

Autocorrelation

Autocorrelation:

In statistics, the autocorrelation of a random process describes the correlation between values of the process at different points in time, as a function of the two times or of the time difference. Let X be some repeatable process, and i be some point in time after the start of that process. (i may be an integer for a discrete-time process or a real number for a continuous-time process.) Then X_i is the value (or realization) produced by a given run of the process at time i . Suppose that the process is further known to have defined values for mean μ_i and variance σ_i^2 for all times i . Then the definition of the autocorrelation between times s and t is

$$R(s, t) = \frac{E[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s},$$

where "E" is the expected value operator. Note that this expression is not well-defined for all time series or processes, because the variance may be zero (for a constant process) or infinite. If the function R is well-defined, its value must lie in the range $[-1, 1]$, with 1 indicating perfect correlation and -1 indicating perfect anti-correlation.

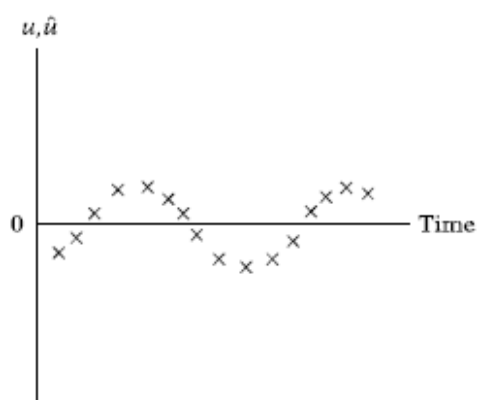
If X_t is a second-order stationary process then the mean μ and the variance σ^2 are time-independent, and further the autocorrelation depends only on the difference between t and s : the correlation depends only on the time-distance between the pair of values but not on their position in time. This further implies that the autocorrelation can be expressed as a function of the time-lag, and that this would be an even function of the lag $\tau = s - t$. This gives the more familiar form

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2},$$

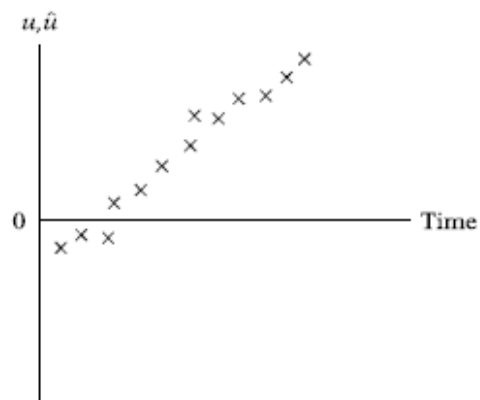
and the fact that this is an even function can be stated as

$$R(\tau) = R(-\tau).$$

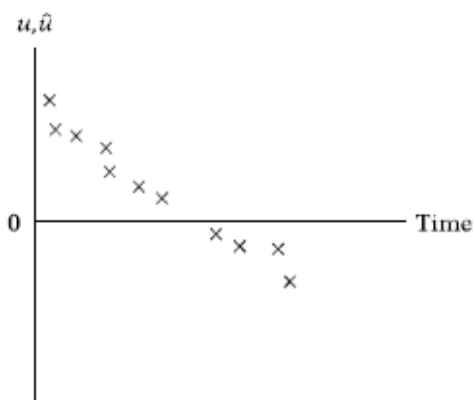
It is common practice in some disciplines, other than statistics and time series analysis, to drop the normalization by σ^2 and use the term "autocorrelation" interchangeably with "auto covariance". However, the normalization is important both because the interpretation of the autocorrelation as a correlation provides a scale-free measure of the strength of statistical dependence, and because the normalization has an effect on the statistical properties of the estimated autocorrelations.



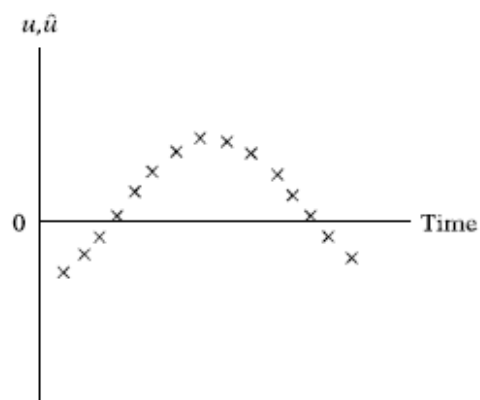
(a)



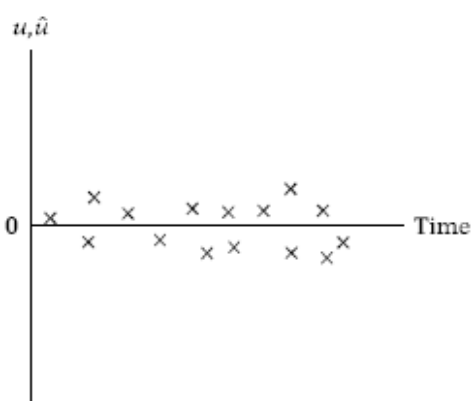
(b)



(c)



(d)



(e)

Patterns of autocorrelation and nonautocorrelation.

What is auto regression?

A regression is known as auto regression if one of the explanatory variables is the lagged value of the dependent variable .

In a time series regression of consumption expenditure on income it is not uncommon to find that the consumption expenditure in the current period depends , among other things , on the consumption expenditure of the previous period .

That is the equation

$$\text{Consumption}_t = \beta_1 + \beta_2 \text{income}_t + \beta_3 \text{consumption}_{t-1} + u_t$$

is known as auto regression .

What is AR1 Model? Write down its assumption?

Let $Y_t = \beta_1 + \beta_2 X_t + u_t$, where t denotes the data or observations at time t . One can assume the disturbance terms as follows

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

Where ρ is known as the coefficient of auto covariance & ε_t is the disturbance term.

The model is known as AR1 model. Because it can be interpreted as the regression of u_t on itself lagged one period i.e. its immediate past value are involved.

Assumptions: (1) $E(u_t) = 0$, (2) $Var(u_t) = \sigma_u^2$, (3) $cov(u_t, u_{t-1}) \neq 0$

What are the sources of auto correlation?

Sources of autocorrelation:

There are several causes which gives rise to autocorrelation. Following are;

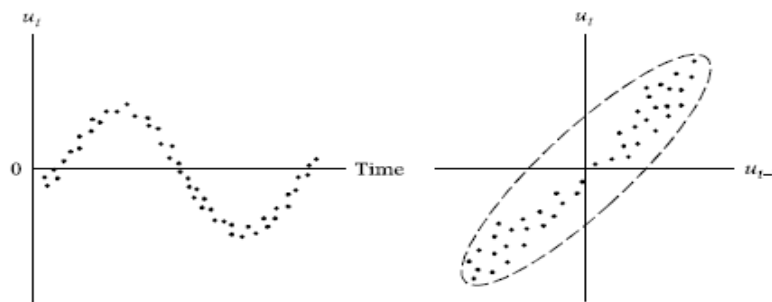
1. Omitted explanatory variables : We know in economics one variable is affected by so many variables .The investigator includes only important and directly fit variables because neither he is competent nor the circumstances allow him to include all variables influencing that phenomena .The error term 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 autocorrelation arises .
2. Misspecification of the mathematical form of the model: If we have adopted a mathematical form which differs from the true form of the relationship, then the disturbance term must show serial correlation.
3. Interpolation in the statistical observation: Most of the published time series data involve some of the interpolation and smoothing process, which do average the true disturbance over successive time periods. As a consequence the successive values are autocorrelated .

4. Misspecification of the true error u : Disturbance term may be autocorrelated because it contains errors, of measurement. If the explanatory variable is measured wrongly the serial disturbance will be auto correlated.

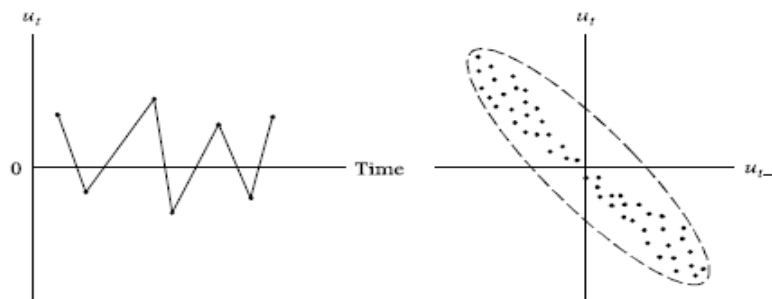
What are the consequences of autocorrelation?

Answer: - The consequences of autocorrelation are given below –

1. When the disturbance term is are seriously correlated then the least square (OLS) estimation are unbiased but optimist property (minimum variance property) is not satisfied .
2. If the disturbance term is are autocorrelated then the OLS variance is greater than the variances of estimate calculated by other method therefore the useal of t and f of significance are no longer valid .
3. The variance of random term u may be seriously underestimated if the u 's are autocorrelated.
4. If the disturbance terms are autocorrelated then the OLS estimate are not asymptotic.
- 5 .Over estimation of ρ^2 .



(a)



(b)

(a) Positive and (b) negative autocorrelation.

Serial correlation [Lagged correlation] :

Sometime autocorrelation is termed as serial correlation. `Tintner defines autocorrelation as “lag correlation of a given series with itself , lagged by a number of time units ” . Where as he reserves the term serial correlation as “lag correlation between”two different time series.

Thus correlation between two time sires $u_1, u_2 \dots, u_{10}$ and v_2, v_3, \dots, v_{11} ,where the former in the latter series lagged by one time period , is autocorrelation where as correlation between such as $u_1, u_2 \dots, u_{10}$ & v_2, v_3, \dots, v_{11} where U and V are two different time series is called serial correlation .

Distinguish between auto correlation and serial correlation:

The distinctions between autocorrelation and serial correlation are given below.

Let us consider two time series data as $u_1, u_2 \dots, u_{10}$, and v_2, v_3, \dots, v_{11}

1. When the correlation occurs in same series then the correlation is called autocorrelation. But when the correlation occurs in different time series then it is called serial correlation.
- 2 .Correlation of a series with itself , lagged by a number of time unit , is know as autocorrelation lagged correlation between two different series is known as serial correlation .

Detecting Autocorrelation: Graphical Method

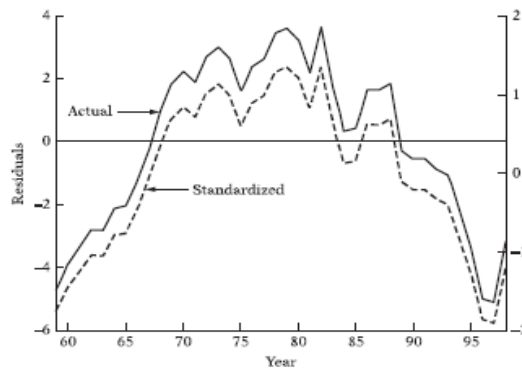


FIGURE 12.8 Residuals and standardized residuals from the wages–productivity regression (12.5.1).

TABLE 12.5 RESIDUALS: ACTUAL, STANDARDIZED, AND LAGGED

Observation	RES1	SRES1	RES1(-1)	Observation	RES1	SRES1	RES1(-1)
1959	-4.703979	-1.758168		1979	3.602089	1.346324	3.444821
1960	-3.874907	-1.448293	-4.703979	1980	3.230723	1.207521	3.602089
1961	-3.359494	-1.255651	-3.874907	1981	2.188868	0.818116	3.230723
1962	-2.800911	-1.046874	-3.359494	1982	3.631600	1.357354	2.188868
1963	-2.828229	-1.057084	-2.800911	1983	1.733354	0.647862	3.631600
1964	-2.112378	-0.789526	-2.828229	1984	0.320571	0.119817	1.733354
1965	-2.039697	-0.762361	-2.112378	1985	0.407350	0.152252	0.320571
1966	-1.252480	-0.468129	-2.039697	1986	1.651836	0.617393	0.407350
1967	-0.280237	-0.104742	-1.252480	1987	1.623640	0.606855	1.651836
1968	0.949713	0.354966	-0.280237	1988	1.838615	0.687204	1.623640
1969	1.835615	0.686083	0.949713	1989	-0.303679	-0.113504	1.838615
1970	2.236492	0.835915	1.835615	1990	-0.560070	-0.209333	-0.303679
1971	1.880977	0.703038	2.236492	1991	-0.559193	-0.209005	-0.560070
1972	2.710926	1.013241	1.880977	1992	-0.885197	-0.330853	-0.559193
1973	3.012241	1.125861	2.710926	1993	-1.056563	-0.394903	-0.885197
1974	2.654535	0.992164	3.012241	1994	-2.184320	-0.816416	-1.056563
1975	1.599020	0.597653	2.654535	1995	-3.355248	-1.254064	-2.184320
1976	2.386238	0.891885	1.599020	1996	-4.996226	-1.867399	-3.355248
1977	2.829847	0.982936	2.386238	1997	-5.137643	-1.920255	-4.996226
1978	3.444821	1.287543	2.829847	1998	-3.278621	-1.225424	-5.137643

Notes: RES 1 = residuals from regression (12.5.1).
SRES 1 = standardized residuals = RES1/2.8755.
RES(-1) = residuals lagged one period.

Note that the important of producing and analyzing plots as a standard part of statistical analysis cannot be overemphasized. Besides occasionally providing an easy to understand summary of a complex problem, they allow the simultaneous examination of the data aggregate while clearly displaying the behavior of individual cases. Here the graph suggests that the residuals are not random.

Run test to detect auto correlation:

A run is defined to be a succession of one or more identical symbols which are followed and proceeded by a different symbol or no symbol at all.

++---++-----++++-----++++--

H_0 : There have no auto correlation problem in the model.

Let n_1 = number of positive symbol (i.e. + residuals)

n_2 = number of negative symbol (i.e. - residuals)

$n = n_1 + n_2$

R = number of runs

- Under the null hypothesis that successive outcomes are independent and assuming that $n_1 \geq 0$ & $n_2 \geq 0$ the number of runs is distributed normally with

$$\text{Mean, } E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\text{Variance, } \sigma^2_R = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

If $n > 25$ then the appropriate test statistics is

$$z = \frac{R - E(R)}{\sqrt{V(R)}} \sim N(0,1)$$

2. If the hypothesis of randomness is sustainable we should expect R the number of runs obtained to lie between $[E(R) \pm 1.96\sigma_R]$ with 95% confidence.
3. Decision rule: Don't reject H_0 of randomness with 95% confidence if $[E(R) - 1.96\sigma_R \leq R \leq E(R) + 1.96\sigma_R]$.
4. Or reject the null hypothesis if the estimated R lies outside the 95% confidence limit.

Durbin Watson d statistic:

Durbin Watson test is appropriate for small sample ($n < 25$). This is the most celebrated test detecting serial correlation which is developed by J. Durbin and G S Watson in 1951.

Hypothesis: Let , $H_0: \rho = 0$ i.e the u 's are not auto correlated.

$H_1: \rho \neq 0$ i.e. the u 's are auto correlated.

Test statistic: the Durbin Watson d statistic's is defined as

$$\begin{aligned} d &= \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^n \hat{u}_t^2} \\ &= \frac{2\sum_{t=2}^n \hat{u}_t^2 - 2\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^n \hat{u}_t^2} \\ &\cong 2 \left[1 - \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sqrt{\sum_{t=2}^n \hat{u}_t^2 \sum_{t=2}^n \hat{u}_{t-1}^2}} \right] \\ &= 2[1 - \rho_{\hat{u}_t, \hat{u}_{t-1}}] \end{aligned}$$

Where $\sum_{t=2}^n \hat{u}_t^2$ & $\sum_{t=2}^n \hat{u}_{t-1}^2$ differ only one observation so they are approximately equal.

$$\therefore d = 2(1 - \rho)$$

When $\rho = 0$ then $d = 2$

$\rho = 1$ then $d = 0$

$\rho = -1$ then $d = 4$

We can see that the value of d lies between 0 to 4.

Mechanisms of Durbin-Watson test

1. Run OLS regression and obtains residuals
2. Compute d statistics by using above formula
3. Find the critical d_L and d_U values for the given sample size and number of explanatory variables.
4. Now follow the decision rules given in the following table or graph.

Conclusion: Durbin Watson has established upper (d_U) & lower (d_L) limits for the significance level of d.

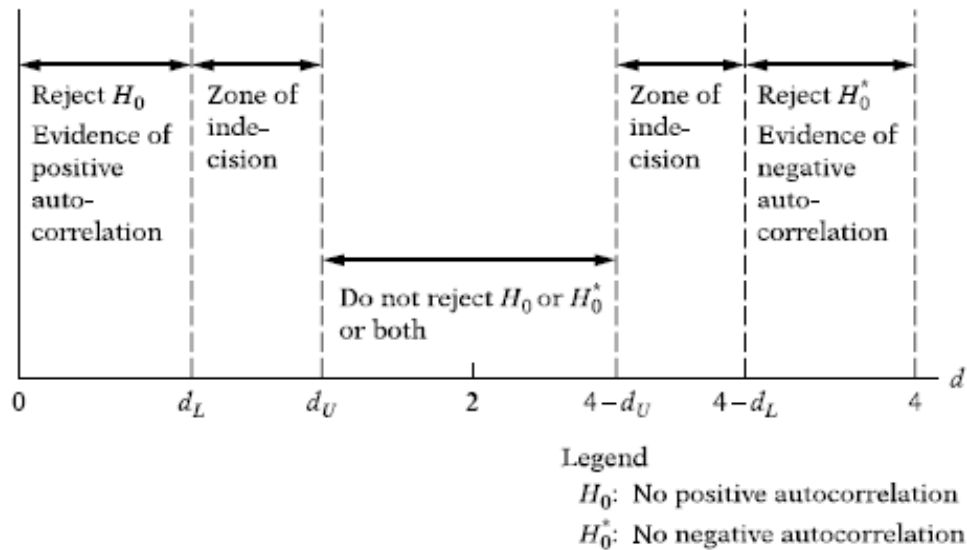


Fig. Decision rules for Durbin-Watson Test

DURBIN-WATSON d TEST: DECISION RULES

Null hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \leq d \leq d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \leq d \leq 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$

²⁵For details, see Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, *Advanced Econometric Methods*, Springer Verlag, New York, 1984, pp. 225–228.

Remedial Measures

If the diagnostic tests suggest that there have an autocorrelation problem then we have some options which are follows as;

1. Try to find out it is pure auto correlation or not.
2. If it is pure autocorrelation then one can transformation of original model so that in the new model we do not have the problem of pure autocorrelation
3. For large sample cases, we can use the New-West method to obtain SE of OLS estimators that are correlated for autocorrelation. It is just extension of White's heteroscedasticity consistent standard errors methods.