Unit IV AUTO-CORRELATION

Auto-Correlation is also called serial Correlation. One assumption in least square model is that successive disturbance terms are drawn at random.

i.e., $E(u_i u_j) = 0$ for $i \neq j = 1, 2, 3, \ldots, n$

It implies that when observation are made over time, the effect of the disturbance occuring at one period does not carry over to another period. Suppose we are studying Consumption function. Here Independence of disturbance term means that change in Consumption due to marriage or some other cause is purely temporary and only the Current month's Consumption is affected. In Cross-section data it means that if the Consumption of one family is disturbed, by the visit of a relative, this does not affect the consumption of any other families.

But this assumption may not be realistic particularly in Case of time series observations one would suspect that a large number of random and independent factors operating in one period would, in part, carry over to following periods. Thus the factors which are assumed as random factors start behaving as permanent factors which causes Auto-Correlation between successive disturbances. In other words it can be said that if the basic assumption of least square estimator does not fulfil or

$E(u_i u_j) \neq 0$ where $i \neq j = 1, 2, \ldots, n$

i.e., there establishes the relationship between u_i and u_j, the auto correlation come in force.

Causes of Autocorrelation (or) Sources of Auto-Correlation

There are several causes which give rise to autocorrelation. Fallowing are the main causes:

i) Omitted Explanatory Variables
ii) Mis-specification of the Mathematical form of the model
iii) Mis-specification of the True random term 'U'.
iv) Interpolations in the statistical Observations.

i) Omitted Explanatory variables

The foremost cause is of omitted variables. As we know in economics one variable is affected by so many variables. The investigator includes only important and directly hit variables. Because neither he is competent nor the circumstances allow him to include all variables influencing that phenomena. The errorterm represents the influence of omitted variables and because of that, an error term in one period may have a relation with the error terms in successive periods. Thus the problem of Auto-Correlation arises.

ii) Mis-specification of the Mathematical form of the model

The other cause is the mis-specification of the relationship. For example, let us suppose the investigator suspects linear relation between dependent (Y) and independent (X) variable while the true relation is quadratic one i.e., relation containing X and X^2 terms. In such a case the disturbance term will contain the X^2 term which causes to inter-relation among successive disturbance term. Thus even if the disturbance term in true relation may not be auto-correlated, the term will be auto-correlated in mis-specified relation.

iii) Mis-specification of the True Random Term U

Disturbance term may be auto-correlated because it contains errors of measurement. Once an investigator has committed an error he will not be able to correct it. If the explanatory variable is measured wrongly, the serial disturbances will be auto-correlated.

Effect of Auto-Correlation

Let us take a simple two variable model,

$$Y_i = \alpha + \beta X_i + u_i$$

where, u_i has first order regression scheme as such that

$$u_i = \rho u_{i-1} + \epsilon_i$$

where ρ = Co-efft of the autocorrelation relationship

ϵ_i = a random term which fulfils all the usual assumptions of a r.v.

In this relationship $-1 < \rho < 1$ and satisfies the following assumptions,

$$\left.\begin{array}{l} E(\epsilon_i) = 0 \\ E(\epsilon_i^2) = \sigma_\epsilon^2 \\ E(\epsilon_i \epsilon_j) = 0 \text{ for } i \neq j \end{array}\right\} \text{for all } i = 1, 2, 3, \ldots, \infty$$

Now since $u_i = \rho u_{i-1} + \epsilon_i$

$$\therefore u_{i-1} = \rho u_{i-2} + \epsilon_{i-1}$$

On putting value of u_{i-1} in previous equation we get

$$u_i = \rho(\rho u_{i-2} + \epsilon_{i-1}) + \epsilon_i$$

$$= \rho^2 u_{i-2} + \rho\epsilon_{i-1} + \epsilon_i$$

$$= \rho^2(\rho u_{i-3} + \epsilon_{i-2}) + \rho\epsilon_{i-1} + \epsilon_i$$

$$= \rho^3 u_{i-3} + \rho^2\epsilon_{i-2} + \rho\epsilon_{i-1} + \epsilon_i$$

$$= \rho^3(\rho u_{i-4} + \epsilon_{i-3}) + \rho^2\epsilon_{i-2} + \rho\epsilon_{i-1} + \epsilon_i$$

$$= \rho^4 u_{i-4} + \rho^3\epsilon_{i-3} + \rho^2\epsilon_{i-2} + \rho\epsilon_{i-1} + \epsilon_i$$

or, $u_i = \epsilon_i + \rho\epsilon_{i-1} + \rho^2\epsilon_{i-2} + \rho^3\epsilon_{i-3} + \ldots$

$$= \sum_{r=0}^{\infty} \rho^r \epsilon_{i-r}$$

(iv) Interpolations in the Statistical Observations

Most of the published time series data involve some interpolation and 'smoothing' processes which do average the true disturbances over successive time periods. As a Consequence the successive values of the 'u' are interrelated and exhibit auto-correlation patterns

Meaning of the Assumption of Serial independence

One of the assumptions of OLS is that the successive values of the r.v 'u' are temporally independent. i.e., the value which 'u' assumes in any one period is independent from the value it assumed in any previous period. This implies that

$$Cov(u_i u_j) = E\{[u_i - E(u_i)][u_j - E(u_j)]\}$$

$$= E(u_i u_j) = 0 \quad \text{for } i \neq j = 1, 2, 3, \ldots, n.$$

$$\text{since } E(u_i) = E(u_j) = 0.$$

If this assumption is not satisfied

i.e., if $E(u_i u_j) \neq 0$ for $i \neq j = 1, 2, \ldots, n$

ie, if the value of 'u' in any particular period is correlated with its own preceding value (or values) we say that there is auto-correlation or serial correlation of the r.v.

Auto-correlation is a special case of correlation. Auto-correlation refers to the relationship, not between two ~~two~~ (or more) different variables, but between the successive values of the same variable. Here we are particularly interested in the auto-correlation of u's.

The linear relationship between any two successive values of u is given by

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

Where u_t = The value of u assumes in period t

u_{t-1} = " " " t-1

ρ = the co-efft of the auto correlation relationship

v_t = a random term which fulfills all the usual assumptions of a r.v.

ie., $E(v)=0$; $E(v^2)=\sigma_v^2$; $E(v_i v_j)=0$ for $i \neq j$

This is known as a first-order auto regressive relationship. We will begin our analysis with this form of simple relationship of the u's. In particular we will deal with the simple autocorrelation co-efficient $\rho_{u_t u_{t-1}}$.

Graphical method of detecting auto-correlation

We may obtain a rough idea of the existence or absence of auto correlation in the u's by plotting the values of the regression residuals, e's, on a two-dimensional diagram. The e's are estimates of the true values of u; thus if the e's are correlated this suggests autocorrelation of the true u's.

In drawing a scatter diagram, the variables whose correlation we attempt to detect in this case are e_t and e_{t-1}.

Variable I	Variable II
e_{t+1} (or) (e_2)	e_t (or) (e_1)
e_{t+2} (e_3)	e_{t+1} (e_2)
e_{t+3} (e_4)	e_{t+2} (e_3)
⋮	⋮
	$e_{t+(n-2)}$ (e_{n-2})
$e_{t+(n-1)}$ (e_{n-1})	
e_{t+n} (e_n)	$e_{t+(n-1)}$ (e_{n-1})

The observational points to be plotted are $e_t e_{t-1}$, or $(e_1, e_2), (e_2, e_3), (e_3, e_4) \ldots (e_n, e_{n-1})$.

By definition, mean of both variables is zero ($\bar{e} = 0$). Hence the perpendiculars which pass through the means are actually the two orthogonal axes

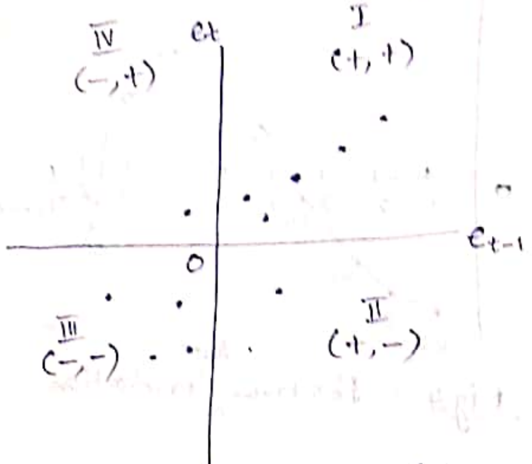


Fig 1: Positive auto Correlation

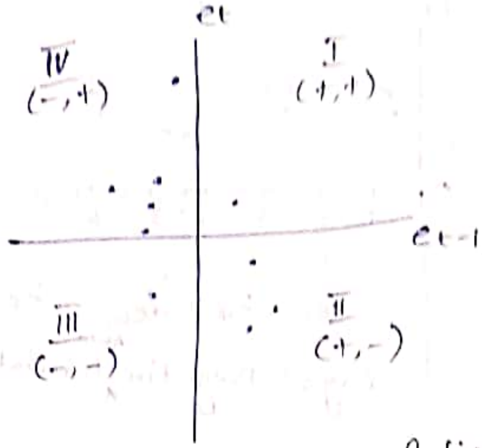


Fig 2: Negative auto correlation.

It is clear from the Figures 1 & 2 that, if most of the points (e_t, e_{t-1}) fall in quadrants I and III (Fig 1), the auto-Correlation will be positive, since the products $(e_t)(e_{t-1})$ are positive. If most of the points (e_t, e_{t-1}) fall in quadrants II & IV (Fig 2) the auto-Correlation will be negative, since the products $(e_t)(e_{t-1})$ are negative. In practice auto Correlation is in most cases positive, because of the economic growth and cyclical movements of economy.

Another method Commonly used in applied econometric research for the detection of auto Correlation is to plot the values of e's against time. If the e's in successive periods show a regular time pattern, we Conclude that there is auto-Correlation in the function.

In general if the successive values of the e's change sign frequently (Fig 3) auto correlation is negative. If the e's do not change sign frequently so that several positive e's are followed by several negative values of e (Fig 4) auto-correlation is positive.

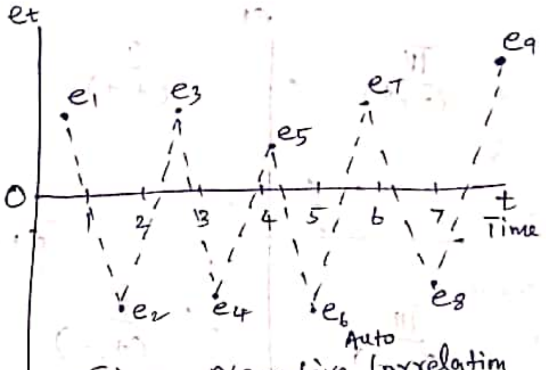


Fig 3: Negative Auto Correlation

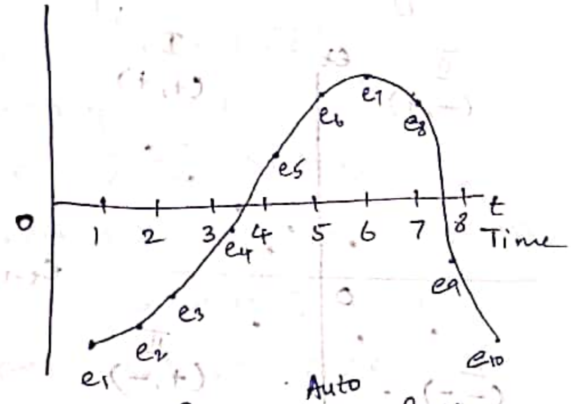


Fig 4: Positive Auto Correlation

The first-order auto regressive Scheme – Autocorrelation problem

Here we analyse the auto-correlation problem to the simple first-order auto regressive scheme. First we establish the mean, variance and co-variance of u when its values are correlated with simple Markov process.

The autoregressive structure is

$$u_t = \rho u_{t-1} + V_t \quad \text{with } |\rho| < 1$$

Where ρ = the co-efft of the auto correlation relationship (ie., ρ = first order auto correlation co-efft $\rho_{u_t u_{t-1}}$)

V_t = a random term which fulfills all the usual assumptions of a r.v

ie. $E(V) = 0$; $E(v^2) = \sigma_v^2$; $E(v_i v_j) = 0$ for $i \neq j$

The Complete form of the first-order Markov process (the pattern of auto correlation for all the values of u), is

$$u_t = \rho u_{t-1} + v_t$$
$$u_{t-1} = \rho u_{t-2} + v_{t-1}$$
$$u_{t-2} = \rho u_{t-3} + v_{t-2}$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$u_{t-r} = \rho u_{t-(r+1)} + v_{t-r}$$

In order to define the error term in any particular period 't' we work as follows:

The auto correlation relationship in period 't' is

$$u_t = \rho u_{t-1} + v_t$$
$$\therefore \quad u_{t-1} = \rho u_{t-2} + v_{t-1}$$

On putting value of u_{t-1} in previous equation, we get

$$u_t = \rho(\rho u_{t-2} + v_{t-1}) + v_t$$
$$= \rho^2 u_{t-2} + \rho v_{t-1} + v_t$$
$$= \rho^2(\rho u_{t-3} + v_{t-2}) + \rho v_{t-1} + v_t$$
$$= \rho^3 u_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t$$
$$= \rho^3(\rho u_{t-4} + v_{t-3}) + \rho^2 v_{t-2} + \rho v_{t-1} + v_t$$
$$= \rho^4 u_{t-4} + \rho^3 v_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t$$

In general

$$u_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \dots$$
$$= \sum_{r=0}^{\infty} \rho^r v_{t-r}$$

This is the value of the error term when it is auto-correlated with a first-order auto regressive scheme.

Scanned with CamScanner

(i) Mean of the autocorrelated U's

$$E(U_t) = E\left[\sum_{r=0}^{\infty} \rho^r V_{t-r}\right]$$

$$= \sum_{r=0}^{\infty} \rho^r E(V_{t-r})$$

$$= E(V_t) + \rho E(V_{t-1}) + \rho^2 E(V_{t-2}) + \cdots$$

$$= 0 \quad \text{since } E(V_t) = 0$$

$\therefore E(U_t) = 0 \quad , \quad t = 1, 2, \ldots n.$

(ii) Variance of the autocorrelated U's

$$E(U_t^2) = E\left[\sum_{r=0}^{\infty} \rho^r V_{t-r}\right]^2$$

$$= \sum_{r=0}^{\infty} (\rho^r)^2 E(V_{t-r})^2$$

$$= \sum_{r=0}^{\infty} (\rho^r)^2 Var(V_{t-r})$$

$$= \sum_{r=0}^{\infty} \rho^{2r} \sigma_v^2$$

$$= \sigma_v^2 (1 + \rho^2 + \rho^4 + \rho^6 + \cdots)$$

$$= \frac{\sigma_v^2}{1-\rho^2} \quad \left(\text{since in G.P. } 1 + r + r^2 + \cdots = \frac{1}{1-r}\right)$$

$\therefore E(U_t^2) = \sigma_v^2 \left[\frac{1}{1-\rho^2}\right]$ (the progression converges since $|\rho| < 1$).

(or) $Var(U_t) = \dfrac{\sigma_v^2}{1-\rho^2}$

(iii) The Covariance of the autocorrelated U's

Given that $U_t = V_t + \rho V_{t-1} + \rho^2 V_{t-2} + \cdots$

and $U_{t-1} = V_{t-1} + \rho V_{t-2} + \rho^2 V_{t-3} + \cdots$

We obtain

$$Cov(U_t U_{t-1}) = E\{[U_t - E(U_t)][U_{(t-1)} - E(U_{(t-1)})]\}$$

$$= E[U_t U_{t-1}]$$

$$\text{Cov}(u_t u_{t-1}) = E\left[(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \dots)(v_{t-1} + \rho v_{t-2} + \rho^2 v_{t-3} + \dots)\right]$$

$$= E\left[\{v_t + \rho(v_{t-1} + \rho v_{t-2} + \dots)\}(v_{t-1} + \rho v_{t-2} + \rho^2 v_{t-3} + \dots)\right]$$

$$= E\left[(v_t)(v_{t-1} + \rho v_{t-2} + \rho^2 v_{t-3} + \dots)\right] + E\left[\rho(v_{t-1} + \rho v_{t-2} + \dots)^2\right]$$

$$= 0 + \rho\, E(v_{t-1} + \rho v_{t-2} + \dots)^2$$

$$= \rho\, E(v_{t-1}^2 + \rho^2 v_{t-2}^2 + \dots + \text{Cross products})$$

$$= \rho(\sigma_v^2 + \rho^2 \sigma_v^2 + \dots + 0)$$

$$= \rho\left[\sigma_v^2 (1 + \rho^2 + \rho^4 + \rho^6 + \dots)\right]$$

$$= \rho\, \sigma_v^2 \frac{1}{1-\rho^2} \quad (\text{for } |\rho| < 1)$$

$$= \rho\, \sigma_u^2$$

|||^{ly}, $\text{Cov}(u_t u_{t-2}) = \rho^2 \sigma_u^2$ and in general

$\text{Cov}(u_t u_{t-s}) = \rho^s \sigma_u^2 \quad (\text{for } s \neq t)$

Hence, When there is autocorrelation of the simple form of first-order autoregressive scheme, the autocorrelated disturbance term has the following characteristics.

i) $u_t = \sum_{r=0}^{\infty} \rho^r v_{t-r}$

ii) $E(u_t) = 0$

iii) $\text{Var}(u_t) = \sigma_u^2 \frac{1}{1-\rho^2} = \sigma_u^2$

iv) $\text{Cov}(u_t u_{t-s}) = \rho^s \sigma_u^2 \neq 0$.

Where v is the random term and ρ is the autocorrelation co-effᵗ in the first order autoregressive scheme

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

Scanned with CamScanner

Effects (or) Consequences of Auto-Correlation

When the disturbance term exhibits serial Correlation the value as well as the standard errors of the parameter estimates are affected. In particular

1. Even when the residuals are serially correlated the parameter estimates of OLS are statistically unbiased.

Consider the eqn. $Y_i = \alpha + \beta X_i + u_i$

Where u_i is auto regressive, the least square estimator of β is

$$\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$= \frac{\Sigma x_i (\beta x_i + u_i)}{\Sigma x_i^2}$$

$$= \beta \cdot \frac{\cancel{\Sigma x_i^2}}{\cancel{\Sigma x_i^2}} + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$$

Taking expectations,

$$E(\hat{\beta}) = \beta \quad \text{since } E(u_i) = 0.$$

2. With autocorrelated values of the disturbance term the OLS variances of the parameter estimates, are likely to be larger than those of other econometric methods.

$$Var(\hat{\beta}) = E\left[\hat{\beta} - \beta\right]^2$$

$$= E\left[\frac{\Sigma x_i u_i}{\Sigma x_i^2}\right]^2$$

$$= \left(\frac{1}{\Sigma x_i^2}\right)^2 E\left[\Sigma x_i u_i\right]^2$$

$$= \frac{1}{(\Sigma x_i^2)^2} E\left[\Sigma x_i^2 u_i^2 + (2\Sigma x_i x_{i-1} u_i u_{i-1}) + 2\Sigma x_i x_{i-2} u_i u_{i-2} + \dots\right]$$

$$Var(\hat{\beta}) = \frac{1}{(\Sigma x_i^2)^2}\left[\sigma^2 \Sigma x_i^2 + 2\rho\sigma_u^2 \Sigma x_i x_{i-1} + 2\rho^2\sigma_u^2 \Sigma x_i x_{i-2} + \cdots\right]$$

$$= \frac{\sigma_u^2}{\Sigma x_i^2} + \frac{\sigma_u^2}{\Sigma x_i^2}\left[2\rho \frac{\Sigma x_i x_{i-1}}{\Sigma x_i^2} + 2\rho^2 \frac{\Sigma x_i x_{i-2}}{\Sigma x_i^2} + \cdots\right]$$

Thus $Var(\hat{\beta})$ of model Containing autocorrelation differs from those given by least squares.

3. The variance of the random term 'ui' may be seriously underestimated if the u's are autocorrelated.

4. If the values of 'u' are autocorrelated, the predictions based on OLS estimate will be ~~an~~ inefficient, in the sense that they will have a larger variance as Compared with predictions based on estimates obtained from other econometric methods.

5. Due to wrong estimation 't' and 'F' test will no longer be valid.

Tests for Auto-Correlation

We Can get some rough idea of the existence and the pattern of auto correlation may be gained by plotting the regression residuals either against their own lagged value(s), or against time.

However, there are more accurate tests for the incidence of auto-correlation. The important tests are

1. Von Neumann ratio and
2. Durbin-Watson test

The Durbin – Watson Test

Durbin-Watson ~~base~~ test gives us a technique which can detect the presence of auto correlation. It will be applicable to small samples. However the test is appropriate only for the first-order auto regressive scheme ($u_t = \rho u_{t-1} + v_t$). The test may be outlined as follows.

The null hypothesis is $H_0 : \rho = 0$

(or) H_0 : the u's are not autocorrelated with a first-order scheme

against the alternative hypothesis $H_1 : \rho \neq 0$.

(or) H_1 : the u's are autocorrelated with a first-order scheme.

To test the null hypothesis we use the Durbin – Watson statistic

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

Where 'e' represents the OLS residuals. Calculated 'd' is to be compared with pre-established lower and upper critical values d_L and d_u for various values of X's and sample sizes.

The decision rules are

a) Reject the Null hypothesis if $d < d_L$

b) Do not reject the hypothesis if $d > d_u$

c) The test is inconclussive if $d_L < d < d_u$

If in case the value of d exceeds 2, also test it against the alternative hypothesis of negative first order auto-correlation.

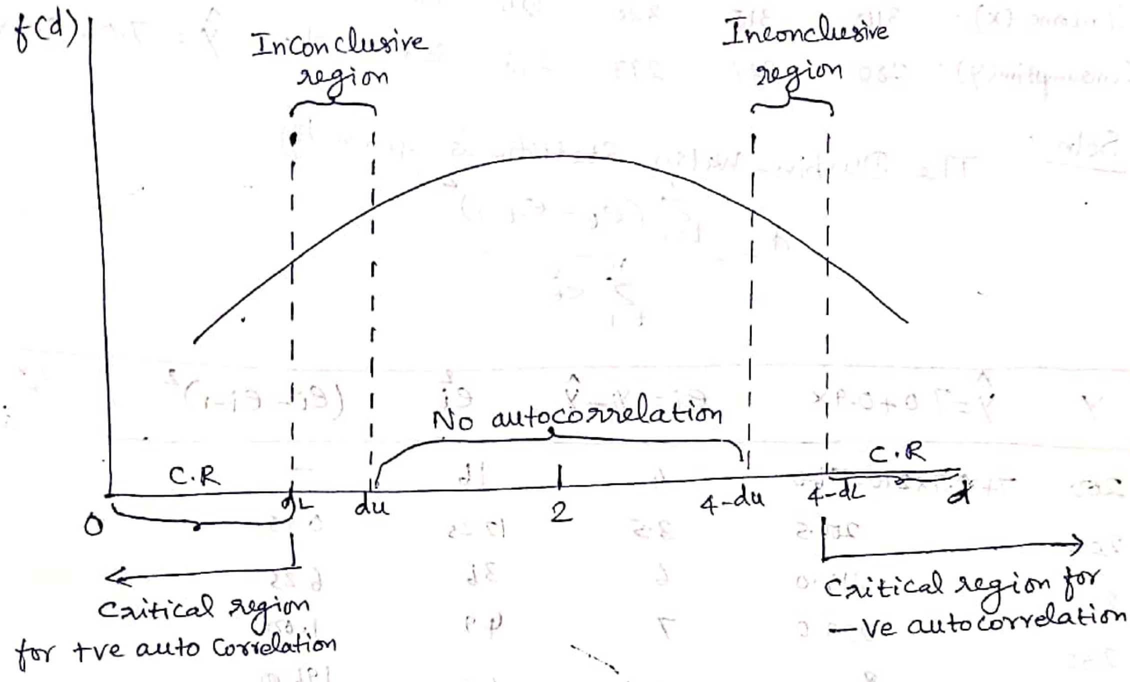
Then decision rules will be

a) Reject the Null hypothesis if $d > 4-d_L$

b) Do not reject it if $d < 4-d_u$

c) Test is inconclussive if $4-d_u < d < 4-d_L$

The critical regions of d-test are shown in the following figure:



Shortcomings of d Statistic

There are several shortcomings of d-statistic.

1. It is not an appropriate technique to measure auto-correlation, if among the explanatory variables there are lagged values of endogenous variables.

2. The Durbin-Watson test is inconclusive if $d_L < d < d_u$ and $4-d_u < d < 4-d_L$. Therefore the result does not have a universal application.

3. The Durbin-Watson test is inappropriate for testing for higher order serial correlation or for other form of autocorrelation (eg. nonlinear forms of serial dependence of the values u_t).

Example :

⊙ On the basis of the following data test the N.H. $\rho=0$ Vs the alternative $\rho>0$ at 5% L.O.S.

Year :	1948	49	50	51	52	53	54	55	56	57	58
Income (X) :	210	215	230	240	250	255	260	270	280	290	300
Consumption (Y) :	200	204	220	230	225	235	240	255	260	270	275

Year :	1959	60	61	62	63
Income (X) :	310	315	320	325	330
Consumption (Y) :	280	289	293	298	303

where $\hat{Y} = 7.0 + 0.9X$.

Soln :

The Durbin-Watson Statistic is given by

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

Y	$\hat{Y}=7.0+0.9X$	$e_i = Y_i - \hat{Y}$	e_i^2	$(e_i - e_{i-1})^2$
200	7+0.9×210=196.0	4	16	–
204	200.5	3.5	12.25	0.25
220	214.0	6	36	6.25
230	223.0	7	49	1.00
225	232.0	−7	49	196.00
235	236.5	−1.5	2.25	30.25
240	241.0	−1	1	0.25
255	250.0	5	25	36.00
260	259.0	1	1	16.00
270	268.0	2	4	1.00
275	277.0	−2	4	16.00
280	286.0	−6	36	16.00
289	290.5	−1.5	2.25	20.25
293	295.0	−2	4	0.25
298	299.5	−1.5	2.25	0.25
303	304.0	−1	1	0.25
			245	340.00

$$\therefore d = \frac{\Sigma (e_i - e_{i-1})^2}{\Sigma e_i^2} = \frac{340}{245} = 1.387$$

From the given table, for $k = 1$ and $n = 16$

$d_L = 1.10$ and $d_U = 1.37$

Since $d = 1.387 > d_U$, the N.H. is true or we can say that there is no autocorrelation between error terms.