CHAPTER FIVE: VIOLATIONS OF THE CLASSICAL ASSUMPTIONS

5.1 MULTICOLLINEARITY

a) The nature of the problem

One of the assumptions of the classical linear regression model (CLRM) is that there is no perfect multicollinearity among the regressors included in the regression model. Although the assumption is said to be violated only in the case of exact multicollinearity (i.e. an exact linear relationship among some of the regressors), the presence of multicollinearity (an approximate linear relationship among some of the regressors) leads to estimation problems.

Multicollinearity does not depend on any theoretical or actual linear relationship among any of the regressors; it depends on the existence of an approximate linear relationship in the data set at hand. Unlike most other estimation problems, this problem is caused by the particular sample available.

The existence of multicollinearity will affect seriously the parameter estimates. Intuitively, when any two explanatory variables are changing in nearly the same way, it becomes extremely difficult to establish the influence of each regressor on the dependent variable separately.

Consider the consumption–income model

$Consumption_i = \beta_0 + \beta_1 Income_i + \beta_2 Wealth_i + u_i$

It may happen that when we obtain data on income and wealth, the two variables may be highly, if not perfectly, correlated: Wealthier people generally tend to have higher incomes. Thus, although in theory income and wealth are logical candidates to explain the behavior of consumption expenditure, in practice (i.e., in the sample) it may be difficult to disentangle the separate influences of income and wealth on consumption expenditure.

Ideally, to assess the individual effects of wealth and income on consumption expenditure we need a sufficient number of sample observations of wealthy individuals with low income, and high-income individuals with low wealth. Although this may be possible in cross sectional studies (by increasing the sample size), it is very difficult to achieve in aggregate time series work.

In general, the problem of multicollinearity arises when individual effects of explanatory variables cannot be isolated and the corresponding parameter magnitudes cannot be determined with the desired degree of precision. Though it is quite frequent in cross section data as well, multicollinearity tends to be more common and more serious problem in time series data.

b) Sources of multicollinearity

1. *The data collection method employed,* for example, sampling over a limited range of the values taken by the regressors in the population.
2. *Constraints on the model or in the population being sampled.* For example, in the regression of electricity consumption on income and house size there is a physical constraint in the population in that families with higher incomes generally have larger homes than families with lower incomes.
3. *Model specification,* for example, adding polynomial terms to a regression model, especially when the range of the *X* variable is small.
4. *An overdetermined model.* This happens when the model has more explanatory variables than the number of observations. This could happen in medical research where there may be a small number of patients about whom information is collected on a large number of variables.
5. An additional reason for multicollinearity, especially in time series data, may be that the regressors included in the model share a *common trend,* that is, they all increase or decrease over time. Thus, in the regression of consumption expenditure on income, wealth, and population, the regressors income, wealth, and population may all be growing over time at more or less the same rate, leading to collinearity among these variables.

c) Consequences of Multicollinearity

i) The case of perfect multicollinearity

consider a three-variable regression model in deviation form

$$y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{u}_i$$

From which it can be obtained

$$\hat{\beta}_1 = \frac{\left(\sum x_{1i} y_i\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{2i} y_i\right)\left(\sum x_{1i} x_{2i}\right)}{\left(\sum x_{1i}^2\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{1i} x_{2i}\right)^2} \qquad \hat{\beta}_2 = \frac{\left(\sum x_{2i} y_i\right)\left(\sum x_{1i}^2\right) - \left(\sum x_{1i} y_i\right)\left(\sum x_{1i} x_{2i}\right)}{\left(\sum x_{1i}^2\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{1i} x_{2i}\right)^2}$$

Suppose $X_{2i} = kX_{1i}$ *where k is non zero constant.*

Then $\hat{\beta}_1 = \dfrac{\left(\sum x_{1i}y_i\right)\left(k^2\sum x_{1i}^2\right) - \left(k\sum x_{1i}y_i\right)\left(k\sum x_{1i}^2\right)}{\left(\sum x_{1i}^2\right)\left(k^2\sum x_{1i}^2\right) - k^2\left(\sum x_{1i}^2\right)^2} = \dfrac{0}{0}$

which is an indeterminate expression. It can also be shown that the expression for $\hat{\beta}_2$ is indeterminate.

Recall that $\hat{\beta}_1$ gives the rate of change in the average value of Y as X_1 changes by a unit, holding X_2 constant. But if X_1 and X_2 are perfectly collinear, there is no way X_2 can be kept constant: As X_1 changes, so does X_2 by the factor k. What it means, then, is that there is no way of disentangling the separate influences of X_1 and X_2 from the given sample.

Moreover, for a three variable model:

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$\text{Var}(\hat{\beta}_2) = \sigma^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

Substituting

$X_{2i} = kX_{1i}$

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \frac{k^2\sum x_1^2}{k^2\sum x_1^2 \sum x_2^2 - k^2\left(\sum x_1^2\right)^2} = \frac{\sigma^2\sum x_1^2}{0} = \infty$$

Which shows that the variances of the estimates become infinite.

Therefore, in the case of perfect multicollinearity, the regression coefficients remain *indeterminate* and their standard errors are *infinite*.

ii) High but imperfect multicollinearity

The perfect multicollinearity situation is a pathological extreme. Generally, there is no exact linear relationship among the X variables, especially in data involving economic time series.

Suppose in the three variable model

$X_{2i} = kX_{1i} + v_i$ where k≠ 0 *and* v_i is the stochastic error term such that $\sum x_{1i}v_i = 0$.

In this case, it is poosiible to estimate the coefficients. However, in the case of near or high multicollinearity, one is likely to encounter the following consequences:

1. Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult. For example, in a three variable linear regression,

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2(1-r_{12}^2)}$$

It is apparent from the above formula that as r_{12} (which is the coefficient of correlation between X_1 and X_2) tends towards 1, that is, as collinearity increases, the variance of the estimator increases. The same holds for Var($\hat{\beta}_2$) and the cov ($\hat{\beta}_1, \hat{\beta}_2$).

2. Because of consequence (1), the confidence intervals tend to be much wider, leading to the acceptance of the "Zero null hypothesis" (i.e., the true population coefficient is zero).
3. Because of consequence (1), the t-ratio of one or more coefficients tends to be statistically insignificant.
4. Although the t-ratio of one or more coefficients is statistically insignificant, R^2, the overall measure of goodness of fit, can be very high.
5. The OLS estimators and their standard errors can be sensitive to small changes in the data.

d) Detecting Multicollinearity

Note that multicollinearity is a *question of degree and not of a kind*. It should also be noted that since multicollinearity refers to the condition of the explanatory variables that are assumed to be nonstochastic, it is a *feature of the sample and not of the population*. Therefore, we do not "test for multicollinearity" but can, if we wish, measure its degree in any particular sample. The following are some rules of thumb and formal rules to detection of multicollinearity.

i) High R^2 but few significant t-ratios. If R^2 is high, say in excess of 0.8, the F-test in most cases will reject the hypothesis that the partial slope coefficients are simultaneously equal to zero, but the individual t tests will show that none or very few of the partial slope coefficients are statistically different from zero.

ii) High pair-wise correlation among regressors. If the pair-wise correlation coefficient among two regressors is high, say in excess of 0.8, then multicollinearity is a serious problem.

iii) Auxiliary Regression. Since multicollinearity arises because one or more of the regressors are exact or approximately linear combinations of the other regressors, one way of finding out which X variable is related to other X variables is to regress each X_i on the remaining X variables and compute the corresponding R^2. As a rule of thumb, multicollinearity may be a troublesome problem only if the R^2 obtained from an auxiliary regression is greater than the overall R^2 (that obtained from the regression of Y on all the regressors).

iv) Tolerance(TOL) and variance inflation factor (VIF)

For a linear regression with two explanatory variables(X_1 and X_2)

$$VIF = \frac{1}{(1-r_{12}^2)}$$

VIF shows how the variance of an estimator is inflated by the presence of multicollinearity. As r_{12}^2 approaches 1, the VIF approaches infinity. That is, as the extent of collinearity increases, the variance of an estimator increases, and in the limit it can become infinite. If there is no collinearity between X_1 and X_2, VIF will be 1.

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2(1-r_{12}^2)} = \frac{\sigma^2}{\sum x_{1i}^2} VIF$$

and

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-r_{12}^2)} = \frac{\sigma^2}{\sum x_{2i}^2} VIF$$

shows that the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ are directly proportional to the VIF.

Similarly, for *k*-variable model

$$\text{Var}(\hat{\beta}_i) = \frac{\sigma^2}{\sum x_i^2(1-R_i^2)} = \frac{\sigma^2}{\sum x_i^2} VIF_i$$

Where $\hat{\beta}_i$ =(estimated) partial regression coefficient of regressor *Xi*

R_i^2 = R^2 in the regression of *Xi* on the remaining regressions

It is also possible to use TOL as a measure of multicollinearity in view of its intimate connection with VIF.

$$TOL_i = \frac{1}{VIF_i} = (1-R_i^2)$$

The larger the value of the VIF_i, the more "troublesome" or collinear the variable X_i. As a rule of thumb, if the VIF of a variable exceeds 10, which will happen if R_i^2 exceeds 0.90, that variable is said to be highly collinear. In other words, the closer is TOL_i to zero, the greater the degree of collinearity of that variable with the other regressors. On the other hand, the closer TOL_i is to 1, the greater the evidence that X_i is not collinear with the other regressors.

e) Remedial Measures

The existence of multicollinearity in a data set does not necessarily mean that the coefficient estimators in which the researcher is interested have unacceptably high variance. Because multicollinearity is essentially a sample problem there are no infallible guides. However one can

try the following rules of thumb, the success of which depends on the severity of the collinearity problem.

a) ***Obtain more data***: - Because the multicollinearity is essentially a data problem, additional data that do not contain the multicollinearity feature could solve the problem. For example, in the three variable model we saw that

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2(1-r_{12}^2)}$$

Now as the sample size increases, $\sum x_{1i}^2$ will generally increases. Thus, for a given r_{12}, the variance of $\hat{\beta}_1$ decreases (thus, the standard error decreases), which will enable us to estimate β_1 more precisely.

b) ***Transformation of variables***: - In time series analysis, one reason for high multicollinearity between two variables is that over time both variables tend to move in the same direction. One way of minimizing then dependence is to transform the variables.

Suppose $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$

This relation must also hold at time t-1 because the origin of time is arbitrary anyway. Therefore we have

$$Y_{t-1} = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + U_{t-1}.$$

Subtracting this from the above gives

$$Y_t - Y_{t-1} = \beta_1(X_{1t} - X_{1t-1}) + \beta_2(X_{2t} - X_{2t-1}) + V_t$$

This is known as the *first difference* form because we run the regression, not on the original variables, but on the difference of successive values of the variables. The first difference regression model often reduces the severity of multicollinearity. Although the levels of X_1 and X_2 may be highly correlated, there is no a priori reason to believe that their difference will also be highly correlated.

Another commonly used transformation in practice is the **ratio transformation.**

Consider the model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

where *Y* is consumption expenditure in real dollars, *X*1 is GDP, and *X*2 is total population. Since GDP and population grow over time, they are likely to be correlated. One "solution" to this problem is to express the model on a per capita basis, that is,

$$\frac{Y_t}{X_{2t}} = \beta_0(\frac{1}{X_{2t}}) + \beta_1(\frac{X_{1t}}{X_{2t}}) + \beta_2 + (\frac{u_t}{X_{2t}})$$

Such a transformation may reduce collinearity in the original variables.

c) ***Formalize relationships among regressors***: - If it is believed that the multicollinearity arises not from an unfortunate data set but from an actual approximate linear relationship

among some of the regressors, this relationship could be formalized and the estimation could then proceed in the context of a simultaneous equation estimation problem.

d) **Combining cross-sectional and time series data:** that is, the combination of cross sectional and time-series data, known as pooling the data.

e) ***Drop a variable***: - when faced with severe multicollinearity, one of the "simplest" thing to do is to drop one of the collinear variables. But note that in dropping a variable from the model we may be committing a specification bias or specification error. Specification bias arises from incorrect specification of the model used in the analysis. Thus, if economic theory requires some variables to be included in the model, dropping one of the variables due to multicollinearity problem would constitute specification bias.

5.2 HETEROSCEDASTICITY

A) The Nature of Heteroscedasticity

The assumption of homoscedasticity (or constant variance) about the random variable u is that its probability distribution remains the same over all observations of X, and in particular that the variance of each U_i is the same for all values of the explanatory variable. That is, the variation of each u_i around its zero mean does not depend on the value of X. Symbolically we have

$$Var(u_i) = E\{(u_i - E(u_i)\}^2 = E(u_i^2) = \sigma^2$$

If the above condition is not satisfied in any particular case, we say that the u_i's are hetroscedastic. That is,$Var\ (u_i) = \sigma_i^2$.

The problem of heteroscedasticity is more serious in cross section data rather than time series data. Suppose we have a cross-section sample of family budget from which we want to measure the savings function. That means Saving = f(income). In this case, the assumption of constant variance of the u_i's is not appropriate, because high-income families show a much greater variability in their saving behavior than do low income families. Families with high income tend to stick to a certain standard of living and when their income falls they cut down their savings rather than their consumption expenditure. But this is not the case in low income families. Hence, the variance of u_i's increase as income increases.

B) Causes of Heteroscedasticity

Heteroscedasticity can arise as a result of several cases.

- **The presence of outliers:** An outlying observation, or outlier, is an observation that is much different (either very small or very large) in relation to the observations in the sample. More precisely, an outlier is an observation from a different population to that generating the remaining sample observations. The inclusion or exclusion of such an observation, especially if the sample size is small, can substantially alter the results of regression analysis.
- **Incorrect specification of the regression model:** Very often what looks like heteroscedasticity may be due to the fact that some important variables are omitted from the model. In such situation the residuals obtained from the regression may give the distinct impression that the error variance may not be constant. But if the omitted variables are included in the model, the impression may disappear.
- **Skewness in the distribution of one or more regressors included in the model**: Examples are economic variables such as income, wealth, and education. It is well known that the distribution of income and wealth in most societies is uneven, with the bulk of the income and wealth being owned by a few at the top.
- **Incorrect data transformation** (e.g., ratio or first difference transformations)**.**
- **Incorrect functional form** (e.g., linear versus log–linear models).

In summary, we may say that on a priori grounds there are reasons to believe that the assumption of homoscedasticity may often be violated in practice. It is therefore, important to examine the consequences of heteroscedaticity.

C) The consequence of Heteroscedasticity

If the assumption of homoscedastic disturbance is not fulfilled we have the following consequences:

i) Heteroscedasticity does not destroy the unbiasedness and consistency properties of OLS estimators.

ii) The OLS estimates do not have the minimum variance property in the class of unbiased estimators. That is, they are not BLUE.

iii) In the presence of heteroscedasticity, the variances of OLS estimators are not provided by the usual OLS formulas. But if we persist in using the usual OLS formulas, the *t* and *F* tests based on them can be highly misleading, resulting in erroneous conclusions.

D) Detecting the problem of Heteroscedasticity

As in the case of multicollinearity, there are no hard-and-fast rules for detecting heteroscedasticity, only a few rules of thumb.

i) Informal method

Nature of the problem

As a matter of fact, in cross-sectional data involving heterogeneous units, heteroscedasticity may be the rule rather than the exception. For example, in a cross-sectional analysis involving the investment expenditure in relation to sales, rate of interest, etc., heteroscedasticity is generally expected if small, medium and large-size firms are sampled together.

Visual Inspection of Residuals / graphical method

This is a postmortem approach when there is no a priori information as the existence of heteroscdasticity. Hence, this approach examines whether the error term depicts some systematic pattern or not. To this end, the residuals are plotted against the dependent or independent variable to which it is suspected the disturbance variance is related. Although $\hat{u}_i^2$ are not the same thing as u_i^2, they can be used as proxies especially if the sample size is sufficiently large.

ii) Formal methods

Park Test

Park formalizes the graphical method by suggesting that σ_i^2 is some function of the explanatory variable X_i.

The functional form he suggests is

$$Var(u_i) = \sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$$

which could be written in logarithmic form as

$$\ln\sigma_i^2 = \ln\sigma^2 + \beta \ln X_i + v_i$$

where v_i is the stochastic disturbance form.

Since σ_i^2 is generally not known, park suggests using $\hat{u}_i^2$ as a proxy and running the following regression:

$$\ln\hat{u}_i^2 = \ln\sigma^2 + \beta \ln X_i + v_i$$

$$= \alpha + \beta \ln X_i + v_i$$

If β turns out to be statistically significant, it would suggest that heteroscedasticity is present in the data. If it turns out to be insignificant, we may accept the assumption of homoscedasticity. The park test is thus a two-stage procedure. In the first stage we run the OLS regression disregarding the heteroscedasticity question. We obtain $\hat{u}_i$ from this regression, and then in the second stage we run the regression $\ln\hat{u}_i^2 = \ln\sigma^2 + \beta \ln X_i + V_i$.

Example: Consider a relationship between Compensation (Y) and Productivity (X). To illustrate the Park approach, the following regression function is used.

$$\boldsymbol{Yi = \beta_0 + \beta_1 X_i + u_i}$$

$$\mathbf{\hat{Y}_i = 1992.35 + 0.23\ X_i}$$

s.e = (936.48) (0.099)

t = (2.13) (2.33) $r^2 = 0.44$

Suppose that the residuals obtained from the above regression were regressed on X_i giving the following results.

$$\ln\hat{u}_i^2 = 35.82 - 2.81 \ln X_i$$

s.e = (38.32) (4.22)

t = (0.93) (-0.67) $r^2 = 0.46$

In the above result, the coefficient of $\ln X_i$ is not significant. That is, there is no statistically significant relationship between the two variables. Following the Park test, one may conclude that there is no heteroscedasticity in the error variance.

Although empirically appealing, the Park test has some problems. For instance, the error term, V_i may not satisfy the OLS assumptions and may itself be heteroscedastic. Nonetheless, as a strictly exploratory method, one may use the Park test.

Spearman's Rank Correlation Test

This test requires calculating rank correlation where its coefficient can be used to detect heteroscedasticity. The rank correlation coefficient is given by

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)}\right]$$

where d_i = difference in the ranks assigned to two different characteristics of the i^{th} individual or phenomenon and n = number of individuals or phenomena ranked. The steps required in this test are stated as follows.

Assume $Y_i = \beta_0 + \beta_1 X_i + u_i$

Step 1. Fit the regression to the data on Y and X and obtain the residuals $\hat{u}_i$

Step 2. Ignoring the sign of $\hat{u}_i$, that is, taking their absolute value $|\hat{u}_i|$, rank both $|\hat{u}_i|$ and X_i (or $\hat{Y}_i$) according to an ascending or descending order and compute the Spearman's rank correlation coefficient.

Step 3. Assuming that the population rank correlation coefficient ρ_s is zero and $n > 8$, the significance of the sample r_s can be tested by the t test as follows:

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

with $df = n - 2$

If the computed t value exceeds the critical t value, we may accept the hypothesis of heteroscedasticity; otherwise we may reject it. If the regression model involves more than one X variables, r_s can be computed between $|\hat{u}_i|$ and each of the X variable separately and can be tested for statistical significance by the t-test given.

Example To illustrate the rank correlation test consider the regression $Y_i = \beta_0 + \beta_1 X_i$. Suppose 10 observations are used to this equation. The following table makes use of the rank correlation approach to test the hypothesis of heteroscedasticity. Notice that column 6 and 7 put rank of $|\hat{u}_i|$ and X_i in an ascending order.

Table 4.1 Rank Correlation Test of Hetroscedasticity

Observ-ation	Y	X	$\hat{Y}$	$\hat{U}=(Y-\hat{Y})$	Rank of $\hat{U}i$	Rank of Xi	d (difference between the two ranking)	d^2
1	12.4	12.1	11.37	1.03	9	4	5	25
2	14.4	21.4	15.64	1.24	10	9	1	1
3	14.6	18.4	14.4	0.20	4	7	-3	9
4	16	21.7	15.78	0.22	5	10	-5	25
5	11.3	12.5	11.56	0.26	6	5	1	1
6	10.0	10.4	10.59	0.59	7	2	5	25
7	16.2	20.8	15.37	0.83	8	8	0	0
8	10.4	10.2	10.50	0.10	3	1	2	4
9	13.1	16.0	13.16	0.06	2	6	-4	16
10	11.3	12.0	11.33	0.03	1	3	-2	4
	TOTAL						0	110

Then,

$$r_s = 1 - 6\left[\frac{110}{10(100-1)}\right]$$

$$= 0.33$$

and

$$t = \frac{(0.33)\sqrt{8}}{\sqrt{1-0.11}}$$

$$= 0.99$$

Note that for 8 (=10-2) df, this t-value is not significant even at the 10% level of significance. Thus, there is no evidence of systematic relationship between the explanatory variable and the absolute value of the residuals, which might suggest that there is no heteroscedasticity.

The Goldfeld – Quandt Test

This popular method is applicable if one assumes that the heteroscedastic variance, σ_i^2 is positively related to *one* of the explanatory variables in the regression model. The test is commonly applicable to large samples. The observation must be at least twice as many as the parameters to be estimated. The test assumes normality and serially independent disturbance term, U_i's. Consider the following:

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots \beta_k X_{ki} + U_i$

The hypothesis to be tested is

H_0: U_i's are homoscedastic

H_1: Ui's are heteroscedastic (with increasing variance)

To test this, Goldfeld-Quandt perform the following steps.

Step 1: The observations are ordered according to the magnitude of the independent variable thought to be related to the variance of the disturbances.

Step 2: A certain number of central observations (represented by c) are omitted, leaving two equal-sized groups of observations, one group corresponding to low values of the chosen independent variable and the other group corresponding to high values. Note that the observations are omitted to sharpen or accentuate the difference between the small variance and the large variance group.

Step 3. We fit separate regression to each sub-sample, and we obtain the sum of squared residuals from each of them and the ratio of their sum of squared residuals is formed. That is,

$\sum \hat{U}_i^2$ = residuals form the sub-sample of low values of X_1 with [(n-c)/2] – k degrees of freedom, where k is the total number of parameters in the model.

$\sum \hat{U}_i^2$ = residual from the sub sample of high values of X, with the sample degree of freedom, [(n-c)/2] – k

If each of these sums is divided by the appropriate degrees of freedom, we obtain estimates of the variances of the $\hat{U}'s$ in the two sub samples.

Step IV : Compute the ratio of the two variances given by

$$F^* = \frac{\sum U_2^2 / [\{(n-c)/2\} - k]}{\sum \hat{U}_1^2 / [\{(n-c)/2\} - k]} = \frac{\sum \hat{U}_2^2}{\sum \hat{U}_1^2}$$

has an F distribution (with numerator and denominator each [{n-c-2k}/2] degrees of freedom, where n = total number of observations, c = central observations omitted, k = number of parameters estimated from each regression). If the two variances are the same (that is, if the $\hat{U}'s$ are homoscedasticc) the value of F^* will tend to one. If the variance differ, F^* will have a large value (given that by the design of the test $\sum \hat{U}_2^2 > \sum \hat{U}_1^2$. Generally, the observed F^* is compared with the theoretical value of F with (n-c-2k)/2 degrees of freedom (at a chosen level of significance. The theoretical value of F (obtained from the F-tables) is the value of F that defines the critical region of the test.

If F* > F we accept that there is heteroscedasticity (that is we reject the null hypothesis of no difference between the variances of U's in the two sub samples). If F^* < F, we accept that the U's are homoscedastic (in other words we accept the null hypothesis). The higher the observed F^* ratio the stronger the heteroscedasticity of the U's.

Example: Suppose that we have data on consumption expenditure in relation to income for a cross section of 30 families. Suppose we postulate that consumption expenditure is linearly related to income but that heteroscedasticity is present in the data. Suppose further that the middle 4 observations are dropped after the necessary reordering of the data. Suppose we obtain the following result after we perform a separate regression based on the two 13 observations.

$$F^* = \left[\frac{1536.8/11}{377.17/11} \right]$$

$$F^* = 4.07$$

Note from the F- table in the appendix that the critical F value for 11 numerator and 11 denominator df at the 5% level is 2.82. Since the estimated F* value exceeds the critical value, we may conclude that there is heteroscedasticity in the error variance.

Note, however, that the ability of the Goldfeld-Quadent test to perform successfully depends on how c is chosen. Moreover, its success depends on identifying the correct X (i.e., independent) variable with which to order the observations.

Breusch-Pagan(BP) test

The Breusch–Pagan test is a Lagrange multiplier test for heteroscedasticity. The main characteristics of Lagrange multiplier tests are that they do not require that the model is estimated under the alternative and that they are often simply computed from the R^2 of some auxiliary regression.

To illustrate this test, consider the k-variable linear regression model

$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + U_i$ ……………………………….. (1)

The Breusch-Pagan test for heteroscedasticity is carried out as follows:

1. Estimate equation (1) by OLS, and obtain the residual, $\hat{u}_i$.
2. Run the auxiliary regression $\hat{u}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + v$ and obtain the R-squared from this regression, $R^2_{\hat{u}^2}$.
3. Form either the F statistic: F = $\dfrac{R^2/(k-1)}{(1-R^2/(n-k)}$

   or the LM statistic: LM = n$R^2_{\hat{u}^2}$

   Under the null hypothesis, LM is distributed asymptotically as $\chi 2_{k-1}$, that is, chi-square with degrees of freedom equal to number of regressors (excluding the intercept). If the p-value is sufficiently small, that is, below the chosen significance level, then we reject the null hypothesis of homoscedasticity.

The White Test

White (1980) proposed a test for heteroscedasticity that adds the squares and cross products of all of the independent variables. The test is explicitly intended to test for forms of heteroscedasticity that invalidate the usual OLS standard errors and test statistics.

Suppose a model with three explanatory variables

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_2 + u$ ……………………………………………..(2)

Steps

1. Run regression on the Eq.(2) and obtain $\hat{u}$
2. Run the auxiliary regression:

   $\hat{u}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + \alpha_6 X_3^2 + \alpha_7 X_1 X_2 + \alpha_8 X_1 X_3 + \alpha_9 X_2 X_3 + u$

   and obtain $R^2_{\hat{u}^2}$.

The rationale of including the independent variables, their squares, and cross products is that the variance may be systematically correlated with either of the independent variables linearly or non-linearly.
The null hypothesis is that: $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_9$

3. Form LM statistic: $LM = n R^2_{\hat{u}^2} \sim \chi 2$ with nine degrees of freedom. If we fail to reject the null, we conclude that there is homoscedasticity. We can also use an F test of this hypothesis; both tests have asymptotic justification.

D) Remedial Measures

Heteroscedacticity does not destroy the unbiasedness and consistency properties of the OLS estimators, but they are no longer efficient, not even asymptotically (i.e., large sample size). This lack of efficiency makes the usual hypothesis testing procedure of dubious value. Therefore, remedial measures are clearly called for. There are two approaches to remediation: when σ_i^2 is known and when σ_i^2 is not known.

When σ_i^2 is Known: The Method of Weighted Least Squares (WLS)

If $\boldsymbol{\sigma_i^2}$ is known, the most straightforward method of correcting heteroscedasticity is by means of weighted least squares. The OLS method assigns equal weight or importance to each observation. However, a method of estimation known as **generalized least squares (GLS)** takes such information into account explicitly and is therefore capable of producing estimators that are BLUE. Though WLS is just a special case of the more general estimating technique, GLS, in the context of heteroscedasticity, one can treat the two terms WLS and GLS interchangeably. To see how this is accomplished, consider the familiar two-variable model:

$Y_i = \beta_0 + \beta_1 X_i + u_i$

$Y_i = \beta_0 X_{0i} + \beta_1 X_i + u_i$ where $X_{0i}=1$ for each i.

Dividing both sides by σ_i

$$\frac{Y_i}{\sigma_i} = \beta_0 (\frac{X_{0i}}{\sigma_i}) + \beta_1 (\frac{X_i}{\sigma_i}) + (\frac{u_i}{\sigma_i})$$

which for ease of exposition, can be written as

$Y_i^* = \beta_0^* + \beta 1^* Xi^* + u_i^*$

What is the purpose of transforming the original model? To see this, notice the following feature of the transformed error term u_i*

$$\text{Var}(u_i^*) = E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2} E(u_i)^2 \quad \text{since } \sigma_i^2 \text{ is known}$$

$$= \frac{1}{\sigma_i^2}\sigma_i^2 \quad \text{since } E(u_i)^2 = \sigma_i^2$$

$$= 1$$

which is a constant. That is, the variance of the transformed disturbance term u_i* is now homoscedastic. Since we are still retaining the other assumptions of the classical model, u_i* is homoscedastic suggests that if we apply OLS to the transformed model it will produce estimators that are BLUE. In short, the estimated β_0* and $\beta 1$* are now BLUE and not the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

To obtain GLS estimators, we minimize

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_0^* X_{0i}^* - \hat{\beta}_1^* X_i^*)^2$$

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_0^* \left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_1^* \left(\frac{X_i}{\sigma_i}\right) \right]^2$$

Which can also be written as:

$$\sum w_i \hat{u}_i^{\,2} = \sum w_i (Y_i - \hat{\beta}_0^* X_{0i} - \hat{\beta}_1^* X_i)^2$$

Where $w_i = \frac{1}{\sigma_i^2}$

Thus, in GLS we minimize a *weighted sum of residual squares* with *wi* = $1/\sigma_i^2$ acting as the weights, but in OLS we minimize non-weighted or (what amounts to the same thing) equally weighted RSS. In GLS, the weight assigned to each observation is inversely proportional to its σ_i, that is, observations coming from a population with larger σ_i will get relatively smaller weight and those from a population with smaller σ_i will get proportionately larger weight in minimizing the RSS.

When σ_i^2 Is Not Known: White's Heteroscedasticity-Consistent Standard Errors

If true σ_i^2 are known, we can use the WLS method to obtain BLUE estimators. Since the true σ_i^2 are rarely known, there is a way of obtaining *consistent* (in the statistical sense) estimates of the variances and covariances of OLS estimators even if there is heteroscedasticity. White has shown that this estimate can be performed so that *asymptotically* valid (i.e., large-sample) statistical inferences can be made about the true parameter values. White's heteroscedasticity-corrected standard errors are also known as **robust standard errors.**

Note: In addition to the above measures, a log transformation such as

$$\ln Y_i = \beta_0 + \beta_1 \ln X_i + U_i$$

very often such transformation reduces heteroscedasticity when compared with the regression

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

This result arises because log transformation compresses the scales in which the variables are measured. For example, log transformation reduces a ten-fold difference between two values (such as between 8 and 80) into a two-fold difference (because ln 80 = 4.32 and ln 8 = 2.08).

5.3 AUTOCORRELATION

a) The Nature of Autocorrelation

The term **autocorrelation** may be defined as "correlation between members of series of observations ordered in time [as in time series data] or space [as in cross-sectional data]. An important assumption of the classical linear model is that there is no autocorrelation or serial correlation among the disturbances entering into the population regression function. This assumption implies that:

$$\text{Cov}(U_i U_j) = E\{[Ui - E(Ui)]\,[Uj - E(Uj)]\}$$
$$= E(U_i U_j)$$
$$= 0 \text{ (for } i \neq j)$$

If this assumption is violated, the disturbances are said to be autocorrelated.

Since autocorrelated errors arise most frequently in time series models, the discussion in the rest of this chapter is couched in terms of time series data.

There are a number of time-series patterns or process that can be used to model correlated errors. The most common is what is known as "***the first order autoregressive process***" or ***AR(1)*** process.

Consider

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

where, t denotes data or observation at time t (i.e., a time series data) with this one can assume that the disturbances are generated as follows.

$$u_t = \rho u_{t-1} + \in_t$$

where ρ is known as the coefficient of autocovariance and $\in_t$ is the stochastic term such that it satisfies the standard OLS assumptions, namely

$$E(\in_t) = 0$$
$$\text{Var}(\in_t) = \sigma^2$$
$$\text{Cov}(\in_t, \in_{t+s}) = 0$$

where subscript 's' represent the exact period of lag.

The above specification is of first order because the regression of u_t is on itself lagged one period (where the coefficient ρ is the *first order coefficient of autocorrelation*). Note that the above specification postulates that the movement or shift in u_t consists of two parts: a part ρu_{t-1}, which accounts for systematic shift, and the other $\in_t$ which is purely random.

Relationships between u_t's can be shown as:

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

by substituting $u_t = \rho u_{t-1} + \in_t$ we obtain:

$= E[(\rho u_{t-1} + \in_t) u_{t-1}]$

$= \rho E[u^2_{t-1}] + E[\in_t u_{t-1}]$

Note that $E(\in_t) = 0$ thus $E(\in_t U_{t-1}) = 0$

Since with the assumption of homoscedasticity (i.e., constant variance) $Var(u_t) = Var(u_{t-1}) = \sigma^2$ the result would be

$$Cov(u_t, u_{t-1}) = \rho\sigma^2$$

Now, correlation of u_t, u_{t-1} is given by

$$Corr(U_t, U_{t-1}) = \frac{Cov(U_t, U_{t-1})}{\sqrt{Var(U_t)\, Var(U_{t-1})}} = \frac{\rho\sigma^2}{Var(U_t)}$$

$$= \frac{\rho\sigma^2}{\sigma^2} = \rho \qquad \text{where } -1 < \rho < 1$$

Hence, ρ(rho) is simple correlation of the successive errors of the original model.

Note that when $\rho > 0$ successive errors are positively correlated and when $\rho < 0$ successive errors are negatively correlated. It can be shown that $corr(U_t, U_{t-s}) = \rho^s$ (where s represents the exact period of lag). It implies that the correlation (be it negative or positive) between any two period diminishes as time goes by; i.e., as s increases.

b) Sources of Autocorrelation

- ***Inertia:*** A salient feature of most economic time series is inertia, or sluggishness. As is well known, time series such as GNP, price indices, production, employment, and unemployment exhibit (business) cycles. Starting at the bottom of the recession, when economic recovery starts, most of these series start moving upward. In this upswing, the value of a series at one point in time is greater than its previous value. Thus, there is a "momentum'' built into them, and it continues until something happens (e.g., increase in interest rate or taxes or both) to slow them down. Therefore, in regressions involving time series data, successive observations are likely to interdependent.
- **Data manipulation**: published data often undergo interpolation or smoothing, procedures that average true disturbances over successive time periods.
- **Specification bias**
  - **Specification Bias: Excluded Variables Case.** In empirical analysis the researcher often starts with a plausible regression model that may not be the most "perfect'' one. After the regression analysis, the researcher does the postmortem to find out whether the results accord with a priori expectations. For example, suppose we have the following demand model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

However, for some reason we run the following regression:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + v_t$$

Now if the first model is the "correct'' model, running the second is tantamount to letting $v_t = \beta_3 X_{3t} + u_t$. To the extent that X_3 affects Y_t, the error term v will reflect a systematic pattern, thus creating (false) autocorrelation. A simple test of this would be to run both models and see whether autocorrelation, if any, observed in the second model, disappears when the first model is run.

- ❖ **Specification Bias: Incorrect Functional Form.** Suppose the "true'' or correct model in a cost-output study is as follows:

  $$\textit{Marginal cost}_i = \beta_0 + \beta_1\, \textit{output}_i + \beta_2\, \textit{output}_i^2 + u_i$$

  But we fit the following model:

  $$\textit{Marginal cost}_i = \alpha_0 + \alpha_1\, \textit{output}_i + v_i$$

  Because the disturbance term v_i is, in fact, equal to $\textit{output}^2 + u_i$, it will catch the systematic effect of the $\textit{output}^2$ term on marginal cost. In this case, v_i will reflect autocorrelation because of the use of an incorrect functional form.

- **Lags.** For instance, in a time series regression of consumption expenditure on income, it is not uncommon to find that the consumption expenditure in the current period depends, among other things, on the consumption expenditure of the previous period. That is,
  $(\text{Consumption})_t = \beta_0 + \beta_1 (\text{income})_t + \beta_2 (\text{consumption})_{t-1} + u_t$

  A regression like this is known as **autoregression** because one of the explanatory variables is the lagged value of the dependent variable. The rationale is consumers do not change their consumption habits readily for psychological, technological, or institutional reasons. Now if we neglect the lagged consumption in the model, the resulting error term will reflect a systematic pattern due to the influence of lagged consumption on current consumption.

c) Consequences of Autocorrelation

i) As in the case of heteroscedasticity, in the presence of autocorrelation the OLS estimators are still linear unbiased as well as consistent and asymptotically normally distributed, but they are no longer efficient (i.e., minimum variance). As a consequence, the usual t, F, and $\chi 2$ tests cannot be legitimately applied.

ii) The prediction based on ordinary least squares estimate will be inefficient with autocorrelated errors. This is because of larger variance as compared with predictions based on estimates obtained from other econometric techniques.

d) Testing (Detecting) for Autocorrelation

Autocorrelation is potentially a series problem. Hence, it is essential to find out whether autocorrelation exists in a given situation. Since the population disturbances U_t, cannot be observed directly, we use its proxy, the residual $\hat{U}_t$ which can be obtained from the usual OLS procedure. The examination of $\hat{U}_t$ can provide useful information not only about autocorrelation but also about heteroscedasticity, model inadequacy, or specification bias.

i) Graphical Method

Some rough idea about the existence of autocorrelation may be gained by plotting the residuals either against time or against their own lagged variables.

For instance, suppose plotting the residual against its lagged variable bring about the following relationship.

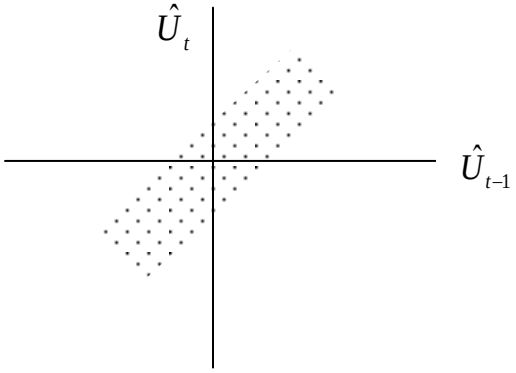


Figure 5.1 $\hat{U}_t$ ***and*** $\hat{U}_{t-1}$

As the above figure reveals, most of the residuals are bunched in the first and the third quadrants suggesting very strongly that there is positive correlation in the residuals. However, the graphical method is essentially subjective or qualitative in nature. There are quantitative tests that can be used to supplement the purely qualitative approach.

ii) Durbin-Watson d Test

The most celebrated test for detecting serial correlation is the one developed by Durbin and Watson. It is popularly known as the Durbin-Watson d-Statistic and it is defined as

$$d = \frac{\sum_{t=2}^{n}\left(\hat{U}_t - U_{t-1}\right)^2}{\sum_{t=1}^{n}\hat{U}_t^2}$$

which is simply the ratio of the sum of squared differences in successive residuals to the residual sum of squares, RSS. Note that in the numerator of the d statistic the number of observations is n-1 because one observation is lost in taking successive differences.

The proof of d-statistic is as follows:

$$d = \frac{\sum_{t=2}^{n}\left(\hat{U}_t^{\,2} + U_{t-1}^{\,2} - 2\hat{U}_t U_{t-1}\right)^2}{\sum_{t=1}^{n}\hat{U}_t^2} = \frac{\sum_{t=2}\hat{U}_t^2 + \sum_{t=2}\hat{U}_{t-1}^2 - \sum_{t=2}2\hat{U}_t\hat{U}_{t-1}}{\sum_{t=1}^{n}\hat{U}_t^2}$$

however, for large samples $\sum_{t=2}\hat{U}_t^2$, $\sum_{t=2}\hat{U}_{t-1}^2$ and $\sum_{t=1}^{n}\hat{U}_t^2$ are approximately equal.

Therefore, it can be written as

$$d \approx \frac{2\sum\hat{U}_{t-1}^2}{\sum\hat{U}_{t-1}^2} - \frac{2\sum\hat{U}_t\hat{U}_{t-1}}{\sum\hat{U}_{t-1}^2}$$

$$d \approx 2\left(1 - \frac{\sum\hat{U}_t\hat{U}_{t-1}}{\sum\hat{U}_{t-1}^2}\right)$$

$$d \approx 2(1-\hat{\rho})$$

The assumptions underlying the d-statistic are:

a) the regression model includes an intercept term
b) the explanatory variables are non-stochastic or fixed in repeated sampling
c) the disturbances U_t are generated by the first order autoregressive scheme. $U_t = \rho U_{t-1} + \in_t$. Therefore, it cannot be used to detect higher-order autoregressive schemes.
d) The error term u_t is assumed to be normally distributed.

e) the regression model does not include lagged value(s) of the dependent variable as one of the explanatory variables
f) There are no missing observations in the data.

In the absence of the autocorrelation we can expect to ***d*** take a value close to 2, when negative autocorrelation is present a value in excess of 2 and may be as high as 4, and when positive autocorrelation is present a value lower than 2 and may be close to zero.

The Durbin-Watson statistic tests the hypothesis that H_0: $\rho = 0$ (implying that the error terms are not autocorrelated with a first order scheme) against the alternative. However, the sampling distribution for the d-statistic depends on the sample size n, the number of explanatory variables and also on the actual sample values of the explanatory variables. Thus, the critical values at which we might, for example, reject the null hypothesis at 5 percent level of significance depend very much on the sample we have chosen. Unfortunately, it is impracticable to tabulate critical values for all possible sets of sample values. What possible however is, for given values of sample size and number of explanatory variables, to find upper and lower bounds such that actual critical values for any set of sample values will fall within these known limits.

The Durbin-Watson test procedure in testing the null hypothesis of $\rho = 0$ against the alternative hypothesis of positive autocorrelation is illustrated in the figure below.

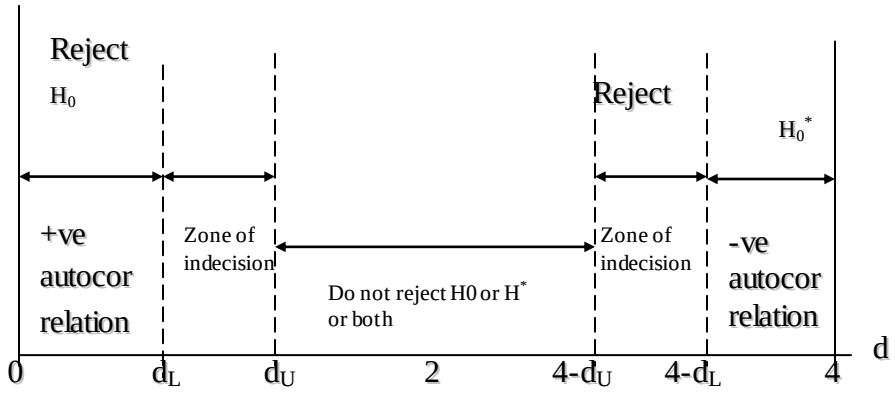
The decision criterion for the Durbin-Watson test is therefore, of the following form.

- for $d < d^*_L$ reject the null hypothesis of no autocorrelation in favor of positive autocorrelation;
- for $d > d^*_U$ do not reject null hypothesis, i.e., insufficient evidence to suggest positive autocorrelation;
- for $d^*_L < d < d^*_U$: inconclusive.

Because of the symmetry of the distribution it is also possible to use the tables for d^*_L and d^*_U to test the null hypothesis of no autocorrelation against the alternative hypothesis of negative autocorrelation, i.e. $\rho < 0$. The decision criterion then takes the form.

- for $d > 4 - d^*_L$ reject the null hypothesis of no autocorrelation in favor of negative autocorrelation.
- for $d < 4 - d^*_U$ do not reject null hypothesis, i.e., insufficient evidence to suggest negative autocorrelation
- for $4 - d^*_L > d > 4 - d^*_U$: inconclusive.

Note that tables for d^*_U and d^*_L are constructed to facilitate the use of one-tail rather than two tail tests. The following representation explains better the actual test procedure which shows that the limits of d are 0 and 4.



Note:
H_0: No positive autocorrelation
H_0^*: No Negative autocorrelation

Note that from the above presentation we can develop the following rule of thumb. That is, if d is found to be closer to 2 in an application, one may assume that there is no first order autocorrelation either positive or negative. If d is closer to 0 it is because ρ is closer to 1 indicating strong positive autocorrelation in the residuals. Similarly the closer d is to 4, the greater the evidence of negative serial correlation. This is because ρ is closer to –1.

Example: Suppose in a regression involving 50 observations with 4 regressors, the estimated d statistic was 1.43. From the Durbin Watson table we find that at the 5% level, the critical d value are d_L = 1.38 and d_U = 1.72. On the basis of the d test we cannot say whether there is positive autocorrelation or not because the estimated d value lies in the indecisive range.

iii) The Breusch–Godfrey (BG) Test

To avoid some of the pitfalls of the Durbin–Watson *d* test of autocorrelation, statisticians Breusch and Godfrey have developed a test of autocorrelation that is general in the sense that it allows for:

- nonstochastic regressors, such as the lagged values of the regressand
- higher-order autoregressive schemes, such as AR(1), AR(2), etc.; and
- simple or higher-order **moving averages** of white noise error terms

Consider $Y_t = \beta_0 + \beta_1 X_t + u_t$ …….(1)

Assume that the error term u_t follows the p^{th} order autoregressive, AR (p), scheme as follows:

$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_p u_{t-p} + \in_t$ *where* $\in_t$ *is a white noise error term.*

The null hypothesis H_0 to be tested is that:
$H_0: \rho_1 = \rho_2 = \ldots = \rho_p = 0$ ……..(2)

The BG test involves the following steps:

i) Estimate equation (1) by OLS and obtain the residuals, $\hat{u}_t$.

ii) Regress $\hat{u}_t$ on the original X_t (if there are more than one X variables in the original model, include them also) and $\hat{u}_{t-1}$, $\hat{u}_{t-2}$, . . . , $\hat{u}_{t-p}$, where the latter are the lagged values of the estimated residuals in step 1. Note that to run this regression we will have only (n − p) observations.

In short, run the following regression:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \ldots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

and obtain R^2 from this (auxiliary) regression.

iii) If the sample size is large , Breusch and Godfrey have shown that $(n-p)R^2 \sim \chi^2_p$.

If $(n - p)R^2$ exceeds the critical chi-square value at the chosen level of significance, we reject the null hypothesis, in which case at least one rho in equation(3) is statistically significantly different from zero. That is, there is autocorrelation.

A drawback of the BG test is that the value of *p*, the length of the lag, cannot be specified a priori.

e) Remedial Measures

With time series data, autocorrelated residuals are often indications of some error in the way we have specified the regression equation than genuine autocorrelation in the disturbances. Mostly, positive autocorrelations in economic data are caused by omission of relevant variables. Incorrect functional form may also be the cause for autocorrelated residuals.

Therefore, we should find out if the autocorrelation is pure autocorrelation and not the result of mis-specification of the model. If the source of the problem is suspected to be due to omission of

important variables, the remedy is to include those omitted variables. Besides if the source of the problem is believed to be the result of misspecification of the model, then the solution is to determine the appropriate mathematical form.

If it is pure autocorrelation, one can use appropriate transformation of the original model so that in the transformed model we do not have the problem of (pure) autocorrelation. As in the case of heteroscedasticity, we will use some type of generalized least-square (GLS) method. In large samples, the Newey–West method can be applied to obtain standard errors of OLS estimators that are corrected for autocorrelation.

I) The method of generalized least squares (GLS)

Knowing the consequences of autocorrelation, especially the lack of efficiency of OLS estimators, we may need to remedy the problem. The remedy depends on the knowledge one has about the nature of interdependence among the disturbances, that is, knowledge about the structure of autocorrelation.

Consider the two-variable regression model:

$Y_t = \beta_0 + \beta_1 X_t + U_t$ …… *(i)*

And assume that the error term follows the AR(1) scheme, namely

$U_t = \rho U_{t-1} + \epsilon_t$ *where* $-1 < \rho < 1$

There are two cases:

- when ρ is known and
- when ρ is not known but has to be estimated.

When ρ *is known*

The serial correlation problem can be satisfactorily resolved if the coefficient of autocorrelation, ρ, is known.

Recall the two variables model.

$Y_t = \beta_0 + \beta_1 X_t + U_t$ …… *(i)*

At time t-1 the above model will be

$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + U_{t-1}$ ……..*(ii)*

Multiplying both sides by ρ, we obtain

$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho U_{t-1}$ ……..*(iii)*

Subtracting (iii) from (i) gives

$$(Y_t - \rho Y_{t-1}) = (\beta_0 - \rho\beta_0) + (\beta_1 X_t - \rho\beta_1 X_{t-1}) + (U_t - \rho U_{t-1})$$

$$= \beta_0 (1-\rho) + \beta_1 (X_t - \rho X_{t-1}) + \epsilon_t \quad \text{.....(iv)}$$

where $\epsilon_t = U_t - \rho U_{t-1}$

The transformed model can be expressed as

$$Y^*_t = \beta^*_0 + \beta^*_1 X^* t + \epsilon_t \quad \text{.....(v)}$$

Where $Y^*_t = Y_t - \rho Y_{t-1}$, $\beta^*_0 = \beta_0 - \rho\beta_0$, $X^*_t = (X_t - \rho X_{t-1})$ and $\beta^*_1 = \beta_1$

Since ϵ_t satisfies the OLS assumptions, one can apply OLS to the transformed variables Y^* and X^* and obtain estimators with all the optimum properties, namely BLUE. Regression of the transformed model is equivalent to using generalized least squares (GLS). In short, GLS is nothing but OLS applied to the transformed model that satisfies the classical assumptions. Regression of equation (iv) is known as the ***generalized,*** or ***quasi, difference equation***. It involves regressing *Y* on *X*, not in the original form, but in the ***difference form***, which is obtained by subtracting a proportion (= ρ) of the value of a variable in the previous time period from its value in the current time period. Note that in this differencing procedure we lose one observation because the first observation has no antecedent.

When ρ is not known

Although straight forward to apply, the method of generalized difference is difficult to run because, ρ, population correlation coefficient is rarely known in practice. Therefore, alternative methods need to be devised.

The First-Difference Method. Since ρ lies between 0 and ±1, one could start from two extreme positions. At one extreme, one could assume that $\rho = 0$, that is, no (first-order) serial correlation, and at the other extreme we could let $\rho = \pm 1$, that is, perfect positive or negative correlation.
As a matter of fact, when a regression is run, one generally assumes that there is no autocorrelation and then lets the Durbin–Watson or other test show whether this assumption is justified. If, however, $\rho = +1$, the generalized difference equation in equation (iv) above reduces to the *first-difference equation:*

$$(Y_t - Y_{t-1}) = \beta 1 (Xt - X_{t-1}) + (U_t - \rho U_{t-1})$$

$$\Delta Y t = \beta 1 \, \Delta X_t + \epsilon_t$$

The first difference transformation may be appropriate if the coefficient of autocorrelation is very high, say in excess of 0.8, or the Durbin–Watson *d* is quite low. Strictly speaking, the first-difference transformation is valid only if $\rho = 1$. Maddala has proposed this rough rule of thumb: *Use the first difference form whenever* $d < R^2$. An interesting feature of the first-difference model is that there is no intercept in it. Hence, we have to use the *regression through the origin*.

Computing ρ from Durbin–Watson d Statistic. If we cannot use the first difference transformation because ρ is not sufficiently close to unity, we have an easy method of estimating it from the relationship between d and ρ as follows:

$$\rho \approx 1 - \frac{d}{2}$$

Thus, in reasonably large samples one can obtain rho and use it to transform the data as shown in the generalized difference equation. However, the relationship between ρ and d may not hold true in small samples.

Estimating ρ from the residuals. If the AR(1) scheme $u_t = \rho u_{t-1} + \varepsilon_t$ is valid, a simple way to estimate rho is to regress the residuals $\hat{u}_t$ on $\hat{u}_{t-1}$, for $\hat{u}_t$ the are consistent estimators of the true u_t. That is, we run the following regression:

$$\hat{u}_t = \rho\, \hat{u}_{t-1} + v_t$$

where, $\hat{u}_t$ are the residuals obtained from the original (level form) regression and where v_t are the error term of this regression. Note that there is no need to introduce the intercept term, because the OLS residuals sum to zero.

Iterative Methods of Estimating ρ. All the methods of estimating ρ explained above provide us with only a single estimate of ρ. But there are the so-called *iterative methods* that estimate ρ iteratively, that is, by successive approximation, starting with some initial value of ρ. Among these methods are:

- Cochrane–Orcutt iterative procedure,
- Cochrane–Orcutt two-step procedure,
- Durbin two–step procedure, and
- Hildreth–Lu scanning or search procedure.

Of these, the most popular is the *Cochrane–Orcutt* iterative method. One advantage of this method is that it can be used to estimate not only an AR(1) scheme, but also higher-order autoregressive schemes. Having obtained the two rhos, one can easily extend the generalized difference equation.

The Cochrane-Orcutt interative Procedure

This procedure helps to estimate ρ from the estimated residuals $\hat{U}_t$ so that information about the unknown ρ will be derived.

To explain the method, consider the two-variable model

$$Y_i = \beta_0 + \beta_1 X_i + U_i \quad \text{..... (a)}$$

and assume that U_t is generated by the AR(1) scheme namely

$$U_t = \rho U_{t-1} + \epsilon_t \quad \text{..... (b)}$$

Cochrane and Orcutt then recommended the following steps to estimate ρ:

Step 1: Estimate the two variables model by the standard OLS routine and obtain the residuals $\hat{U}_t$

Step 2: Using the estimated residuals, run the following regression

$$\hat{U}_t = \hat{\rho}\hat{U}_{t-1} + V_t$$

Step 3: Using $\hat{\rho}$ obtained from step 2 regression, run the generalized difference equation as follows

$$(Y_t - \hat{\rho} Y_{t-1}) = \beta_0 (1 - \hat{\rho}) + \beta_1 (X_t - \hat{\rho} X_{t-1}) + (U_t - \hat{\rho} U_{t-1})$$

or $Y^*_t = \beta^*_0 + \beta^*_1 X^*_t + \hat{U}^*_t$

Step 4: Since a priori it is not known that the $\hat{\rho}$ obtained from the regression in step 2 is the best estimate of ρ, substitute the values of $\hat{\beta}^*_0$ and $\hat{\beta}^*_1$ obtained from the regression in step 3 into the original regression (a) and obtain the new residuals, say $\hat{U}^{**}_t$ as

$$\hat{U}^{**}_t = Y_t - \hat{\beta}^*_0 - \hat{\beta}^*_1 X_t$$

Note that this can be easily computed since Y_t, X_t, $\hat{\beta}^*_0$ and $\hat{\beta}^*_2$ are all known.

Step 5: Now estimate this regression

$$\hat{U}^{**}_t = \hat{\hat{\rho}}\hat{U}^{**}_{t-1} + W_t$$

Where $\hat{\hat{\rho}}$ is the second round estimate of ρ.

Since we do not know whether this second round estimate $\hat{\hat{\rho}}$ is the best estimate of ρ, we can go into the third estimate, and so on. That is why the Cochrane-Orcutt method is said iterative. But how long should we go on? The general procedure is to stop carrying out iterations when the successive estimates of ρ converges. Thus, we select that chosen ρ to transform the model and apply a kind of GLS estimation that minimizes the problem of autocorrelation.

Note that:

1. Since the OLS estimators are consistent despite autocorrelation, in large samples, it makes little difference whether we estimate ρ from the Durbin–Watson d, or from the regression of the residuals in the current period on the residuals in the previous period, or from the Cochrane–Orcutt iterative procedure because they all provide consistent estimates of the true ρ.

2. The various methods discussed above are basically two-step methods. In step 1 we obtain an estimate of the unknown ρ and in step 2 we use that estimate to transform the variables to estimate the generalized difference equation, which is basically GLS. But since we use $\hat{\rho}$ instead of the true ρ, all these methods of estimation are known in the literature as *feasible GLS (FGLS)* or *estimated GLS (EGLS)* methods.

3. It is important to note that whenever we use an *FGLS* or *EGLS* method to estimate the parameters of the transformed model, the estimated coefficients will not necessarily have the usual optimum properties of the classical model, such as BLUE, especially in small samples. In short, whenever we use an estimator in place of its true value, the estimated OLS coefficients may have the usual optimum properties asymptotically, that is, in large samples. Also, the conventional hypothesis testing procedures are, strictly speaking, valid asymptotically. In small samples, therefore, one has to be careful in interpreting the estimated results.

4. In using EGLS, if we do not include the first observation (as was originally the case with the Cochrane Orcutt procedure), not only the numerical values but also the efficiency of the estimators can be adversely affected, especially if the sample size is small and if the regressors are not strictly speaking nonstochastic. Therefore, in small samples it is important to keep the first observation *à la* Prais Winsten.

II) The Newey–West method of correcting the OLS standard errors

Instead of using the FGLS methods, we can still use OLS but correct the standard errors for autocorrelation by a procedure developed by Newey and West. This is an extension of White's heteroscedasticity-consistent standard errors. The corrected standard errors are known as **HAC (heteroscedasticity- and autocorrelation-consistent) standard errors** or simply as **Newey–West standard errors.**

This method is strictly speaking valid in large samples and may not be appropriate in small samples. Therefore, if a sample is reasonably large, one should use the Newey–West procedure to correct OLS standard errors not only in situations of autocorrelation only but also in cases of heteroscedasticity, for the HAC method can handle both, unlike the White method, which was designed specifically for heteroscedasticity.