order the *data* in ascending order according to the values of $X_i$

| $X_i$ | $Y_i$ |
|---|---|
| 21 | 52 |
| 28 | 73 |
| 34 | 56 |
| 39 | 65 |
| 43 | 74 |
| 47 | 74 |
| 52 | 75 |
| 57 | 92 |
| 64 | 82 |
| 75 | 98 |

The first sample

| $X_{1i}$ | $Y_{1i}$ |
|---|---|
| 21 | 52 |
| 28 | 73 |
| 34 | 56 |
| 39 | 65 |

The second sample

| $X_{2i}$ | $Y_{2i}$ |
|---|---|
| 52 | 75 |
| 57 | 92 |
| 64 | 82 |
| 75 | 98 |

# The first sample

| | $X_{1i}$ | $Y_{1i}$ | $\hat{Y}_{1i}$ | $u_{1i} = Y_{1i} - \hat{Y}_{1i}$ | $u_{1i}^2$ |
|---|---|---|---|---|---|
| | **21** | **52** | | | |
| | **28** | **73** | | | |
| | **34** | **56** | | | |
| | **39** | **65** | | | |
| Sum | | | | | 236.359 |

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} = \begin{pmatrix} n & \Sigma X_{1i} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma Y_{1i} \\ \Sigma X_{1i}Y_{1i} \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 122 \\ 122 & 3902 \end{pmatrix}^{-1} \begin{pmatrix} 246 \\ 7575 \end{pmatrix} = \begin{pmatrix} 49.3674 \\ 0.39779 \end{pmatrix}$$

$$\hat{Y}_{1i} = 49.3674 + 0.39779X_{1i}$$

$$S_1^2 = \frac{\Sigma u_{1i}^2}{n_1 - k - 1} = \frac{236.359}{2} = 118.18$$

$$Cal\ F = \frac{S_2^2}{S_1^2} = \frac{118.18}{70.1535} = 1.6845$$

$$Tab\ F(0.05,\ 2,\ 2) = 19$$

$$\because 1.6845 < 19$$

$$\therefore Cal\ F < Tab\ F\ ,\ we\ not\ \ rejected\ H_o$$

This means there is no heterogeneity problem between errors variance.

Or this indicate homogeneity of error variance.

**Ex:** From the following data , test if there is heterogeneity problem between errors variances or not , using **Goldfeld – Quandt** test, if you know,

$n$ = 30 , Tab F(0.05, 9, 9) = 3.18

| $X_i$ | $Y_i$ | | | | | |
|---|---|---|---|---|---|---|
| 1 | 6 | 4 | 3 | 5 | 6 | 2 |
| 2 | 3 | 4 | 4 | 3 | 4 | |
| 3 | 5 | 4 | 7 | 5 | | |
| 4 | 5 | 6 | 7 | 3 | 3 | |
| 5 | 3 | 7 | 8 | 7 | | |
| 6 | 2 | 6 | 4 | 10 | | |
| 7 | 5 | 8 | | | | |

Sol.: $n$ = 30 then c = 8

deleted observations

$n$ – c = 30 – 8 = 22

remained observations

$$n_1 = n_2 = \frac{n-c}{2} = \frac{22}{2} = 11$$

*No. of observation in each sample*

$$H_o : \sigma_1^2 = \sigma_2^2 = ... = \sigma_{11}^2 = \sigma_u^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_{11}^2 \neq \sigma_u^2$$

# The first sample

| | $X_{1i}$ | $Y_{1i}$ | $\hat{Y}_{1i}$ | $u_{1i} = Y_{1i} - \hat{Y}_{1i}$ | $u_{1i}^2$ |
|---|---|---|---|---|---|
| | 1 | 6 | 5.04 | 0.96 | 0.9216 |
| | 1 | 4 | = | - 1.04 | |
| | 1 | 3 | = | - 2.04 | |
| | 1 | 5 | = | - 0.04 | |
| | 1 | 6 | = | 0.96 | |
| | 1 | 2 | = | - 3.04 | |
| | 2 | 3 | 4.66 | - 1.66 | |
| | 2 | 4 | = | - 0.66 | |
| | 2 | 4 | = | - 0.66 | |
| | 2 | 3 | = | - 1.66 | |
| | 2 | 4 | = | - 0.66 | |
| Sum | 16 | 44 | | | 23.1476 |

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} = \begin{pmatrix} n & \Sigma X_{1i} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma Y_{1i} \\ \Sigma X_{1i}Y_{1i} \end{pmatrix}$$

$$= \begin{pmatrix} 11 & 16 \\ 16 & 26 \end{pmatrix}^{-1} \begin{pmatrix} 44 \\ 62 \end{pmatrix}, \; |X'X| = 30$$

$$= \begin{pmatrix} .87 & -.53 \\ -.53 & .37 \end{pmatrix} \begin{pmatrix} 44 \\ 62 \end{pmatrix} = \begin{pmatrix} 5.42 \\ -.38 \end{pmatrix}$$

$$\hat{Y}_{1i} = 5.42 - 0.38X_{1i}$$

$$S_1^2 = \frac{\Sigma u_{1i}^2}{n_1 - k - 1} = \frac{23.1476}{9} = 2.572$$

# The second sample

| | $X_{2i}$ | $Y_{2i}$ | $\hat{Y}_{2i}$ | $u_{2i} = Y_{2i} - \hat{Y}_{2i}$ | $u_{2i}^2$ |
|---|---|---|---|---|---|
| | 4 | 3 | 8.91 | - 5.91 | 34.9281 |
| | 5 | 3 | 10.53 | - 7.53 | |
| | 5 | 7 | = | - 3.53 | |
| | 5 | 8 | = | - 2.53 | |
| | 5 | 7 | = | - 3.53 | |
| | 6 | 2 | 12.15 | - 10.15 | |
| | 6 | 6 | = | - 6.15 | |
| | 6 | 4 | = | - 8.15 | |
| | 6 | 10 | = | - 2.15 | |
| | 7 | 5 | 13.77 | - 8.77 | |
| | 7 | 8 | = | - 5.77 | |
| Sum | 62 | 63 | | | 445.0475 |

$$\underline{\hat{\beta}} = \begin{pmatrix} 11 & 62 \\ 62 & 358 \end{pmatrix}^{-1} \begin{pmatrix} 63 \\ 360 \end{pmatrix}, \quad |X'X| = 94$$

$$= \begin{pmatrix} 3.81 & -.66 \\ -.66 & .12 \end{pmatrix} \begin{pmatrix} 63 \\ 360 \end{pmatrix} = \begin{pmatrix} 2.43 \\ 1.62 \end{pmatrix}$$

$$\hat{Y}_{2i} = 2.43 + 1.62 X_{2i}$$

$$S_2^2 = \frac{\Sigma u_{2i}^2}{n_2 - k - 1} = \frac{445.0475}{9} = 49.45$$

$$Cal\ F = \frac{S_2^2}{S_1^2} = \frac{49.45}{2.572} = 19.23$$

$$Tab\ F(0.05, 9, 9) = 3.18$$

$$\because 19.23 > 3.18$$

$$\therefore Cal\ F > Tab\ F\ ,\ we\ rejected\ H_o$$

This indicate heterogeneity of error variance, i.e., the errors suffering from the problem of heterogeneity of errors variances, it means there is heterogeneity problem between errors variances.

# 3- Spearman-Rank Correlation Coefficient Test

The steps of this test as follows;

**1.** Determination of test hypotheses

$$H_o: \ r_{u.X}^{S} = 0$$

$$H_1: \ r_{u.X}^{S} \neq 0$$

$r_{u.X}^{S}$ : *Correlation coefficient between ranks of errors and explanatory variable.*

**2.** Order the values of residuals ($u_i$'s) with the values of $X_i$ in ascending or descending order, with ignoring the signs of residuals (absolute value).

**3.** Compute Spearman-Rank Correlation Coefficient by the following form;

$$r^{S}_{u\,.\,X} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$$

Where; $D_i$ = Rank of $X_i$ – Rank of $u_i$

If $r^{S}_{u\,.X} \to 1$ indicate existence strong relation between errors and explanatory variable X, therefore there is Heteroscedasticity problem.

**4.** We have two cases

a) If $n < 30$ then we use $t$-test with $(n-2)$ df

$$Cal\ t = \frac{r^{S}_{u.X}\sqrt{n-2}}{\sqrt{1-(r^{S}_{u.X})^{2}}}\ ,\ Tab\ t(\alpha/2\ ,n-2)$$

b) If $n \geq 30$ then we use $z$-test

$$Cal\ z = \frac{r^{S}_{u.X}}{\hat{\sigma}_{r^{S}_{u.X}}}\ ,\quad \hat{\sigma}_{r^{S}_{u.X}} : standard\ deviation\ for\ Spearman-Rank$$

$$\hat{\sigma}_{r^{S}_{u.X}} = \frac{1}{\sqrt{n-1}} \Rightarrow Cal\ z = r^{S}_{u.X}\sqrt{n-1}$$

**5.** Compare between Cal $z$ and (Tab $z_a = \pm 1.96$) value We not rejected $H_o$ if $(-1.96 < \text{Cal } z < +1.96)$, i.e., the errors are homogeneous, there is no Heteroscedasticity problem. And with inverting it we rejected $H_o$, i.e., the errors are heterogeneous, there is Heteroscedasticity problem.

**Note:** We can apply this test if there are two or more explanatory variables in model, by computing rank correlation coefficient between $u_i$ and each $X_i$.

**Ex:** For the following data from (31) observations, test Heteroscedasticity problem at significant level 5% by using
1. Spearman-Rank correlation coefficient test.
2. Bartlett test.

| $n$ | $X_i$ | $Y_i$ | $\hat{Y}_i$ | $u_i$ | Ascending order for $u_i$ | Rank of $u_i$ | Rank of $X_i$ | Rank of $u_i^*$ | $D_i$ $= R_X - R_u$ | $D_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 8777 | 264 | 97.85 | 166.15 | 5.97 | 1 | 1 | 16 | -15 | 225 |
| 2 | 9210 | 105 | 184.46 | -29.46 | -27.45 | 2 | 2 | 3 | -1 | 1 |
| 3 | 9954 | 90 | 197.35 | -107.35 | -29.46 | 3 | 3 | 7 | -4 | 16 |
| 4 | 10508 | 131 | 244.18 | -113.18 | 77.10 | 4 | 4 | 8 | -4 | 16 |
| 5 | 10979 | 122 | 250.18 | -128.18 | -80.04 | 5 | 5 | 10 | -5 | 25 |
| 6 | 11912 | 107 | 362.87 | -255.87 | 100.97 | 6 | 6 | 23 | -17 | 269 |
| 7 | | | | -27.45 | -107.35 | 7 | 7 | 2 | 5 | 25 |
| 8 | | | | 5.97 | -113.18 | 8 | 8 | 1 | 7 | 49 |
| 9 | | | | -131.12 | -127.24 | 9 | 9 | 11 | -2 | 4 |
| 10 | | | | -80.04 | -128.18 | 10 | 10 | 5 | 5 | 25 |
| 11 | | | | 127.84 | -131.12 | 11 | 11 | 9 | 2 | 4 |
| 12 | | | | 100.97 | 133.68 | 12 | 12 | 6 | 6 | 36 |
| 13 | | | | -147.12 | -134.04 | 13 | 13 | 15 | -2 | 4 |
| 14 | | | | -196.73 | -135.54 | 14 | 14 | 20 | -6 | 36 |
| 15 | | | | 77.10 | 147.12 | 15 | 15 | 4 | 11 | 121 |
| 16 | | | | 411.96 | 166.15 | 16 | 16 | 28 | -12 | 144 |
| ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| 31 | 39700 | 2300 | 2585.11 | - 285.11 | | 31 | 31 | 24 | 7 | 49 |
| Sum | | | | | | | | | | 1474 |

**Note:** Rank of $u_i^*$: represents rank of $u_i$ based on the Variable $X_i$ corresponding $u_i$ value and $u_i$ rank.

$$H_o: \ r_{u.X}^S = 0$$

$$H_1: \ r_{u.X}^S \neq 0$$

$$r_{u.X}^S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} = 1 - \frac{6(1474)}{31(31^2-1)} = 0.703$$

The Spearman correlation coefficient is large and approximately approach to one, i.e., we rejected $H_o$, there is significant correlation between ($u_i$, $X_i$), that indicates existence Heteroscedasticity problem. And to test that we use z – test because ($n > 30$).

$H_o$: $u_i$'s are homogeneous

$H_1$: $u_i$'s are heterogeneous

$$Cal\ z = \frac{r^S_{u.X}}{\hat{\sigma}_{r^S_{u.X}}}\ , \quad \hat{\sigma}_{r^S_{u.X}} = \frac{1}{\sqrt{n-1}} = \frac{1}{\sqrt{31-1}} = 0.14$$

$$= \frac{0.703}{0.14} = 5.02 \quad , \quad Tab\, z = \mp 1.96$$

$$\because\ 5.02 > 1.96 \quad \Rightarrow Cal\, z > Tab\, z \quad , \ we\ rejected\ H_o$$

This means that the errors are heterogeneous, which there is Heteroscedasticity problem.

# Park test

- As an exploratory test, log the residuals and regress them on the logged values of the suspected independent variable.

$$\ln \hat{e}_i^2 = \ln \sigma^2 + \beta \ln X_i + u_i$$

$$= a + \beta \ln X_i + u_i$$

- If the $\beta$ is significant, then heteroscedasticity may be a problem.

# Bartlett Test

• The basic idea of this test is partition the sample into (m) partial samples , then computing the error variance for each partial samples ($S_i^2$) with ($n_i - 1$) degrees of freedom.

• Often this type of test apply on the samples which available in it more than one observation for each value from explanatory variable values, therefore then such test will must partition explanatory variable into several levels, assume there are ($n_i$) observations corresponding each level, where ($i$ = 1,2,..., m), then the total of sample observations equals to:–

$$n = \sum_{i=1}^{m} n_i$$

Assume the dependent variable correlated with explanatory variable by the following form:–

$$Y_{ij} = \beta_o + \beta_1 X_{ij} + u_{ij} \quad , \quad j = 1, 2, ..., n_i$$

The steps of this test are as follows:–

1. Determination of test hypotheses

$$H_o : \sigma_{u_1}^2 = \sigma_{u_2}^2 = \ldots = \sigma_{u_m}^2 \quad , or \ \sigma_{u_i}^2 = \sigma_u^2$$

indicate error variances of the partial samples that drawn from population homogeneous

$$H_1 : \sigma_{u_1}^2 \neq \sigma_{u_2}^2 \neq \ldots \neq \sigma_{u_m}^2 \quad , or \ \sigma_{u_i}^2 \neq \sigma_u^2$$

indicate error variances of the partial samples that drawn from population heterogeneous

2. Compute the calculated value for test statistic as follows;

$$Cal\ \chi^2 = \frac{Q}{L} \sim \chi^2_{(m-1)}$$

$$Q = n \ln \left( \frac{\sum_{i=1}^{m} n_i S_i^2}{n} \right) - \sum_{i=1}^{m} n_i \ln S_i^2 \quad , \quad L = 1 + \frac{1}{3(m-1)} \left( \sum_{i=1}^{m} \frac{1}{n_i} - \frac{1}{n} \right)$$

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \quad , \quad \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

3. Compare the calculated value with tabulated value for ($\chi^2_{m-1}$) with ($m$ – 1) degree of freedom and particular Significant level (α);

If; $$Cal\ \chi^2 \geq Tab\ \chi^2_{m-1} \quad , \quad we\ rejected\ H_o$$

This means that the error variances which computed from partial samples and which drawn from population are heterogeneous, i.e., there is Heteroscedasticity problem.

Either; $$Cal\ \chi^2 < Tab\ \chi^2_{m-1} \quad , \quad we\ not\ rejected\ H_o$$

This means that the error variances which computed from partial samples and which drawn from population are homogeneous (constant), i.e., there is no Heteroscedasticity problem.

# Remedial Tools of Heteroscedasticity

- Re – specification of the Model.
- Transformations (Log, …etc)
- The Generalized Least Squares Method(GLS)
  - We covered this in autocorrelation.
- Weighted Least Squares.
- Iteratively weighted least squares (IWLS).
- Whites's corrected standard errors.