

## Chapter 9

### Autocorrelation

One of the basic assumptions in the linear regression model is that the random error components or disturbances are identically and independently distributed. So in the model  $y = X\beta + u$ , it is assumed that

$$E(u_t, u_{t-s}) = \begin{cases} \sigma_u^2 & \text{if } s = 0 \\ 0 & \text{if } s \neq 0 \end{cases}$$

i.e., the correlation between the successive disturbances is zero.

In this assumption, when  $E(u_t, u_{t-s}) = \sigma_u^2$ ,  $s = 0$  is violated, i.e., the variance of disturbance term does not remain constant, then the problem of heteroskedasticity arises. When  $E(u_t, u_{t-s}) = 0$ ,  $s \neq 0$  is violated, i.e., the variance of disturbance term remains constant though the successive disturbance terms are correlated, then such problem is termed as the problem of autocorrelation.

When autocorrelation is present, some or all off-diagonal elements in  $E(uu')$  are nonzero.

Sometimes the study and explanatory variables have a natural sequence order over time, i.e., the data is collected with respect to time. Such data is termed as **time-series data**. The disturbance terms in time series data are serially correlated.

The **autocovariance** at lag  $s$  is defined as

$$\gamma_s = E(u_t, u_{t-s}); s = 0, \pm 1, \pm 2, \dots$$

At zero lag, we have constant variance, i.e.,

$$\gamma_0 = E(u_t^2) = \sigma^2.$$

The **autocorrelation coefficient** at lag  $s$  is defined as

$$\rho_s = \frac{E(u_t u_{t-s})}{\sqrt{\text{Var}(u_t) \text{Var}(u_{t-s})}} = \frac{\gamma_s}{\gamma_0}; s = 0, \pm 1, \pm 2, \dots$$

Assume  $\rho_s$  and  $\gamma_s$  are symmetrical in  $s$ , i.e., these coefficients are constant over time and depend only on the length of lag  $s$ . The autocorrelation between the successive terms  $(u_2$  and  $u_1)$   $(u_3$  and  $u_2), \dots, (u_n$  and  $u_{n-1})$  gives the autocorrelation of order one, i.e.,  $\rho_1$ . Similarly, the autocorrelation between the successive terms  $(u_3$  and  $u_1), (u_4$  and  $u_2), \dots, (u_n$  and  $u_{n-2})$  gives the autocorrelation of order two, i.e.,  $\rho_2$ .

## Source of autocorrelation

Some of the possible reasons for the introduction of autocorrelation in the data are as follows:

1. Carryover of effect, at least in part, is an important source of autocorrelation. For example, the monthly data on expenditure on household is influenced by the expenditure of preceding month. The autocorrelation is present in cross-section data as well as time-series data. In the cross-section data, the neighbouring units tend to be similar with respect to the characteristic under study. In time-series data, time is the factor that produces autocorrelation. Whenever some ordering of sampling units is present, the autocorrelation may arise.
2. Another source of autocorrelation is the effect of deletion of some variables. In regression modeling, it is not possible to include all the variables in the model. There can be various reasons for this, e.g., some variable may be qualitative, sometimes direct observations may not be available on the variable etc. The joint effect of such deleted variables gives rise to autocorrelation in the data.
3. The misspecification of the form of relationship can also introduce autocorrelation in the data. It is assumed that the form of relationship between study and explanatory variables is linear. If there are log or exponential terms present in the model so that the linearity of the model is questionable, then this also gives rise to autocorrelation in the data.
4. The difference between the observed and true values of the variable is called measurement error or errors-in-variable. The presence of measurement errors on the dependent variable may also introduce the autocorrelation in the data.

## Structure of disturbance term:

Consider the situation where the disturbances are autocorrelated,

$$\begin{aligned}
 E(\varepsilon\varepsilon') &= \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} \\
 &= \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix} \\
 &= \sigma_u^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}.
 \end{aligned}$$

Observe that now there are  $(n+k)$  parameters-  $\beta_1, \beta_2, \dots, \beta_k, \sigma_u^2, \rho_1, \rho_2, \dots, \rho_{n-1}$ . These  $(n+k)$  parameters are to be estimated on the basis of available  $n$  observations. Since the number of parameters are more than the number of observations, so the situation is not good from the statistical point of view. In order to handle the situation, some special form and the structure of the disturbance term is needed to be assumed so that the number of parameters in the covariance matrix of disturbance term can be reduced.

The following structures are popular in autocorrelation:

1. Autoregressive (AR) process.
2. Moving average (MA) process.
3. Joint autoregression moving average (ARMA) process.

### 1. Autoregressive (AR) process

The structure of disturbance term in the autoregressive process (AR) is assumed as

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_q u_{t-q} + \varepsilon_t,$$

i.e., the current disturbance term depends on the  $q$  lagged disturbances and  $\phi_1, \phi_2, \dots, \phi_q$  are the parameters (coefficients) associated with  $u_{t-1}, u_{t-2}, \dots, u_{t-q}$  respectively. An additional disturbance term is introduced in  $u_t$  which is assumed to satisfy the following conditions:

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t \varepsilon_{t-s}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } s = 0 \\ 0 & \text{if } s \neq 0. \end{cases}$$

This process is termed as  $AR(q)$  process. In practice, the  $AR(1)$  process is more popular.

## 2. Moving average (MA) process:

The structure of disturbance term in the moving average (MA) process is

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-p},$$

i.e., the present disturbance term  $u_t$  depends on the  $p$  lagged values. The coefficients  $\theta_1, \theta_2, \dots, \theta_p$  are the parameters and are associated with  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p}$ , respectively. This process is termed as  $MA(p)$  process.

## 3. Joint autoregressive moving average (ARMA) process:

The structure of disturbance term in the joint autoregressive moving average (ARMA) process is

$$u_t = \phi_1 u_{t-1} + \dots + \phi_q u_{t-q} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_p \varepsilon_{t-p}.$$

This is termed as  $ARMA(q, p)$  process.

The method of correlogram is used to check that the data is following which of the processes. The correlogram is a two dimensional graph between the lag  $s$  and autocorrelation coefficient  $\rho_s$  which is plotted as lag  $s$  on  $X$ -axis and  $\rho_s$  on  $y$ -axis.

In  $MA(1)$  process

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\rho_s = \begin{cases} \frac{\theta_1}{1 + \theta_1^2} & \text{for } s = 1 \\ 0 & \text{for } s \geq 2 \end{cases}$$

$$\rho_0 = 1$$

$$\rho_1 \neq 0$$

$$\rho_i = 0 \quad i = 2, 3, \dots$$

So there is no autocorrelation between the disturbances that are more than one period apart.

In  $ARMA(1,1)$  process

$$u_t = \phi_1 u_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\rho_s = \begin{cases} \frac{(1 + \phi_1 \theta_1)(\phi_1 + \theta_1)}{(1 + \theta_1^2 + 2\phi_1 \theta_1)} & \text{for } s = 1 \\ \phi_1 \rho_{s-1} & \text{for } s \geq 2 \end{cases}$$

$$\sigma_u^2 = \left( \frac{1 + \theta_1^2 + 2\phi_1 \theta_1}{1 - \phi_1^2} \right) \sigma_\varepsilon^2.$$

The autocorrelation function begins at some point determined by both the  $AR$  and  $MA$  components but thereafter, declines geometrically at a rate determined by the  $AR$  component.

In general, the autocorrelation function

- is nonzero but is geometrically damped for  $AR$  process.
- becomes zero after a finite number of periods for  $MA$  process.

The  $ARMA$  process combines both these features.

The results of any lower order of process are not applicable in higher-order schemes. As the order of the process increases, the difficulty in handling them mathematically also increases.

### Estimation under the first order autoregressive process:

Consider a simple linear regression model

$$y_t = \beta_0 + \beta_1 X_t + u_t, \quad t = 1, 2, \dots, n.$$

Assume  $u_t$ 's follow a first-order autoregressive scheme defined as

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

where  $|\rho| < 1$ ,  $E(\varepsilon_t) = 0$ ,

$$E(\varepsilon_t, \varepsilon_{t+s}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } s = 0 \\ 0 & \text{if } s \neq 0 \end{cases}$$

for all  $t = 1, 2, \dots, n$  where  $\rho$  is the first-order autocorrelation between  $u_t$  and  $u_{t-1}$ ,  $t = 1, 2, \dots, n$ . Now

$$\begin{aligned} u_t &= \rho u_{t-1} + \varepsilon_t \\ &= \rho(u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \vdots \\ &= \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \dots \\ &= \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r} \end{aligned}$$

$$\begin{aligned}
E(u_t) &= 0 \\
E(u_t^2) &= E(\varepsilon_t^2) + \rho^2 E(\varepsilon_{t-1}^2) + \rho^4 E(\varepsilon_{t-2}^2) + \dots \\
&= (1 + \rho^2 + \rho^4 + \dots) \sigma_\varepsilon^2 \quad (\varepsilon_t\text{'s are serially independent}) \\
E(u_t^2) &= \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \text{ for all } t.
\end{aligned}$$

$$\begin{aligned}
E(u_t u_{t-1}) &= E\left[\left(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots\right) \times \left(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \dots\right)\right] \\
&= E\left[\left\{\varepsilon_t + \rho\left(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots\right)\right\} \left\{\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots\right\}\right] \\
&= \rho E\left[\left(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots\right)^2\right] \\
&= \rho \sigma_u^2.
\end{aligned}$$

Similarly,

$$E(u_t u_{t-2}) = \rho^2 \sigma_u^2.$$

In general,

$$\begin{aligned}
E(u_t u_{t-s}) &= \rho^s \sigma_u^2 \\
E(uu') &= \Omega = \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.
\end{aligned}$$

Note that the disturbance terms are no more independent and  $E(uu') \neq \sigma^2 I$ . The disturbances are nonspherical.

### Consequences of autocorrelated disturbances:

Consider the model with first-order autoregressive disturbances

$$\begin{aligned}
y &= X\beta + u \\
n \times 1 & \quad n \times k \quad k \times 1 \quad n \times 1 \\
u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n
\end{aligned}$$

with assumptions

$$\begin{aligned}
E(u) &= 0, \quad E(uu') = \Omega \\
E(\varepsilon_t) &= 0, \quad E(\varepsilon_t \varepsilon_{t+s}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } s = 0 \\ 0 & \text{if } s \neq 0 \end{cases}
\end{aligned}$$

where  $\Omega$  is a positive definite matrix.

The ordinary least squares estimator of  $\beta$  is

$$\begin{aligned} b &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + u) \\ b - \beta &= (X'X)^{-1}X'u \\ E(b - \beta) &= 0. \end{aligned}$$

So OLSE remains unbiased under autocorrelated disturbances.

The covariance matrix of  $b$  is

$$\begin{aligned} V(b) &= E(b - \beta)(b - \beta)' \\ &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &\neq \sigma_u^2(X'X)^{-1}. \end{aligned}$$

The residual vector is

$$\begin{aligned} e &= y - Xb = \bar{H}y = \bar{H}u \\ e'e &= y'\bar{H}y = u'\bar{H}u \\ E(e'e) &= E(u'u) - E[u'X(X'X)^{-1}X'u] \\ &= n\sigma_u^2 - \text{tr}(X'X)^{-1}X'\Omega X. \end{aligned}$$

Since  $s^2 = \frac{e'e}{n-1}$ , so

$$E(s^2) = \frac{\sigma_u^2}{n-1} - \frac{1}{n-1} \text{tr}(X'X)^{-1}X'\Omega X,$$

so  $s^2$  is a biased estimator of  $\sigma^2$ . In fact,  $s^2$  has a downward bias.

Application of OLS fails in case of autocorrelation in the data and leads to serious consequences as

- overly optimistic view from  $R^2$ .
- narrow confidence interval.
- usual  $t$ -ratio and  $F$ -ratio tests provide misleading results.
- prediction may have large variances.

Since disturbances are nonspherical, so generalized least squares estimate of  $\beta$  yields more efficient estimates than OLSE.

The GLSE of  $\beta$  is

$$\begin{aligned}\hat{\beta} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \\ E(\hat{\beta}) &= \beta \\ V(\hat{\beta}) &= \sigma_u^2 (X' \Omega^{-1} X)^{-1}.\end{aligned}$$

The GLSE is best linear unbiased estimator of  $\beta$ .

## Tests for autocorrelation:

### Durbin Watson test:

The Durbin-Watson ( $D-W$ ) test is used for testing the hypothesis of lack of first-order autocorrelation in the disturbance term. The null hypothesis is

$$H_0 : \rho = 0$$

Use OLS to estimate  $\beta$  in  $y = X\beta + u$  and obtain the residual vector

$$e = y - Xb = \bar{H}y$$

where  $b = (X'X)^{-1}X'y$ ,  $\bar{H} = I - X(X'X)^{-1}X'$ .

The  $D-W$  test statistic is

$$\begin{aligned}d &= \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \\ &= \frac{\sum_{t=2}^n e_t^2}{\sum_{t=1}^n e_t^2} + \frac{\sum_{t=2}^n e_{t-1}^2}{\sum_{t=1}^n e_t^2} - 2 \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}.\end{aligned}$$

For large  $n$ ,

$$\begin{aligned}d &\approx 1 + 1 - 2r \\ d &\approx 2(1 - r)\end{aligned}$$

where  $r$  is the sample autocorrelation coefficient from residuals based on OLSE and can be regarded as the regression coefficient of  $e_t$  on  $e_{t-1}$ . Here



positive autocorrelation of  $e_t$ 's  $\Rightarrow d < 2$

negative autocorrelation of  $e_t$ 's  $\Rightarrow d > 2$

zero autocorrelation of  $e_t$ 's  $\Rightarrow d \approx 2$

As  $-1 < r < 1$ , so

if  $-1 < r < 0$ , then  $2 < d < 4$  and

if  $0 < r < 1$ , then  $0 < d < 2$ .

So  $d$  lies between 0 and 4.

Since  $e$  depends on  $X$ , so for different data sets, different values of  $d$  are obtained. So the sampling distribution of  $d$  depends on  $X$ . Consequently, exact critical values of  $d$  cannot be tabulated owing to their dependence on  $X$ . Durbin and Watson, therefore, obtained two statistics  $\underline{d}$  and  $\bar{d}$  such that

$$\underline{d} < d < \bar{d}$$

and their sampling distributions do not depend upon  $X$ .

Considering the distribution of  $\underline{d}$  and  $\bar{d}$ , they tabulated the critical values as  $d_L$  and  $d_U$  respectively. They prepared the tables of critical values for  $15 < n < 100$  and  $k \leq 5$ . Now tables are available for  $6 < n < 200$  and  $k \leq 10$ .

The test procedure is as follows:

$H_0 : \rho = 0$			
Nature of $H_1$	Reject $H_0$ when	Retain $H_0$ when	The test is inconclusive when
$H_1 : \rho > 0$	$d < d_L$	$d > d_U$	$d_L < d < d_U$
$H_1 : \rho < 0$	$d > (4 - d_L)$	$d < (4 - d_U)$	$(4 - d_U) < d < (4 - d_L)$
$H_1 : \rho \neq 0$	$d < d_L$ or $d > (4 - d_L)$	$d_U < d < (4 - d_U)$	$d_L < d < d_U$ or $(4 - d_U) < d < (4 - d_L)$
Values of $d_L$ and $d_U$ are obtained from tables.			

## Limitations of $D$ - $W$ test

1. If  $d$  falls in the inconclusive zone, then no conclusive inference can be drawn. This zone becomes fairly larger for low degrees of freedom. One solution is to reject  $H_0$  if the test is inconclusive. A better solution is to modify the test as

- Reject  $H_0$  when  $d < d_U$ .
- Accept  $H_0$  when  $d \geq d_U$ .

This test gives a satisfactory solution when values of  $x_t$ 's change slowly, e.g., price, expenditure etc.

2. The  $D$ - $W$  test is not applicable when the intercept term is absent in the model. In such a case, one can use another critical value, say  $d_M$  in place of  $d_L$ . The tables for critical values  $d_M$  are available.

3. The test is not valid when lagged dependent variables appear as explanatory variables. For example,

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_r y_{t-r} + \beta_{r+1} x_{t1} + \dots + \beta_k x_{t,k-r} + u_t,$$

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

In such case, **Durbin's  $h$  test** is used, which is given as follows.

## Durbin's $h$ -test

Apply OLS to

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_r y_{t-r} + \beta_{r+1} x_{t1} + \dots + \beta_k x_{t,k-r} + u_t,$$

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

and find OLSE  $b_1$  of  $\beta_1$ . Let its variance be  $Var(b_1)$  and its estimator is  $\widehat{Var}(b_1)$ . Then the Durbin's  $h$ -statistic is

$$h = r \sqrt{\frac{n}{1-n \widehat{Var}(b_1)}}$$

which is asymptotically distributed as  $N(0,1)$  and

$$r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}.$$

This test is applicable when  $n$  is large. When  $\left[1 - n\widehat{\text{Var}}(b_1)\right] < 0$ , then test breaks down. In such cases, the following test procedure can be adopted.

Introduce a new variable  $\varepsilon_{t-1}$  to  $u_t = \rho u_{t-1} + \varepsilon_t$ . Then

$$e_t = \delta \rho_{t-1} + y_t.$$

Now apply OLS to this model and test  $H_{0A} : \delta = 0$  versus  $H_{1A} : \delta \neq 0$  using  $t$ -test. If  $H_{0A}$  is accepted then accept  $H_0 : \rho = 0$ .

If  $H_{0A} : \delta = 0$  is rejected, then reject  $H_0 : \rho = 0$ .

4. If  $H_0 : \rho = 0$  is rejected by  $D$ - $W$  test, it does not necessarily mean the presence of first-order autocorrelation in the disturbances. It could happen because of other reasons also, e.g.,

- distribution may follow higher-order  $AR$  process.
- some important variables are omitted.
- dynamics of the model is misspecified.
- functional term of the model is incorrect.

## Estimation procedures with autocorrelated errors when autocorrelation coefficient is known

Consider the estimation of regression coefficient under first-order autoregressive disturbances and the autocorrelation coefficient is known. The model is

$$y = X\beta + u,$$

$$u_t = \rho u_t + \varepsilon_t$$

and assume that  $E(u) = 0$ ,  $E(uu') = \psi \neq \sigma^2 I$ ,  $E(\varepsilon) = 0$ ,  $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I$ .

The OLSE of  $\beta$  is unbiased but not, in general, efficient, and the estimate of  $\sigma^2$  is biased. So we use generalized least squares estimation procedure, and GLSE of  $\beta$  is

$$\hat{\beta} = (X'\psi^{-1}X)^{-1}X'\psi^{-1}y$$

where

$$\psi^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & & \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}.$$

To employ this, we proceed as follows:

1. Find a matrix  $P$  such that  $P'P = \psi^{-1}$ . In this case

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}.$$

2. Transform the variables as

$$y^* = Py, X^* = PX, \varepsilon^* = P\varepsilon.$$

Such transformation yields

$$y^* = \begin{bmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_n - \rho y_{n-1} \end{bmatrix}, \quad X^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2} x_{12} & \cdots & \sqrt{1-\rho^2} x_{1k} \\ 1-\rho & x_{22} - \rho x_{12} & \cdots & x_{2k} - \rho x_{1k} \\ 1-\rho & x_{32} - \rho x_{22} & \cdots & x_{3k} - \rho x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & x_{n2} - \rho x_{n-1,2} & \cdots & x_n - \rho x_{n-1} \end{bmatrix}.$$

Note that the first observation is treated differently than other observations. For the first observation,

$$\left(\sqrt{1-\rho^2}\right) y_1 = \left(\sqrt{1-\rho^2}\right) x_1' \beta + \left(\sqrt{1-\rho^2}\right) u_1$$

whereas for other observations

$$y_t = \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + (u_t - \rho u_{t-1}); t = 2, 3, \dots, n$$

where  $x_t'$  is a row vector of  $X$ . Also,  $\sqrt{1-\rho^2} u_1$  and  $(u_t - \rho u_{t-1})$  have the same properties. So we expect these two errors to be uncorrelated and homoscedastic.

If the first column of  $X$  is a vector of ones, then the first column of  $X^*$  is not constant. Its first element is  $\sqrt{1-\rho^2}$ .

Now employ OLSE with observations  $y^*$  and  $X^*$ , then the OLSE of  $\beta$  is

$$\beta^* = (X^{*'} X^*)^{-1} X^{*'} y^*,$$

its covariance matrix is

$$\begin{aligned} V(\hat{\beta}) &= \sigma^2 (X^{*'} X^*)^{-1} \\ &= \sigma^2 (X' \Psi^{-1} X)^{-1} \end{aligned}$$

and its estimator is

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (X' \Psi^{-1} X)^{-1}$$

where

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})' \Psi^{-1} (y - X\hat{\beta})}{n - k}.$$

## Estimation procedures with autocorrelated errors when autocorrelation coefficient is unknown

Several procedures have been suggested to estimate the regression coefficients when autocorrelation coefficient is unknown. The feasible GLSE of  $\beta$  is

$$\hat{\beta}_F = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y$$

where  $\hat{\Omega}^{-1}$  is the  $\Psi^{-1}$  matrix with  $\rho$  replaced by its estimator  $\hat{\rho}$ .

### 1. Use of sample correlation coefficient

The most common method is to use the sample correlation coefficient  $r$  between successive residuals as the natural estimator of  $\rho$ . The sample correlation can be estimated using the residuals in place of disturbances as

$$r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}$$

where  $e_t = y_t - x_t' b$ ,  $t = 1, 2, \dots, n$  and  $b$  is OLSE of  $\beta$ .

Two modifications are suggested for  $r$  which can be used in place of  $r$ .

1.  $r^* = \left(\frac{n-k}{n-1}\right)r$  is the Theil's estimator.
2.  $r^{**} = 1 - \frac{d}{2}$  for large  $n$  where  $d$  is the Durbin Watson statistic for  $H_0 : \rho = 0$ .

## 2. Durbin procedure:

In Durbin procedure, the model

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t, \quad t = 2, 3, \dots, n$$

is expressed as

$$\begin{aligned} y_t &= \beta_0(1 - \rho) + \rho y_{t-1} + \beta x_t - \rho \beta x_{t-1} + \varepsilon_t \\ &= \beta_0^* + \rho y_{t-1} + \beta x_t + \beta^* x_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, n \quad (*) \end{aligned}$$

where  $\beta_0^* = \beta_0(1 - \rho)$ ,  $\beta^* = -\rho\beta$ .

Now run a regression using OLS to model (\*) and estimate  $r^*$  as the estimated coefficient of  $y_{t-1}$ .

Another possibility is that since  $\rho \in (-1, 1)$ , so search for a suitable  $\rho$  which has smaller error sum of squares.

## 3. Cochrane-Orcutt procedure:

This procedure utilizes  $P$  matrix defined while estimating  $\beta$  when  $\rho$  is known. It has following steps:

(i) Apply OLS to  $y_t = \beta_0 + \beta_1 x_t + u_t$  and obtain the residual vector  $e$ .

(ii) Estimate  $\rho$  by  $r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$ .

Note that  $r$  is a consistent estimator of  $\rho$ .

(iii) Replace  $\rho$  by  $r$  is

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t$$

and apply OLS to the transformed model

$$y_t - r y_{t-1} = \beta_0^* + \beta(x_t - r x_{t-1}) + \text{disturbance term}$$

and obtain estimators of  $\beta_0^*$  and  $\beta$  as  $\hat{\beta}_0^*$  and  $\hat{\beta}$  respectively.

This is Cochrane-Orcutt procedure. Since two successive applications of OLS are involved, so it is also called as **two-step procedure**.

This application can be repeated in the procedure as follows:

(I) Put  $\hat{\beta}_0^*$  and  $\hat{\beta}$  in the original model.

(II) Calculate the residual sum of squares.

(III) Calculate  $\rho$  by  $r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$  and substitute it in the model

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t$$

and again obtain the transformed model.

(IV) Apply OLS to this model and calculate the regression coefficients.

This procedure is repeated until convergence is achieved, i.e., iterate the process till the two successive estimates are nearly same so that stability of estimator is achieved.

This is an iterative procedure and is numerically convergent procedure. Such estimates are asymptotically efficient and there is a loss of one observation.

#### 4. Hildreth-Lu procedure or Grid-search procedure:

The Hildreth-Lu procedure has the following steps:

(i) Apply OLS to

$$(y_t - \rho y_{t-1}) = \beta_0(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t, t = 2, 3, \dots, n$$

using different values of  $\rho$  ( $-1 \leq \rho \leq 1$ ) such as  $\rho = \pm 0.1, \mp 0.2, \dots$ .

(ii) Calculate the residual sum of squares in each case.

(iii) Select that value of  $\rho$  for which residual sum of squares is smallest.

Suppose we get  $\rho = 0.4$ . Now choose a finer grid. For example, choose  $\rho$  such that  $0.3 < \rho < 0.5$  and consider  $\rho = 0.31, 0.32, \dots, 0.49$  and pick up that  $\rho$  with the smallest residual sum of squares. Such iteration can be repeated until a suitable value of  $\rho$  corresponding to minimum residual sum of squares is obtained.

The selected final value of  $\rho$  can be used and for transforming the model as in the case of Cochrane-Orcutt procedure. The estimators obtained with this procedure are as efficient as obtained by Cochrane-Orcutt procedure and there is a loss of one observation.

## 5. Prais-Winston procedure

This is also an iterative procedure based on two-step transformation.

(i) Estimate  $\rho$  by  $\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=3}^n e_{t-1}^2}$  where  $e_t$ 's are residuals based on OLSE.

(ii) Replace  $\rho$  by  $\hat{\rho}$  is the model as in Cochrane-Orcutt procedure

$$\left(\sqrt{1-\hat{\rho}^2}\right) y_1 = \left(\sqrt{1-\hat{\rho}^2}\right) \beta_0 + \beta \left(\sqrt{1-\hat{\rho}^2}\right) x_1 + \left(\sqrt{1-\hat{\rho}^2}\right) u_1$$

$$y_t - \hat{\rho} y_{t-1} = (1-\hat{\rho})\beta_0 + \beta(x_t - \hat{\rho}x_{t-1}) + (u_t - \hat{\rho}u_{t-1}), t = 2, 3, \dots, n.$$

(iii) Use OLS for estimating the parameters.

The estimators obtained with this procedure are asymptotically as efficient as the best linear unbiased estimators. There is no loss of any observation.

## (6) Maximum likelihood procedure

Assuming that  $y \sim N(X\beta, \sigma_\varepsilon^2 \psi)$ , the likelihood function for  $\beta, \rho$  and  $\sigma_\varepsilon^2$  is

$$L = \frac{1}{(2\pi\sigma_\varepsilon^2)^{\frac{n}{2}} |\psi|^{\frac{n}{2}}} \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} (y - X\beta)' \psi^{-1} (y - X\beta) \right].$$

Ignoring the constant and using  $|\psi| = \frac{1}{1-\rho^2}$ , the log-likelihood is

$$\ln L = \ln L(\beta, \sigma_\varepsilon^2, \rho) = -\frac{n}{2} \ln \sigma_\varepsilon^2 + \frac{1}{2} \ln(1-\rho^2) - \frac{1}{2\sigma_\varepsilon^2} (y - X\beta)' \psi^{-1} (y - X\beta).$$

The maximum likelihood estimators of  $\beta, \rho$  and  $\sigma_\varepsilon^2$  can be obtained by solving the normal equations

$$\frac{\partial \ln L}{\partial \beta} = 0, \quad \frac{\partial \ln L}{\partial \rho} = 0, \quad \frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = 0.$$

These normal equations turn out to be nonlinear in parameters and can not be easily solved.

One solution is to

- first derive the maximum likelihood estimator of  $\sigma_\varepsilon^2$ .
- Substitute it back into the likelihood function and obtain the likelihood function as the function of  $\beta$  and  $\rho$ .
- Maximize this likelihood function with respect to  $\beta$  and  $\rho$ .



Thus

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = 0 &\Rightarrow -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^2} (y - X\beta)' \psi^{-1}(y - X\beta) = 0 \\ &\Rightarrow \hat{\sigma}_\varepsilon^2 = \frac{1}{n} (y - X\beta)' \psi^{-1}(y - X\beta)\end{aligned}$$

is the estimator of  $\sigma_\varepsilon^2$ .

Substituting  $\hat{\sigma}_\varepsilon^2$  in place of  $\sigma_\varepsilon^2$  in the log-likelihood function yields

$$\begin{aligned}\ln L^* = \ln L^*(\beta, \rho) &= -\frac{n}{2} \ln \left[ \frac{1}{n} (y - X\beta)' \psi^{-1}(y - X\beta) \right] + \frac{1}{2} \ln(1 - \rho^2) - \frac{n}{2} \\ &= -\frac{n}{2} \left[ \ln \left\{ (y - X\beta)' \psi^{-1}(y - X\beta) \right\} - \frac{1}{n} \ln(1 - \rho^2) \right] + k \\ &= k - \frac{n}{2} \ln \left[ \frac{(y - X\beta)' \psi^{-1}(y - X\beta)}{(1 - \rho^2)^{\frac{1}{n}}} \right]\end{aligned}$$

where  $k = \frac{n}{2} \ln n - \frac{n}{2}$ .

Maximization of  $\ln L^*$  is equivalent to minimizing the function

$$\frac{(y - X\beta)' \psi^{-1}(y - X\beta)}{(1 - \rho^2)^{\frac{1}{n}}}.$$

Using the optimization techniques of non-linear regression, this function can be minimized and estimates of  $\beta$  and  $\rho$  can be obtained.

If  $n$  is large and  $|\rho|$  is not too close to one, then the term  $(1 - \rho^2)^{-1/n}$  is negligible and the estimates of  $\beta$  will be same as obtained by nonlinear least-squares estimation.