Chapter 15

Problem of Autocorrelation in R

15.1 Nature of Autocorrelation

In empirical research works, when we are dealing with time series data, a very common problem we face is that the observations for a particular variable are likely to correlate among themselves, especially if the time interval between successive observations is short such as a day, a week or even a month. For example, stock price indexes such as S&P 500, BSE etc over a period of time move up or down for a period of time in succession and there found to have strong correlation between these movement.

In this chapter we will try to focus on the following points:

i. what is the problem of autocorrelation?
ii. What are the consequences of autocorrelation?
iii. How to detect the problem of autocorrelation?
iv. How to get out of the autocorrelation problem?

Autocorrelation is a special case of correlation and refers not to the relationship between two or more variable, but to the relationship between successive values of the same variable. One of the basic assumptions in the regression model is that the value of the successive disturbance term is independent of each other, so that

$$cov(u_i, u_j) = 0 \quad for \quad i \neq j$$
$$E(u_i, u_j) = 0 \quad for \quad i \neq j$$

If this assumption is violated, that is, the value of *u* in any particular period is correlated with its own preceding values, it is termed as autocorrelation or serial correlation of the error term.

Autocorrelation is found to occur generally in time series data. So, we will use the subscript *t* denoting time instead of *i* as used in earlier chapters.

Now the question arises that why does autocorrelation occur in time series data? Some of the possible reasons are:

i. Most of the time series data such as Gross Domestic Product, price indexes, production, employment, and unemployment exhibit business cycles. Al of these variables after starting at the bottom of the recession, starts moving with economic recovery. In this upswing phase of business cycle, the value of a series at one point in time is greater than its previous value which automatically in-cult a ‘momentum’ into them, and it continues until something happens to slow them down. Thus, in the time series data there is a built-in tendency to correlate among themselves.
ii. Omitted Explanatory Variables: If this occurs, since it is known that most economic variables are auto-correlated, then the error term will also be auto-correlated.
iii. Misspecification of the mathematical form of the model can also create autocorrelation problem in the time series analysis.

To throw more light on the autocorrelation, we start with the following two variable model:

$$Y_t = \beta_o + \beta_1 X_t + u_t \qquad where \quad t = \text{time perdiod running from } 1 \text{ to } n$$

A measure of the first order autocorrelation coefficient can be obtained from the regression residuals using the following formula

$$r_{u_t u_{t-1}} = \frac{\sum u_t u_{t-1}}{\sqrt{u_t^2}\sqrt{u_{t-1}^2}}$$

$$\text{Or, } r_{\hat{u}_t \hat{u_{t-1}}} = \frac{\sum \hat{u}_t \hat{u_{t-1}}}{\sqrt{\hat{u}_t^{\,2}}\sqrt{\hat{u_{t-1}}^{\,2}}}$$

An estimate of the true correlation coefficient is denoted by $\hat{\rho}_{u_t u_{t-1}}$ which measures the correlation of the true population of *u's*. Sometimes u_t and $\hat{u}_t$ are used interchangeably.

Several forms of autocorrelation is possible, the simplest being

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

this is first order autocorrelation. The parameter ρ measures the correlation between u_t and u_{t-1} and v_t is another random term that *iid.*

Now if $\rho > 1$ then it is called the case of positive autocorrelation whereas if $\rho < 1$ it is called the case of negative autocorrelation.

Higher order autocorrelation is also possible. For example, the second order autocorrelation is written as

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$$

15.2 OLS Estimation in the Presence of Autocorrelation

How the presence of autocorrelation affects the OLS estimates in multiple regression model? To answer this, we start with the simplest form of two variable model with first order autocorrelation.

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad .....(15.1)$$

with $u_t = \rho u_{t-1} + v_t$, $|\rho| < 1$ and v_t satisfying al the assumptions of a random variable ..(15.2)

where $\rho =$ Coefficient of autocovariance

$$E(v_t) = 0$$

$$var(v_t) = \sigma_v^2$$

$$cov(v_t, v_{t+s}) = 0, \quad s \neq 0$$

So, equation (15.2) tells us that the value of u_t in period t is equal to ρ times its value in the previous period plus a purely random error term. The scheme (15.2) is known as Markov's first order autoregressive scheme and denoted by AR(1).

Given the AR(1) scheme, it can be shown that

$$var(u_t) = E(u_t^2) = \frac{\sigma_v^2}{1-\rho^2} \qquad ....(15.3)$$

$$cov(u_t, u_{t+s}) = E(u_t, u_{t+s}) = \rho^2 \frac{\sigma_v^2}{1-\rho^2} \qquad ....(15.4)$$

$$cov(u_t, u_{t+s}) = \rho^2 \qquad ....(15.5)$$

As ρ is constant and it lies between -1 to +1, we see from equation (15.3) that u_t is still homoscedastic, but u_t is correlated with its past values.

It is important to note that $|\rho| < 1$.

When $\rho = 1$, the variance and covariance as given by equation (15.3) and (15.4) cannot be defined.

When $|\rho| < 1$, the AR(1) process as defined by (15.1) is stationary and so, its mean and variance will remain constant over time and the covariance will go on dominising as we go into distinct past.

Now let us come back to two variable model,

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

The OLS estimators of the parameter β_1 are given as:

$$\beta_1 = \frac{\sum x_t y_t}{\sum x_t^2} \qquad ....(15.6)$$

$$var(\beta_1) = \frac{\sigma^2}{\sum x_t^2} \qquad ....(15.7)$$

But when the disturbance term follows AR(1) process, the variance of β_1 can be written as:

$$var(\hat{\beta}_1)_{AR1} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + ...+ 2\rho^{n-1}\frac{\sum x_1 x_n}{\sum x_t^2}\right] \qquad ....(15.8)$$

By comparing the equation (15.7) and (15.8), it is clear that the former is equal to the latter times a term which depends on ρ as well as the sample autocorrelations between the values taken by the regressor X at different lags.

Also in general, it is not possible to tell whether $var(\hat{\beta}_1)$ is less than or greater than $var(\hat{\beta}_1)_{AR1}$.

However, if ρ = 0, both the estimates are equal to each other.

Also if the correlations among the successive values of regressors are very small, the usual OLS variance of the slope estimator will not be seriously biased. But in general, the two variances will not be the same[1].

Thus, the consequences of autocorrelation can be summed as under:

1. The OLS estimators are still unbiased and consistent.
2. They are still normally distributed in large samples.
3. But they are no longer efficient. That is, they are no longer BLUE (best linear unbiased estimator). In most cases OLS standard errors are underestimated, which means the estimated *t* values are inflated, giving the appearance that a coefficient is more significant than it actually may be.
4. As the residual variance is likely to be underestimated, it is likely to overestimate R^2.
5. Consequently, the usual *t* and *F* tests of significance will no longer be valid, and if applied, they are likely to produce seriously misleading conclusions about the statistical significance of the estimated regression coefficients.

15.3 Testing for Autocorrelation

(I) Graphical Method

The assumption of autocorrelation relates to the residuals of the population regression function. However, the population residuals are not directly measurable. Instead, we can use their proxies, that is, sample residuals $\hat{u}_t$, which are obtained by the OLS method. A visual inspection of the sample residuals $(\hat{u}_t)$ will enable us to make judgment about the presence of autocorrelation as well as heteroscedasticity. There are several ways of plotting the residuals.

1. Plot the residuals against time and observe the pattern.
2. Plot the standardized residuals against time and observe the pattern. Standardized residuals are simply the residuals divided by the standard error of regression, that is, $\hat{u}_t/\hat{\sigma}$.

If the scatter plot shows any specific pattern, then there is autocorrelation in the data.

1 Basic Econometrics by Damodar Gujarati

(II) Durbin-Watson d Test

Durbin and Watson have developed a test on the basis of assumption that the error term in the regression model are generated by a first-order autoregressive process observed at equally spaced time period, that is,

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

Durbin and Watson have used the following test statistic

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^{\,2}} \qquad \ldots.(15.9)$$

which is simply the ratio of the sum of squared differences in successive residuals to the 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.

Before proceeding further, let us look at the assumptions on which the Durbin-Watson test is based on. These are:

i. The regression model must include the intercept term.
ii. The explanatory variables are non-stochastic in nature.
iii. The disturbance term follows the first order autoregressive scheme only.
iv. The error term is assumed to follow Normal distribution.
v. The regression model should not include any lagged values of the dependent variable.
vi. There is no missing observation in the data set.

Now the Durbin-Watson's d statistic can be rewritten as

$$d = \frac{\sum \hat{u}_t^{\,2} + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^{\,2}} \qquad \cdots(15.10)$$

Since $\sum \hat{u}_t^{\,2}$ and $\sum \hat{u}_{t-1}^2$ differ only in one observation in the sequence of the data, we may consider that they are approximately equal. So,

$$d \approx 2\left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^{\,2}}\right) \qquad \cdots(15.11)$$

We can define the first-order autocorrelation coefficient as

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \qquad \ldots.(15.12)$$

Substitution of (15.12) into (15.11) yields

$$d \approx 2(1 - \hat{\rho}) \qquad \cdots(15.13)$$

Since $-1 \le \rho \le 1$, equation (15.13) implies that

$$0 \le d \le 4$$

Thus, the values of d statistic will lie between 0 and 4.

More specifically, when there is no serial correlation, $\hat{\rho} = 0$, and from equation (15.13), we get that $d = 2$.

therefore, as a rule of thumb, if in any application, d is found to be 2, we can assume that there is no first-order autocorrelation.

If $\hat{\rho} = +1$, indicating perfect positive correlation in the residuals, d ≈ 0. Therefore, the closer d is to 0, the greater the evidence of positive serial correlation.

If $\hat{\rho}=-1$, that is, there is perfect negative correlation among successive residuals, d ≈ 4. Hence, the closer *d* is to 4, the greater the evidence of negative serial correlation.
The distribution of *d* is reported by Durbin and Watson(1951). To perform the Durbin-Watson test, we define the critical values of *d.* The hypotheses are framed as

$$H_0: \rho=0, \ (\textit{No autocorrelation})$$
$$\textit{against}$$
$$H_1: \rho\neq 0, \ (\textit{Autocorrelation of first-order})$$

The distribution tables can be obtained in most of text book. The following table shows the critical regions for given values of d_L and d_U which depends on the sample size and number of explanatory variables in the model. The tabulated critical values allow the following inferences to be made the population parameter ρ.

Table 15.1: Regions of Acceptance and Rejection of the Null Hypothesis

$0<d<d_L$	$d_L\leq d\leq d_U$	$d_U<d<4-d_U$	$4-d_U\leq d\leq 4-d_L$	$4-d_L<d<4$
Reject H_0 Positive autocorrelation	No decision about the presence of autocorrelation	Accept H_0 No autocorrelation (Positive or negative)	No decision about the presence of autocorrelation	Reject H_0 Negative autocorrelation

The Durbin–Watson *d* test has become so venerable that practitioners often forget the assumptions underlying the test. In particular, the assumptions that

(1) the explanatory variables, or regressors, are non-stochastic;

(2) the error term follows the normal distribution;

(3) the regression models do not include the lagged value(s) of the regressand; and

(4) only the first-order serial correlation is taken into account are very important for the application of the *d* test. It should also be added that a significant d statistic may not necessarily indicate autocorrelation. Rather, it may be an indication of omission of relevant variables from the model.

If a regression model contains lagged value(s) of the regressand, the *d* value in such cases is often around 2, which would suggest that there is no (first-order) autocorrelation in such models. Thus, there is a built-in bias against discovering (first-order) autocorrelation in such models. This does not mean that autoregressive models do not suffer from the autocorrelation problem. As a matter of fact, Durbin has developed the so-called *h* test to test serial correlation in such models. But this test is not as powerful, in a statistical sense, as the Breusch–Godfrey test, so there is no need to use the *h* test.

(III) The Breusch–Godfrey (BG) Test
The Durbin-Watson test described above though very simple and extremely popular, it has one great drawback in that, if it falls in the indecisive zone, one cannot conclude that (first-order) autocorrelation does or does not exit. exist. Also it cannot detect the autocorrelation of higher order. To avoid these drawbacks of the Durbin–Watson test of autocorrelation, statisticians Breusch and Godfrey have developed a general test of autocorrelation which will allow for

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

To illustrate the BG test, we start with the following two-variable model.

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad ....(15.1)$$

For autocorrelation of order *p*, we have

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + v_t \qquad \cdots(15.14)$$

The hypotheses are framed as:

$$H_0 : \rho_1 = \rho_2 = ... = \rho_p$$

against

$$H_1 : \textit{Not all } \rho_s \textit{ are zero}$$

The steps in the B-G test are as under:

Step 1: Estimate the model (1) by OLS and obtain the residuals, $\hat{u}_t$

Step 2: Regress $\hat{u}_t$ on the original X values and lagged values of the residuals

$\hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-p}$,i.e., we run the following regression

$$\hat{u}_t = \alpha_0 + \alpha_1 X_t + \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + v_t \qquad .....(15.15)$$

Note down the value of R^2 from this auxiliary regression.

Step 3: When the sample size becomes large, Breusch-Godfrey have shown that

$$(n-p) R^2 \sim \chi_p^2 \qquad ....(15.16)$$

If in an application, $(n-p) R^2$ exceeds the critical value of chi-square at the chosen level of significance, it is wise to reject the null hypothesis, in which case at least one ρ in equation (15.15) is statistically significantly different from zero.

The following points should be remembered by the researchers while applying the B-G test for autocorrelation:

i. The regressors included in the regression model may contain lagged values of the regressand Y, that is, Y_{t-1}, Y_{t-2} etc., may appear as explanatory variables. Contrast this model with the Durbin-Watson test restriction that there may be no lagged values of the regressand among the regressors.

ii. As noted earlier, the B-G test is applicable even if the disturbances follow a p-th order moving average (MA) process, that is, when the disturbance term is generated as:

$$u_t = \epsilon_1 + \lambda_1 \epsilon_{t-1} + \lambda_2 \epsilon_{t-2} + ... + \lambda_p \epsilon_{t-p} \qquad ....(15.17)$$

where ϵ_t is the white noise term with OLS assumptions.

iii. A drawback of the B-G test is that the value of p, the length of the lag, cannot be specified a priori. Some experimentation with the p value is inevitable. Researchers can also use the so-called Akaike and Schwarz information criteria to select the lag length, that is, the model with lowest Akaike and Schwarz information will be selected.

15.4 Remedial Measures for Autocorrelation

In the previous section we have seen the consequences of autocorrelation in the context of OLS estimation of the model parameters. It is, therefore, necessary to find the remedial measures for this autocorrelation problem.

The remedial measures will be easy once the nature of autocorrelation is known to us. For example, if there is first-order autocorrelation and the autocorrelation coefficient (ρ) is known to us, then the solution becomes very easy. However, in most of the cases, the autocorrelation coefficient is not known to us. Anyway let us first see the remedial measures when there is first-order autocorrelation

and the autocorrelation coefficient (ρ) is known. We take the case of simple two-variable model, though it can be easily extended for multiple variable case.

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad ....(15.1)$$

When the above equation (15.1) holds true for time period *t*, it will also true hold for the period $(t-1)$. So, we can write

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1} \qquad ....(15.18)$$

Now multiplying both sides of equation (15.18) by the autocorrelation coefficient (ρ) and subtracting it from (15.1), we get

$$(Y_t - \rho Y_{t-1}) = (1-\rho)\beta_0 + \beta_1 (X_t - \rho X_{t-1}) + v_t \qquad \textit{where} \quad v_t = u_t - u_{t-1} \qquad ....(15.20)$$

Further, the equation (15.20) can also be expressed as

$$Y_t^* = \beta_0^* + \beta_1^* X_t^* + v_t \qquad ....(15.21)$$

where $Y_t^* = (Y_t - \rho Y_{t-1})$, $\beta_0^* = (1-\rho)\beta_0$, $X_t^* = (X_t - \rho X_{t-1})$, $\beta_1^* = \beta_1$

As the error term in equation (15.21) satisfies all the usual OLS assumptions, we can apply OLS to the transformed equation involving the variables Y^* and X^* and obtain estimators which will satisfy the BLUE properties.

Applying OLS to the transformed model of equation (15.21 or 15.20) is known as Generalized Regression or differences equation, as it involves in regressing *Y* on *X*, not in the original form, but in the difference form.

The transformation involves in subtracting a proportion (ρ) of the value of the variable in the previous time period from its value in the current time period. However, in this procedure of transformation, we will loose one observation at the beginning of time series data-set. Prais-Winston have developed one method for avoiding the loss of observation.

To preserve the first observation $(t=1)$ in the model, the first observation on *Y* and *X* is transformed as follows:

$$Y_t^* = \sqrt{1-\rho^2}\, Y_1$$
$$X_t^* = \sqrt{1-\rho^2}\, X_1$$

The above procedure is conceptually very simple and straightforward. But in practice we face the problem of applying this method, as the autocorrelation coefficient (ρ) is not known to us. So, we need to find the ways of estimating the autocorrelation coefficient (ρ).

A Priori Information on ρ:

In most of the empirical research works, the researcher makes an educated guess about the value of the autoregressive coefficient(s) (ρ) on the basis of the his knowledge and intuition about the relationship between the variables in question. Most often the researcher assumes that $\rho = 1$. Then we can proceed as follows:

We start with the model given in equation (15.1).

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad ....(15.1)$$

where $u_t = \rho u_{t-1} + v_t$ with v_t satisfying all the assumption requirements for OLSM.

Taking one period lagged form of equation (15.1) and then multiplying by ρ we have,

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho u_t \qquad ....(15.22)$$

By subtracting (15.22) from (15.1), we get

$$Y_t - \rho Y_{t-1} = (\beta_0 - \rho\beta_0) + (\beta_1 X_t - \rho\beta_1 X_{t-1}) + (u_t - \rho u_{t-1}) \qquad \cdots(15.23)$$

Assuming $\rho = 1$,

$$Y_t - Y_{t-1} = (\beta_0 - \beta_0) + (\beta_1 X_t - \beta_1 X_{t-1}) + (u_t - u_{t-1})$$
$$\text{or, } \Delta Y_t = \beta_1 \Delta X_t + u_t^* \qquad \cdots(15.24)$$

This method rests on the assumption that either there is perfect positive or negative autocorrelation in the model. But the true ρ may actually lies somewhere between -1 and +1. Suppose that $\rho = -1$ instead of $\rho = 1$, in that case the equation (15.23) can be written as:

$$Y_t + Y_{t-1} = 2\beta_0 + \beta_1 (X_t + X_{t-1}) + (u_t + u_{t-1})$$
$$\text{or, } \frac{Y_t + Y_{t-1}}{2} = \beta_0 + \frac{\beta_1 (X_t + X_{t-1})}{2} + \frac{(u_t + u_{t-1})}{2}$$
$$\text{or, } \frac{Y_t + Y_{t-1}}{2} = \beta_0 + \frac{\beta_1 (X_t + X_{t-1})}{2} + u_t^*$$

The above model is known as two-period moving average regression, since we are regressing the value of one moving average $\left(\frac{Y_t - Y_{t-1}}{2}\right)$ over another moving average $\left(\frac{X_t - X_{t-1}}{2}\right)$.

Thus, we find that when there is perfect positive autocorrelation, taking the first differences of the variables makes the computational procedure simplified. Only thing is that we loose the very first observation on the data-set which, however, can be replaced by the Prais-Winston method.

However, when ρ is not known to us, we can take the two-step feasible GLS procedure. Firstly, we estimate the relationship among the variables and obtain the random error terms. Assuming that the error term follows a first-order autoregressive scheme, we can write:

$u_t = \rho u_{t-1} + v_t$ *with* v_t *satisfying all the classical assumptions of OLS*

(I) Cochrane-Orcutt Iterative Procedue:

We consider the model (15.1):

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad ....(15.1)$$

and the AR(1) scheme $u_t = \rho u_{t-1} + v_t \quad -1 < \rho < 1 \qquad \cdots(15.25)$

Th steps in Cochrane-Orcutt methods are as follows:

Step 1: Apply OLS to (15.1) and obtain the estimates of coefficients and residuals as $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{u}_t$

Step 2: Using the residuals obtained in step 1, run the following regression to obtain the first round estimate of ρ.

$$\hat{u}_t = \hat{\rho} u_{t-1} + v_t^*$$

Step 3: Using $\hat{\rho}$ obtained in step 2, transform the original variables and apply OLS to the following generalized difference equation

$$(Y_t - \hat{\rho} Y_{t-1}) = \beta_0(1-\hat{\rho}) + \beta_1(X_t - \hat{\rho} X_{t-1}) + (u_t - \hat{\rho} u_{t-1})$$
$$\text{or, } Y_t^* = \beta_0^* + \beta_1 X_t^* + u_t^* \qquad \cdots(15.26)$$

Again apply OLS to (15.26) and obtain the estimates of β_0^* and β_1.

Let us denote these second round estimates by $\hat{\hat{\beta}}_o^*$ and $\hat{\hat{\beta}}_1$.

Step 4: Substitute $\hat{\hat{\beta}}_o^*$ and $\hat{\hat{\beta}}_1$ obtained in step 3, in the original regression (15.1) and compute the new residuals, say $\hat{\hat{u}}_t$ as follows:

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

Step 5: Now using $\hat{\hat{u}}_t$ estimate the following regression to obtain the second round estimate of ρ.

$$\hat{\hat{u}}_t = \hat{\hat{\rho}} u_{t-1} + w_t^*$$

Step 6: Using $\hat{\hat{\rho}}$ to transform the original variables and apply OLSM to the generalized difference equation (15.27) as follows:

$$(Y_t - \hat{\hat{\rho}} Y_{t-1}) = \hat{\hat{\beta}}_0(1-\hat{\hat{\rho}}) + \hat{\hat{\beta}}_1(X_t - \hat{\hat{\rho}} X_{t-1}) + (u_t - \hat{\hat{\rho}} u_{t-1}) \qquad \cdots(15.27)$$

Since we do not know whether this second round estimate of ρ is the best estimate of true ρ, we repeat this iterative procedure until the value of the estimate of ρ converges. Generally it is recommended that the iterations may be stopped when the successive estimates of ρ differ s small amount say by less than 0.01 or 0.005. Often researchers stop the iteration at the second round estimates of ρ , then it is called Cochrane-Orcutt two step procedure. The limitation of this procedure is the dropping of the first observation in the data-set. Discarding the first observation may cause serious loss of efficiency particularly in small samples.

(II) Prais-Winsten Procedure:

Prais-Winsten method is very similar to the Cochrane-Orcutt method except that the first observation which was missing in the Cochrane-Orcutt method is now replaced by by some approximation. The first observation is replaced by the following formula:

$Y_1^* = Y_1\sqrt{1-\hat{\rho}^2}$ for the series of Y

$X_1^* = X_1\sqrt{1-\hat{\rho}^2}$ for the series of X

Methods of Estimating ρ:

There are several ways of estimating ρ. All yield biased estimates of the coefficients, but consistent estimates of both the coefficients and their standard errors. We will consider some of these methods in this section.

(I) Estimation from the Residuals:

A measure of the first-order autocorrelation coefficient can be obtained from the residuals of the OLSM by using the following formula:

$$r_{u_t u t-1} = \frac{\sum u_t u_{t-1}}{\sqrt{\sum u_t^2}\sqrt{\sum u_{t-1}^2}} = \hat{\rho}_{u_t u_{t-1}}$$

(II) Estimation from the Residuals:

If the assumption of AR(1) scheme is valid for the residuals, a simple way to estimate ρ is to regress the estimated residuals $\hat{u}_t$ on the lagged values of the estimate residual $\hat{u}_{t-1}$, that is, run the following regression model

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t \qquad \textit{where} \quad v_t = \text{random error term}$$

Please note that there is no constant term in the above regression model.

(III) Estimation based on Durnin-Watson *d* Statistic:

Earlier in section 15.3, we have seen the following relationship:

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

Hence a possible estimate of ρ can be given by:

$$\hat{\rho} = 1 - \frac{d}{2}$$

(IV) Theil and Nagar Method:

The estimation of ρ based on Durbin-Watson *d* statistic will be accurate only when the sample size is large as the above relationship is true only for large sample. For data-set consisting of small number of observations, Theil and Nagar (1961) have suggested the following formula:

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

where *n = Total number of observations, d = Durbin-Watson d-statistic and*

k = Number of parameters including the intercept term in the regression.

When all the explanatory variables will change their values smoothly, the above formula will provide an improvement in the estimation of ρ.

(V) Durbin's Two Step Method:

Durbin has suggested a two steps method for correcting autocorrelation and estimating ρ. For the two variable model, Durbin's two step method can be outlined briefly as follows:

Let us start with the two variable model (15.1)

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad ....(15.1)$$

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

Assuming the first-order autoregressive schemes, we consider the equation

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

Rearranging the above equation, we obtained

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

Or, $Y_t = \beta_0(1-\rho) + \rho Y_{t-1} + \beta_1 X_t - \beta_1 \rho X_{t-1} + u_t^*$ *where* $u_t^* = (u_t - \rho u_{t-1})$

Treating the above equation as a multiple linear regression model, we obtain an estimate of ρ, $\hat{\rho}$ which is the coefficient of Y_{t-1}. This estimate of $\hat{\rho}$ is used to define new variables $(Y_t - \hat{\rho} Y_{t-1})$ and $(X_t - \hat{\rho} X_{t-1})$. In the second step, estimators of β_0 and β_1 are obtained from the following regression equation.

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

where $\beta_0^* = \beta_0(1-\rho)$ and $u_t^* = (u_t - \rho u_{t-1})$

The extension of the Durbin's two-step method is quite simple and straight forward. This can be extended to higher order autoregressive schemes and more than one explanatory variable.

15.5 Dealing with Autocorrelation in R

To illustrate, we will consider the Blaisdell Company example from page 489 of *Applied Linear Regression Models* (4th ed) by Kutner, Nachtsheim, and Neter. In this example we will try to explain the company sales, 'comsales' ($ millions) with the help of single predictor variable, viz. 'indsales' by fitting a simple linear regression model. The data is given in the text format (blaisdell.txt). Mathematically we write the model as follows:

$comsales_t = \beta_0 + \beta_1(indsales_t) + u_t$ where subscript t denotes the time period

Let us first import the data into R by making use of the following commands.

```
> blaisdell<-read.table(file.choose(),header = TRUE)
> View(blaisdell)
> attach(blaisdell)
```

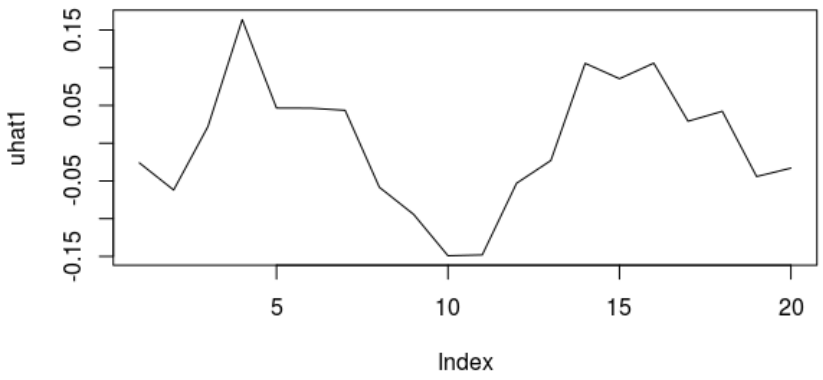
We now fit the model in R and view the results using the commands below:

```
> fit1<-lm(comsales~indsales)
> summary(fit1)
Call:
lm(formula = comsales ~ indsales)
Residuals:
    Min      1Q  Median      3Q     Max
-0.149142 -0.054399 -0.000454  0.046425  0.163754
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.454750   0.214146  -6.793 2.31e-06 ***
indsales     0.176283   0.001445 122.017  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.08606 on 18 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9987
F-statistic: 1.489e+04 on 1 and 18 DF,  p-value: < 2.2e-16
```

The above results show that the company sales is well predicted by the industry sales as the positive regression coefficient of 0.176 is highly significant with a t-value of 122.017. The coefficient of multiple determination (R-squared) is also very high 0.99 and the F-statistic is also statistically significant at 5% level of significance. So, apparently the model is well accepted.
Now we will diagnose the model with respect to autocorrelation problem only.
Tests of autocorrelation: Although there are several tests of autocorrelation, we will discuss only a few here, namely, the graphical method and the Durbin–Watson test.
Graphical Method.
In evaluating regression results it is always good practice to plot the residuals from the estimated model for clues regarding possible violation of one or more OLS assumptions.
Since autocorrelation refers to correlation among the error terms (u_t), a rough and ready method of testing for autocorrelation is to simply plot the values of (u_t), chronologically. Unfortunately, we do not observe (u_t), directly. What we observe are their proxies, $\hat{u}_t$, which we can observe after we estimate the regression model.
By plotting the data on $\hat{u}_t$, chronologically we can get a visual impression of the possibility of autocorrelation. Doing so, we obtain figure 15.1. The codes for obtaining the fig.-15.1 are given below:

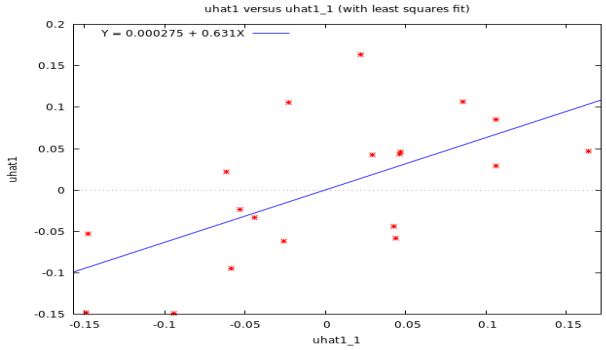
```
> uhat1<-resid(fit1)
> plot(uhat1, type = "l")
```

Fig.-15.1: Residuals Plot of Company Sales model



The curve shows a see-saw pattern, suggesting that the residuals are correlated. This can be seen more clearly if we plot residuals at time *t* against residuals at time *(t – 1)*, as in Figure 15.2.

Fig.-15.2: Scatter plot of Residuals against Lag of Residuals



Durbin-Watson *d* test.

Recall that if we have a first-order autocorrelation with the errors, then the errors are modeled as:

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

where $|\rho| < 1$ and $v_t \sim iid\, N(0, \sigma^2)$

If we suspect first-order autocorrelation with the errors, then a formal test does exist regarding the parameter ρ. In particular, the **Durbin-Watson test** is constructed as:

$$H_0: \rho = 0$$
against
$$H_1: \rho \neq 0$$

So the null hypothesis of $\rho = 0$ means that $u_t = v_t$ or that the error term in one period is not correlated with the error term in the previous period, while the alternative hypothesis of $\rho \neq 0$ means the error term in one period is either positively or negatively correlated with the error term in the previous period. Often times, a researcher will already have an indication of whether the errors are positively or negatively correlated. For example, a regression of oil prices (in dollars per barrel) versus the gas price index will surely have positively correlated errors. When the researcher has an indication of the direction of the correlation, then the Durbin-Watson test also accommodates the one-sided alternatives $H_1: \rho < 0$ for negative correlations or $H_1: \rho > 0$ for positive correlations (as in the oil example).

The test statistic for the Durbin-Watson test on a data set of size *n* is given by:

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

where $\hat{u}_t = y_t - \hat{y}_t$ are the residuals from the ordinary least squares fit. Exact critical values are difficult to obtain, but tables (for certain significance values) can be used to make a decision. The tables provide a lower and upper bound, called d_L and d_U, respectively. In testing for positive autocorrelation, if $d < d_L$ then reject H_0, if $d > d_U$ then fail to reject H_0, or if $d_L \leq d \leq d_U$, then the test is inconclusive. While the prospect of having an inconclusive test result is less than desirable, there are some programs which use exact and approximate procedures for calculating a p-value. These procedures require certain assumptions on the data which we will not discuss. One "exact" method is based on the beta distribution for obtaining p-values.

To obtian the Durbin-Watson test in R, it is required to use the package ‘car’. The test is conducted using the following codes:

```
> library(car)
> durbinWatsonTest(fit1)
 lag Autocorrelation D-W Statistic p-value
   1       0.6260046     0.7347256       0
 Alternative hypothesis: rho != 0
```

Since the value of the Durbin-Watson Statistic falls below the lower bound at a 0.01 significance level (obtained from a table of Durbin-Watson test bounds), there is strong evidence the error terms are positively correlated.

Remedial Measures

When autocorrelated error terms are found to be present, then one of the first remedial measures should be to investigate the omission of a key predictor variable. If such a predictor does not aid in reducing/eliminating autocorrelation of the error terms, then certain transformations on the variables can be performed. We discuss two transformations which are designed for AR(1) errors. Methods for dealing with errors from an AR(*k*) process do exist in the literature, but are much more technical in nature.

Cochrane-Orcutt Procedure

The first of the three transformation methods we discuss is called the Cochrane-Orcutt procedure, which involves an iterative process (after identifying the need for an AR(1) process):

The method has been discussed in details in the theoretical section. Here we will just see the estimation procedure in R. The package called 'orcutt' will implement the Cochrane-Orcutt estimation in R. So, it is required to load the package 'orcutt'.

```
> library(orcutt)

Loading required package: lmtest
Loading required package: zoo
Attaching package: 'zoo'
The following objects are masked from 'package:base':
  as.Date, as.Date.numeric
```

Now the Cochrane-Orcutt estimation can be performed using the following commands:

It is evident from the results that the coefficient of the explanatory variable 'indsales' has changed slightly from 0.176 in the original OLS model to 0.161 in the AR(1) model estimation and still it is highly significant. Other model fit criteria such as adjusted R-squared, F-statistic etc. all are still very high and statistically significant. The most significant improvement is in the Durbin-Watson's *d* statistic. In the original OLS model there was strong evidence of positive first-order autocorrelation. In modified model, the d-statistic has improved significantly to 1.72 using Cochrane-Orcutt estimation. Since the value of the Durbin-Watson Statistic falls above the upper bound at a 0.01 significance level (obtained from a table of Durbin-Watson test bounds), there is no evidence the error terms are positively correlated in the model with the transformed variables.

```
> fit2<-cochrane.orcutt(fit1)
> summary(fit2)
Call:
lm(formula = comsales ~ indsales)
              Estimate Std. Error t value  Pr(>|t|)
(Intercept) 1.7389010  0.2141461  -6.793 2.314e-06 ***
indsales    0.1605234  0.0014447 122.017 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0649 on 17 degrees of freedom
Multiple R-squared:  0.9702 ,  Adjusted R-squared:  0.9684
F-statistic: 553.1 on 1 and 17 DF,  p-value: < 2.085e-14

Durbin-Watson statistic
(original):    0.73473 , p-value: 1.748e-04
(transformed): 1.72442 , p-value: 2.801e-01
```

First Differences Procedure

Suppose autocorrelation is of AR(1) type, then we can write

$$u_t - \rho u_{t-1} = v_t$$

If we know the value of ρ, we can subtract from the current value of the error term ρ times the previous value of the error term. The resulting error term, v_t will satisfy the standard OLS assumptions. Therefore, we can transform the original regression as:

$$comsales_t - \rho * comsales_{t-1} = \beta_0(1-\rho) + \beta_2(indsales_t - \rho * indsales_{t-1}) + (u_t - \rho u_{t-1})$$

The last term in this equation is simply v_t, which now is free from autocorrelation. The transformed model can therefore be estimated by OLS. All we have to do is transform each variable by subtracting from its current value ρ times its previous value and run the regression. The estimators obtained from the transformed model are BLUE.

But note that in this transformation we lose one observation, because for the very first observation there is no antecedent. If the sample is reasonably large, loss of one observation may not matter much. But if the sample size is small, the loss of the first observation means the estimators will not be BLUE. However, there is a procedure, called the Prais–Winsten transformation, that can take into account the first observation. Now the question is: how do we estimate ρ ? We know that $-1<\rho<1$. Therefore, any value in this range can be used to transform the original model. But which one value should we choose, for literally there is an infinite number of values in this range?

Many economic time series are highly inter-correlated, suggesting that perhaps a value $\rho=1$ may be appropriate to transform the original model. If this is indeed the case, the transformed model can be written as:

$$\Delta comsales_t = \beta_1 \Delta indsales_t + v_t$$

where Δ is the first-difference operator. The last equation is called the first-difference transformation, while the original model is called the level form regression.

Please note that in the first-difference transformation there is no intercept in it. Therefore, in estimating this model, we need to suppress the intercept term. R can easily do this.

```
> fit3<-lm(diff(comsales,1)~0+diff(indsales,1))
> summary(fit3)
Call:
lm(formula = diff(comsales, 1) ~ 0 + diff(indsales, 1))

Residuals:
     Min       1Q   Median       3Q      Max
-0.08958 -0.03231  0.02412  0.05344  0.15139

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
diff(indsales, 1) 0.168488   0.005096   33.06   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.06939 on 18 degrees of freedom
Multiple R-squared:  0.9838,    Adjusted R-squared:  0.9829
F-statistic:  1093 on 1 and 18 DF,  p-value: < 2.2e-16
```

If we test this regression for autocorrelation using the Durbin-Watson or any other test, we find that there is no evidence of autocorrelation.

```
> durbinWatsonTest(fit3)
 lag Autocorrelation D-W Statistic p-value
   1       0.1223879      1.738895   0.662
 Alternative hypothesis: rho != 0
```

If we compare the two regression models (level form and transformed form), we find that the regression coefficient of the explanatory variable 'indsales' are more or less the same.

It should be noted that the R2 values in the level form and in the first-difference form are not directly comparable because the dependent variable in the two models is different. As noted before, to compare two or more R2 values, the dependent variable must be the same.

15.6 Advanced Autocorrelation Problem in R

In this section we will reproduce the case of US Consumption function, 1947-2000 discussed by Gujarati in his book *Econometrics by Example.* The data-set consists of data on real consumption expenditure (C), real disposable income (Y), real wealth (W) and real interest rate (r) for the USA economy during the period 1947-2000. The model to start with is given below:

$$lnC_t = \beta_0 + \beta_1 lnY_t + \beta_2 lnW_t + \beta_3 r_t + u_t \qquad \cdots(i)$$

where the subscript *t* denotes the time and *ln* stands for natural logarithm of the variables.

All the variables except the real interest are introduced in logarithm form. This is so because some of the real interest rates are negative.

As far as the partial regression coefficients are concerned it can be said that β_1 and β_2 are the elasticities of real consumption expenditure with respect to real disposable income and real wealth respectively and β_3 is the semi-elasticity with respect to real interest rate. β_1 and β_2 are expected to hold positive signs and β_3 is expected to have negative sign.

The data-set is given in excel form (us_consumption.xls).

Firstly, we import the data-set in R using the commands given below.

```
> library(readxl)
> us_consumption<-read_excel(file.choose(), sheet = "Sheet1", col_names = TRUE,
      col_types = NULL, skip = 0)
> View(us_consumption)
> attach(us_consumption)
```

Now we run the model in R and the results are shown below:

```
> fit1<- lm(lnconsump~lndpi+lnwealth+interest)
> summary(fit1)
Call:
lm(formula = lnconsump ~ lndpi + lnwealth + interest)

Residuals:
    Min        1Q    Median        3Q       Max
-0.018441 -0.010001  0.000337  0.007038  0.032579

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.4677120  0.0427780 -10.933 7.33e-15 ***
lndpi        0.8048728  0.0174978  45.998  < 2e-16 ***
lnwealth     0.2012702  0.0175926  11.441 1.43e-15 ***
interest    -0.0026891  0.0007619  -3.529 0.000905 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.01193 on 50 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9995
F-statistic: 3.783e+04 on 3 and 50 DF,  p-value: < 2.2e-16
```

Interpretation of the result:

All the slope coefficients have their expected signs and they are highly statistically significant as the estimated p-values are much lower than the level of significance (5% or 0.05). The income elasticity of 0.8 suggests that, holding other variables constant, if real personal disposal income goes up by 1%, mean real consumption expenditure goes up by about 0.8%. similarly we can interpret the regression coefficient of real wealth. The interest semi-elasticity suggests that if interest rate goes up by one percentage point (not 1%), mean real consumption expenditure goes down by about 0.26%, ceteris paribus.

The R^2 and other relevant statistic (F-value) suggest that the model is very well fit. In fact, the value of adjusted R-squared is nearly unity which gives rise to the suspect of possible spurious regression. Also as the data-set is a time series data, if there is problem of autocorrelation, the estimated t-values of the regression coefficients and the statistical significance of the coefficients will be in question mark.

Therefore, we should go for the diagnostic checking of autocorrelation of the model, before they are accepted for further analysis for prediction.

Tests for Autocorrelation.

We will restrict ourselves into graphical tests, Durbin-Watson’ test and Breusch-Godfrey test for the detection of autocorrelation in the model.

Graphical Method.

In graphical method we will plot the residuals of the model fit against time and check the pattern therein. If the residuals follow any particular trend, then we may doubt about the validity of the fulfillment of the autocorrelation assumption. To save the residual of the model and declare it as time series object, use the following R command.

```
uhat1<-ts(resid(fit1))
```

So, the residuals are saved as time series object called ‘uhat1’. We plot the residuals of the model using the command below and the fig. -15.3 is produced.

```
> plot(uhat1, type="l")
```

Fig. - 15.3: Time Series Plot of Residual

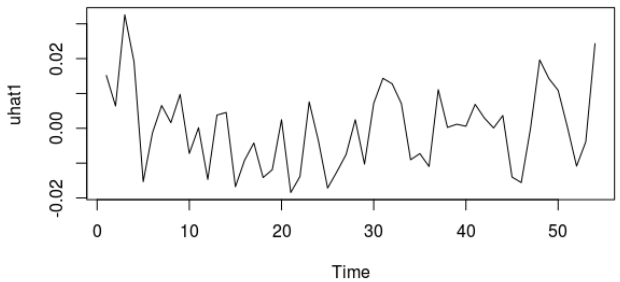
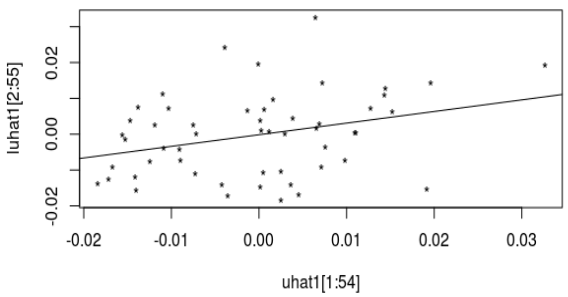


Fig.-15.4: Current vs. lagged Residuals



The residuals curve suggests that the residuals are correlated. This can be seen more clearly if we plot residuals at time t against residuals at time (t – 1), as shown in Figure 15.4. The codes to be used to generate the fig.-15.4 are shown below.

```
> luhat1<-lag(uhat1, -1)
> plot(uhat1[1:54], luhat1[2:55], pch="*")
> abline(lm(luhat1[2:55]~uhat1[1:54]))
```

The stretched regression line in fig.-15.4 shows that the residuals are positively correlated.

Durbin-Watson *d* test.

We have already discussed in details about this test repeatedly. So, without going into the much elaborated theorem, we will concentrate on the empirical results only. The Durbin-Watson's *d* statistic is calculated for the current example and is found to be 1.28.

```
> library(car)
> durbinWatsonTest(fit1)
 lag Autocorrelation D-W Statistic p-value
   1       0.2977555      1.289232   0.004
 Alternative hypothesis: rho != 0
```

What do we do with this value? As noted earlier that the *d* value lies between 0 and 4. The closer it is to zero, the greater is the evidence of positive autocorrelation, and the closer it is to 4, the greater is the evidence of negative autocorrelation. If *d* is about 2, there is no evidence of positive or negative first-order autocorrelation.

In our current problem, we have $n = 54$, X (number of regressors) = 3. The 5% critical *d* values for this combination are (using n = 55): (1.452, 1.681). Since the computed *d* value is about 1.28, it lies below the lower limit, leading to the conclusion that we probably have positive autocorrelation in the error term. The 1% critical *d* values are: (1.284, 1.506). The computed *d* value is slightly below the lower limit, again suggesting that our regression probably suffers from positive first-order autocorrelation.

Breusch–Godfrey (BG) test of autocorrelation.

To overcome some of the limitations of the d test, Breusch and Godfrey have developed a test of autocorrelation that is more general in that it allows for (i) lagged values of the dependent variables to be included as regressors, (ii) higher-order autoregressive schemes, such as AR (2) and AR (3), and (iii) moving average terms of the error term, such as u_{t-1}, u_{t-1} and so on.

In our consumption function problem, suppose that the error term of the model (i) follows the following pattern:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots\ldots + \rho_p u_{t-p} + v_t \qquad \cdots(ii)$$

The above equation is an AR(p) autoregressive scheme where the error term in the current period depends on the lagged error terms upto p-th period.

The hypotheses to be tested are given as under:

$H_0: \rho_1 = \rho_2 = ... = \rho_p = 0$

against

H_1: Not all the ρ_s are zero

So, the null hypothesis is set as that there is autocorrelation of any order.

The steps in the B – G test have already discussed in section 15.3 earlier. For large sample size, BG have shown that

$(n-p)R^2 \sim \chi_p^2$ where R^2 is the R-squared of the following auxiliary regression

$$\hat{u}_t = \alpha_0 + \alpha_1 lnY_t + \alpha_2 lnW_t + \alpha_3 r_t + \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + v_t \qquad .....(iii)$$

That is, in large sample, *(n – p)* times R^2 follows the chi-square distribution with *p* degrees of freedom.

If in an application, $(n-p)R^2$ exceeds the critical value of chi-square at the chosen level of significance, it is wise to reject the null hypothesis, in which case at least one ρ in equation (iii) is statistically significantly different from zero.

A practical problem in the application of the BG test is the choice of the number of lagged error terms, *p*, in equation (iii). The value of *p* may depend on the type of time series. For monthly data, we may include 11 lagged error terms, for quarterly data we may include three lagged error terms, and for annual data, one lagged error term may suffice. Of course, we can choose the lag length by trial and error and choose the value of *p* based on the Akaike Information Criteria (AIC) and Schwarz information criteria (SIC). The lower the value of these criteria, the better is the model.

Returning to our consumption function problem, we perform the B-G test by choosing one period lag. The codes and results in R are shown below:

```
> library(lmtest)
> bgtest(fit1, order = 1, type = "Chisq")
    Breusch-Godfrey test for serial correlation of order up to 1
data:  fit1
LM test = 5.312, df = 1, p-value = 0.02118
```

The above results show that there is strong evidence of first-order autocorrelation, for chi-squared value is highly significant because their p values are so low.

Remedial Measures for Autocorrelation.

To deal with the autocorrelation problem, it will be easier if the nature of the correlation among the successive error terms is known to us. But on very rare occasion we know the correlation structure of the error terms u_t , since they are not directly observable. Hence, we need to resort to some educated guesswork or some kind of transformation of the original regression model so that in the transformed model we do not face the serial correlation problem. There are several methods that we could try. Some of these methods will be discussed here.

First Differences Method.

As already discussed in the last empirical problem that taking the first-difference of the variables can often solve the problem of autocorrelation. Taking the first-difference of the original model as given in equation (I) can be written as:

$$\Delta lnC_t = \beta_1 \Delta lnY_t + \beta_2 \Delta lnW_t + \beta_3 \Delta r_t + v_t \qquad \cdots (iv)$$

where Δ is the first-difference operator, that is, $\Delta lnC_t = (lnC_t - lnC_{t-1})$ and similarly we can write for other variables as well.

Now to estimate the model given by equation (iv), use the following codes in R.

```
> fit2<- lm(diff(lnconsump,1)~0+diff(lndpi,1)+diff(lnwealth,1)+diff(interest,1))
> summary(fit2)
Call:
lm(formula = diff(lnconsump, 1) ~ 0 + diff(lndpi, 1) + diff(lnwealth,
  1) + diff(interest, 1))

Residuals:
    Min        1Q    Median        3Q       Max
-0.027810 -0.007000  0.002783  0.009686  0.023537

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
diff(lndpi, 1)    0.8489940  0.0515376  16.473  < 2e-16 ***
diff(lnwealth, 1) 0.1063550  0.0368543   2.886  0.00575 **
diff(interest, 1) 0.0006530  0.0008261   0.790  0.43297
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.01113 on 50 degrees of freedom
Multiple R-squared:  0.9237,    Adjusted R-squared:  0.9191
F-statistic: 201.6 on 3 and 50 DF,  p-value: < 2.2e-16
```

Now we should test this model for serial correlation by B-G test. The null hypothesis for B-G test is given as:

H_0: There is no autocorrelation

against

H_1: There is autocorrelation

The R code and results for the B-G test are shown below:

```
> bgtest(fit2,order = 1, type = "Chisq")

   Breusch-Godfrey test for serial correlation of order up to 1
data:  fit2
LM test = 0.86275, df = 1, p-value = 0.353
```

The result shows that the transformed model doesn’t suffer from the autocorrelation problem.

If we compare the regression results of the original regression and those obtained from first difference transformation, we see that the income elasticity is more or less the same, but the wealth elasticity, although statistically significant, is almost half in value and the interest rate semi-elasticity is practically zero and has the wrong sign. This outcome could be due to the wrong value of ρ chosen for transformation. But more fundamentally it may have to do with the stationarity of one or more variables, a topic that is beyond the scope of this book. It should be emphasized that

the R^2 values in the level form and in the first-difference form are not directly comparable because the dependent variable in the two models is different.

Generalized Transformation.

For the large sample size and for the AR(1) scheme, we can regress $\hat{u}_t$ on $\hat{u}_{t-1}$ using $\hat{u}_t$ as a proxy for the original residual, because in large samples $\hat{u}_t$ is a consistent estimator of ρ. So, we estimate the following equation:

$$\hat{u}_t = \hat{\rho}\,\hat{u}_{t-1} + random\ error\ term$$

where $\hat{\rho}$ is an estimator of ρ

After getting the estimator of ρ, we can transform the original model as given in equation (i) and estimate the transformed model.

Use the following R codes, we can estimate the value of ρ.

To save the residuals of the original model (i).

```
> uhat1<-resid(fit1)
```

To generate the one-period lag of the residuals.

```
> library(dplyr)
> luhat1<-lag(uhat1, n=1L)
```

For estimating ρ we regress $\hat{u}_t$ on $\hat{u}_{t-1}$ without intercept term by using the following codes.

```
> fit3<-lm(uhat1~0+luhat1)
> summary(fit3)
Call:
lm(formula = uhat1 ~ 0 + luhat1)
Residuals:
     Min       1Q   Median       3Q      Max
-0.0215503 -0.0078554  0.0002091  0.0066878  0.0305027
Coefficients:
     Estimate Std. Error t value Pr(>|t|)
luhat1   0.3247     0.1351   2.402   0.0199 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01092 on 52 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.09991,   Adjusted R-squared:  0.0826
F-statistic: 5.772 on 1 and 52 DF,  p-value: 0.01989
```

Thus, we get $\hat{\rho} = 0.3247$.

Alternatively we can estimate ρ from the Durbin-Watson d statistic of the original model by using the following relationship: $\rho \approx 1 - \frac{d}{2}$

This is what is known as Generalized Least Squares Estimate.

Coming back to R, use the following codes to get the Durbin-Watson d statistic of the original model.

```
> durbinWatsonTest(fit1)
```

```
 lag Autocorrelation D-W Statistic p-value
   1       0.2977555      1.289232   0.004
 Alternative hypothesis: rho != 0
```

Substituting the value of *d = 1.289* in the above relationship, we get

$$\hat{\rho}=1-\frac{1.289}{2}=0.355$$

Now we can transform the original model using this estimated value of ρ as follows:

$$lnC_t-\rho\, lnC_{t-1}=\beta_0(1-\rho)+\beta_1(lnY_t-\rho\, lnY_{t-1})+\beta_2(lnW_t-\rho\, lnW_{t-1})+\beta_3(lnr_t-\rho\, lnr_{t-1})+(u_t-\rho\, u_{t-1})$$

$$\text{Or, } C_t^*=\beta_0^*+\beta_1^* Y_t^*+\beta_2^* W_t^*+\beta_3^* r_t^*+v_t^*$$

$$\textit{where}\quad C_t^*=lnC_t-\rho\, C_{t-1}, \beta_0^*=\beta_0(1-\rho), \textit{etc}.$$

The complete R codes and the results of the transformed regression are shown below:

```
> Cstar<- lnconsump-0.3247*lag(lnconsump,n=1L)
> Ystar<-lndpi-0.3247*lag(lndpi,n=1L)
> Wstar<-lnwealth-0.3247*lag(lnwealth,n=1L)
> rstar<-interest-0.3247*lag(interest,n=1L)
> fit4<-lm(Cstar~Ystar+Wstar+rstar)
> summary(fit4)
Call:
lm(formula = Cstar ~ Ystar + Wstar + rstar)
Residuals:
    Min        1Q    Median        3Q       Max
-0.017901 -0.006608 -0.001012  0.006385  0.025941
Coefficients:
        Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.797e-01  3.373e-02  -8.294 6.81e-11 ***
Ystar        8.187e-01  2.110e-02  38.806  < 2e-16 ***
Wstar        1.836e-01  2.099e-02   8.749 1.40e-11 ***
rstar       -1.782e-05  9.691e-04  -0.018    0.985
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.01042 on 49 degrees of freedom
 (1 observation deleted due to missingness)
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9992
F-statistic: 2.133e+04 on 3 and 49 DF,  p-value: < 2.2e-16
```

Now the residuals from this regression when diagnosed for serial correlation, using, the BG test using 1 and 2 lagged terms, there is not any evidence of serial correlation in the residuals of AR(1) transformed model.

By comparing the results of the original regression model and that of the last transformed model, we see that the standard errors of the coefficients in the two tables are substantially different, but remember that the original model does not correct for autocorrelation, whereas the transformed model does. The magnitudes of the income and wealth elasticities are about the same in the two models, although the standard errors, and therefore the *t* values, are different. The lower absolute *t* values in the transformed model suggest that the original OLS standard errors were underestimated, which follows our discussion of the consequences of OLS estimation in the presence of autocorrelation. The interest rate coefficient in the transformed model has the correct sign, but it is statistically insignificant.