

Chapter 10

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 the disturbance term is not constant, then heteroskedasticity arises. When $E(u_t, u_{t-s}) = 0, s \neq 0$ is violated, i.e., the variance of the disturbance term remains constant even though successive disturbance terms are correlated, then such a problem is termed an autocorrelation problem.

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

Sometimes the study and explanatory variables have a natural temporal sequence (sequence order over time), i.e., the data are collected over 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

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

Assume ρ_s and γ_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 $u_2)$, ..., $(u_n$ and $u_{n-1})$ gives the autocorrelation of order one, i.e., ρ_1 . Similarly, the autocorrelation between the successive terms $(u_3$ and $u_1)$, $(u_4$ and $u_2)$, ..., $(u_n$ and $u_{n-2})$ provides the autocorrelation of order two, i.e., ρ_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, monthly household expenditure data is influenced by the preceding month's expenditure. Autocorrelation is present in both cross-sectional and time-series data. In the cross-sectional 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 deleting some variables. In regression modelling, it is not possible to include all variables. There can be various reasons for this, e.g., some variables may be qualitative, direct observations may not be available for some variables, etc. The joint effect of these deleted variables leads to autocorrelation in the data.
3. The misspecification of the form of the relationship can also introduce autocorrelation in the data. It is assumed that the form of the relationship between the study and explanatory variables is linear. If the model contains log or exponential terms, which can make its linearity questionable, this can also introduce autocorrelation in the data.
4. The difference between the observed and true values of variable is called measurement error or errors-in-variables. The presence of measurement errors on the dependent variable may also introduce autocorrelation in the data.

Structure of disturbance term

Consider the situation where the disturbances are autocorrelated,

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

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 exceeds the number of observations, the situation is not ideal from a statistical perspective. To handle the situation, some special forms and structures for the disturbance term are assumed so that the number of parameters in its covariance matrix 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.

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

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

$$E(u_t) = 0$$

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

$$E(u_t^2) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \text{ for all } i.$$

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

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

In general,

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

$$E(uu') = \Omega = \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.$$

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

Consequences of autocorrelated disturbances

Consider the model with first-order autoregressive disturbances

$$y = X\beta + u$$

$$u_t = \rho u_{t-1} + \varepsilon_t, t = 1, 2, \dots, n$$

with assumptions

$$E(u) = 0, E(uu') = \Omega$$

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

where Ω is a positive definite matrix.

The ordinary least squares estimator of β is

$$b = (X'X)^{-1} X'y$$

$$= (X'X)^{-1} X'(X\beta + u)$$

$$b - \beta = (X'X)^{-1} X'u$$

$$E(b - \beta) = 0.$$

So OLSE remains unbiased under autocorrelated disturbances.

The covariance matrix of b is

$$V(b) = E(b - \beta)(b - \beta)'$$

$$= (X'X)^{-1} X' E(uu') X (X'X)^{-1}$$

$$= (X'X)^{-1} X' \Omega X (X'X)^{-1}$$

$$\neq \sigma_u^2 (X'X)^{-1}.$$

The residual vector is

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

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 σ^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

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

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

The GLSE of β 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 the best linear unbiased estimator of β .

Durbin-Watson test

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

$$H_0 : \rho = 0$$

Use OLS to estimate β 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

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

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

For large n ,

$$d \approx 1 + 1 - 2r$$

$$d \approx 2(1 - r)$$

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, the 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 the $D-W$ 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 the values of x_i '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 a case, **Durbin's h test** is used.

Estimation procedures with autocorrelated errors when the autocorrelation coefficient is known

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

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

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

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

The OLSE of β is unbiased but not, in general, efficient and an estimate of σ^2 is biased. So we use generalized least squares estimation procedure and GLSE of β is

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

where

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

Estimation procedures with autocorrelated errors when the autocorrelation coefficient is unknown

Several procedures have been proposed for estimating regression coefficients when the autocorrelation coefficient is unknown. The feasible GLSE of β is

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

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

The following are some of the methods used to estimate ρ :

1. Use of the sample correlation coefficient

The most common method is to use the sample correlation coefficient r as the natural estimator of ρ . 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 the OLSE of β .

2. Durbin procedure
3. Cochran-Orcutt procedure
4. Hildreth-Lu procedure or Grid-search procedure
5. Prais-Winsten procedure
6. Maximum likelihood procedure