Name of the Programme: M A Economics (Sem. IV)

Name of the Course: EC- 1 Group C: Basic Econometrics

Module 4: Problems of Single Equation Model

Name of the Topic: Autocorrelation

By

Dr. Benoy Kumar Lal

Associate Professor

Department of Economics, Patna University, Patna

Email ID: benoykumarlal@yahoo.in

Mobile No. / WhatsApp No. 9608570960/ 9852059001

Autocorrelation: Meaning, Causes and Methods of Detection

(A) Meaning:

Autocorrelation is a type of correlation which refers to the relationship between the successive values of the same variable.

Let us suppose a linear regression

Model - $Y_i = a + bX_i + u_i$

(where $i = 1, 2, \ldots n$)

In Ordinary Least Squares Model, we assume that the error terms(u_i) and (u_j) are not correlated or covariance between u_i and u_j is equal to zero. i.e

$$Cov.\ (u_i\, u_j) = E\left[\{u_i - E(u_i)\} \cdot \{u_j - E(u_j)\}\right] = 0$$

$$Cov.\ (u_i\, u_j) = E(u_i\, u_j) = 0 \quad \text{(for } i \neq j\text{)}$$

[Since $E(u_i)$ & $E(u_j)=0$ Assumption of OLS]

If the above assumption (assumption on non-autocorrelation) is not satisfied, that is, if the value of error term in any particular period (say u_i) is correlated with its own preceding value or values (say u_j), we say that the error terms is autocorrelated or serially correlated.

(B) Causes of Autocorrelation

There are several causes which are responsible for occurance of autocorrelation. Following are the main causes:

1. <u>Inertia:</u> Existence of Inertia in economic time series data is an important cause of autocorrelation (or Serial correlation). National Income, prices, production, employment related time series exhibit cyclical movements. From recession to recovery, most of these series start moving upward and therefore, successive observations become interdependent and causes autocorrelation.

2. <u>Excluded Variables</u>: If the model under consideration is not truly represented by relevant variable (s), it may cause autocorrelation. In fact, the investigator includes only important variables. In this situation, the error term represents the influence of omitted (excluded) variables and because of this exclusion, an error term in one period may have a relation with the error terms in successive periods.

3. <u>Incorrect Functional Form</u>: Suppose the true relationship between independent (x) and dependent (Y) variable is quadratic ($Y_i = a + b_1 X_i + b_2 X_i^2 + u_i$) but for some reasons, we assume relationship between X and Y as linear (i.e. $Y_i = a + b X_i + u_i$). This type of wrong selection of functional relation between variables causes problem of autocorrelation.

4. <u>Errors in Measurement</u>: Disturbance term may be autocorrelated because it contains errors of measurement. Once an investigator has committed an error, he will not be able to rectify it. If the explanatory (independent) variable is measured wrongly, the serial disturbances will be autocorrelated.

5. Lagged Variables: Exclusion of lagged dependent variables as independent variable in the model may result as the problem of autocorrelation. For instance,

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 C_{t-1} + u_t$$

where, C_t = consumption at time t
Y_t = disposable income at time t
C_{t-1} = consumption at time t-1
u_t = disturbance term at time t.

The above equation is autoregressive. Here, present level (at time t) of consumption (C_t) is not only influenced by current income (Y_t) but also on immediate past consumption (C_{t-1}). In the above model if the lagged term (C_{t-1}) is neglected, the resultant error term will show a systematic pattern due to the influence of lagged consumption on current consumption.

6. Mis-specification of true random term: Pure random factors like war, Pandemic, drought, strike etc. exert effect that is spread over more than one period of time. For example strike will have effect on the output which will continue for several future periods.

Detection of Autocorrelation

Various methods or tests are used to detect the presence of autocorrelation. Here, we shall discuss two important methods i.e. (a) Durbin-Watson Test and (b). Von-Neumann Ratio Test in some details.

(a) Durbin-Watson Test: In 1951, J. Durbin and G.S. Watson proposed 'd' statistic to detect the presence of autocorrelation. The method/test is applicable to small samples and appropriate for first order autoregressive (AR) Scheme i.e. $u_t = \rho u_{t-1} + v_t$ where ρ is the autoregression coefficient.

Steps or Procedure of the DW Test

1. Formulation of the Hypothesis i.e.

   $H_0 : \rho = 0$ (i.e. u's are non autocorrelated by first order AR Scheme)

   $H_1 : \rho \neq 0$ (i.e. u's are autocorrelated)

2. Calculation of 'd' statistic with the help of following formula suggested by Durbin and watson:

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

here $e_t = Y_t - \hat{Y}_t$

3. **Interpretation:** Calculated value of 'd' is to be compared with pre-established lower and upper critical values d_L and d_u for number of explanatory variable(s) (k) and number of items/observations in the sample (n) at required level of significance. The Decision rules are:

If d is less than 2,

(a) Reject the Null Hypothesis (H_0) if $d < d_L$

(b) Do not reject the H_0 if $d > d_u$

(c) The test is inconclusive if $d_L < d < d_u$

If in case the value of d exceeds 2, test it against the alternative hypothesis of negative first order autocorrelation. In this case

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

(b) Do not reject the null hypothesis if $d < 4 - d_u$

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

Numerical Example on Durbin-Watson Test –

Consider the model $Y_t = a + bX_t + u_t$ with the following observations on Y on X and test the presence of autocorrelation using Durbin-Watson Test.

X	8	10	12	14	16	18	20	22	24	26
Y	14	13	18	19	22	23	24	28	31	30

Solution: The estimated model will be $\hat{Y}_t = 4.8906 + 1.0182 X_t$.

(Students are required to verify the estimated model.)

Calculation of Durbin-Watson 'd' statistic

X_t	Y_t	$\hat{Y}_t = 4.8906 + 1.0182 X_t$	e_t	$e_t - e_{t-1}$	e_t^2	$(e_t - e_{t-1})^2$
8	14	13.0362	0.9638	–	0.9289	
10	13	15.0726	-2.0726	-3.0364	4.2957	9.2197
12	18	17.1090	0.8910	2.9636	0.7939	8.7829
14	19	19.1454	-0.1454	-1.0364	0.0211	1.0741
16	22	21.1818	0.8182	+0.9636	0.6695	0.9285
18	23	23.2182	-0.2182	-1.0364	0.0476	1.0741
20	24	25.2546	-1.2546	-1.0364	1.5740	1.0741
22	28	27.2910	0.7090	1.9636	0.5027	3.8557
24	31	29.3274	1.6726	0.9636	2.7976	0.9285
26	30	31.3638	-1.3638	-3.0364	1.8600	9.2197
	$\Sigma Y_t = 222$	$\Sigma \hat{Y}_t = 222$			$\Sigma e_t^2 = 13.491$	$\Sigma (e_t - e_{t-1})^2 = 36.1573$

Since $d = \dfrac{\Sigma (e_t - e_{t-1})^2}{\Sigma e_t^2} = \dfrac{36.1573}{13.491} = 2.6801$

at $K = 1$ and $n = 10$; $d_L = 0.604$ & $d_u = 1.001$
(at 1 % level of Significance)

Since, $d > 2$ we have to test the presence of negative first order autocorrelation.

Here, $4 - d_L = 4 - 0.604 = 3.396$

and $4 - d_u = 4 - 1.001 = 2.999$

Since, $d < 4 - d_u$ [as $2.68 < 2.999$]

$\therefore$ Null Hypothesis is true and there is presence of negative first order autocorrelation in the model.

(b) Von-Neumann Ratio Test:

The Von-Neumann Ratio is the ratio of 'Variance of first differences of e' to the 'Variance of e'. The test is used for large samples, ~~and~~ applicable for directly observed series and for variables which are random. For large sample, the Von-Neumann Ratio is

$$\frac{\delta^2}{S_e^2} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2 / N-1}{\sum_{t=1}^{n}(e_t - \bar{e})^2 / N}$$

where $\bar{e} = 0$ by definition.

For large n, $\frac{\delta^2}{S_e^2}$ may be taken as approximately normally distributed with

$$\text{mean} = \frac{2N}{N-1}$$

and $\text{Variance} = \frac{4N^2(N-2)}{(N+1)(N-1)^2}$ or $\sigma = \sqrt{\frac{4N^2(N-2)}{(N+1)(N-1)^3}}$

Now, after computation of the Ratio (δ^2/S^2), mean and σ, if $\frac{\delta^2}{S^2}$ lies between '$\text{mean} \pm 2\sigma$', we accept the hypothesis of presence of autocorrelation otherwise we reject the hypothesis of its presence.

Numerical Example:

Given $\Sigma(e_t - e_{t-1})^2 = 49.56$, $\Sigma(e_t - \bar{e})^2 = 18.2$, $N = 112$.

Estimate Von-Neumann Ratio and examine the existence of autocorrelation.

Solution: Since, Von-Neumann Ratio $\left(\frac{\delta^2}{S_e^2}\right) = \frac{\Sigma(e_t - e_{t-1})^2 / N-1}{\Sigma(e_t - \bar{e})^2 / N}$

$$\therefore \frac{\delta^2}{S_e^2} = \frac{49.56/112-1}{18.8/112} = \frac{0.446486}{0.167857} = 2.66$$

$$\therefore \text{Von-Neumann Ratio}\left(\frac{\delta^2}{S_e^2}\right) = 2.66$$

$$\text{Mean} = \frac{2N}{N-1} = \frac{2 \times 112}{112-1} = \frac{224}{111} = 2.0180$$

$$\text{Variance} = \frac{4N^2(N-2)}{(N+1)(N-1)^3} = \frac{4 \times (112)^2(110)}{(112+1)\ (112-1)^3} = \frac{5519360}{154542303}$$

$$= 0.0357 = 0.$$

$$\therefore \sigma = \sqrt{\text{Variance}} = \sqrt{0.0357} = 0.189$$

Here, $\text{Mean} \pm 2\sigma = 2.018 \pm 2(0.189)$

$$= 2.018 \pm 0.378$$

$\therefore$ The Acceptance Region will be between

$2.018 - 0.378$ and $2.018 + 0.378$

or 1.64 to 2.396

Since, Von-Neumann Ratio (2.66) does not lies within the mean $\pm 2\sigma$ limit (i.e. from 1.64 to 2.396), we reject the null hypothesis and accept the alternative hypothesis of presence of autocorrelation.

ⓐ Short Answer type Questions :

1. Define Autocorrelation.
2. Discuss the main causes of Autocorrelation.

ⓑ Long Answer type Questions

1. What do you mean by Autocorrelation? Discuss the causes of autocorrelation.
2. Describe Durbin-Watson 'd' statistic method of detecting presence of autocorrelation.
3. Explain the methods (any-two) of detecting the presence of autocorrelation.

Suggested Readings:

1. Introduction to Econometrics – G.M.K. Madnani
2. Econometrics: Theory and Applications – Shyamala, Kaur and Pragasam
3. Econometrics & Mathematical Economics - Singh, Parasar & Singh
4. Econometrics : K. Dhanasekaran
5. Basic Econometrics: D.N. Gujarati

Prepared by:

Dr. Benoy Kumar Lal
Associate Professor
Department of Economics
Patna University, Patna

Mobile: 9608570960, 9852059001
Email : benoykumarlal@yahoo.in
