0.1092 seems low. This might seem to be a disappointing value, but as we shall show in the next chapter, such low R^2 values are frequently encountered in cross-sectional data with a large number of observations. Besides, even an apparently low R^2 value can be statistically significant (i.e., different from zero), as we will show in the next chapter.

Source: World Bank World Development Indicators, adjusted to 2000 base and estimated and projected values developed by the Economic Research Service.

*7.11 Partial Correlation Coefficients

Explanation of Simple and Partial Correlation Coefficients

In Chapter 3 we introduced the coefficient of correlation r as a measure of the degree of linear association between two variables. For the three-variable regression model we can compute three correlation coefficients: r_{12} (correlation between Y and X_2), r_{13} (correlation coefficient between Y and X_3), and r_{23} (correlation coefficient between X_2 and X_3); notice that we are letting the subscript 1 represent Y for notational convenience. These correlation coefficients are called **gross** or **simple correlation coefficients,** or **correlation coefficients of zero order.** These coefficients can be computed by the definition of correlation coefficient given in Eq. (3.5.13).

But now consider this question: Does, say, r_{12} in fact measure the "true" degree of (linear) association between Y and X_2 when a third variable X_3 may be associated with both of them? This question is analogous to the following question: Suppose the true regression model is (7.1.1) but we omit from the model the variable X_3 and simply regress Y on X_2, obtaining the slope coefficient of, say, b_{12}. Will this coefficient be equal to the true coefficient β_2 if the model (7.1.1) were estimated to begin with? The answer should be apparent from our discussion in Section 7.7. In general, r_{12} is not likely to reflect the true degree of association between Y and X_2 in the presence of X_3. As a matter of fact, it is likely to give a false impression of the nature of association between Y and X_2, as will be shown shortly. Therefore, what we need is a correlation coefficient that is independent of the influence, if any, of X_3 on X_2 and Y. Such a correlation coefficient can be obtained and is known appropriately as the **partial correlation coefficient.** Conceptually, it is similar to the partial regression coefficient. We define

$r_{12.3}$ = partial correlation coefficient between Y and X_2, holding X_3 constant
$r_{13.2}$ = partial correlation coefficient between Y and X_3, holding X_2 constant
$r_{23.1}$ = partial correlation coefficient between X_2 and X_3, holding Y constant

These partial correlations can be easily obtained from the simple or zero-order, correlation coefficients as follows (for proofs, see the exercises):[19]

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \tag{7.11.1}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{23}^2\right)}} \tag{7.11.2}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{13}^2\right)}} \tag{7.11.3}$$

*Optional.

[19] Most computer programs for multiple regression analysis routinely compute the simple correlation coefficients; hence the partial correlation coefficients can be readily computed.

The partial correlations given in Eqs. (7.11.1) to (7.11.3) are called **first-order correlation coefficients.** By *order* we mean the number of secondary subscripts. Thus $r_{12.34}$ would be the correlation coefficient of order two, $r_{12.345}$ would be the correlation coefficient of order three, and so on. As noted previously, r_{12}, r_{13}, and so on are called *simple* or *zero-order correlations*. The interpretation of, say, $r_{12.34}$ is that it gives the coefficient of correlation between Y and X_2, holding X_3 and X_4 constant.

Interpretation of Simple and Partial Correlation Coefficients

In the two-variable case, the simple r had a straightforward meaning: It measured the degree of (linear) association (and not causation) between the dependent variable Y and the single explanatory variable X. But once we go beyond the two-variable case, we need to pay careful attention to the interpretation of the simple correlation coefficient. From Eq. (7.11.1), for example, we observe the following:

1. Even if $r_{12} = 0$, $r_{12.3}$ will not be zero unless r_{13} or r_{23} or both are zero.

2. If $r_{12} = 0$ and r_{13} and r_{23} are nonzero and are of the same sign, $r_{12.3}$ will be negative, whereas if they are of the opposite signs, it will be positive. An example will make this point clear. Let Y = crop yield, X_2 = rainfall, and X_3 = temperature. Assume $r_{12} = 0$, that is, no association between crop yield and rainfall. Assume further that r_{13} is positive and r_{23} is negative. Then, as Eq. (7.11.1) shows, $r_{12.3}$ will be positive; that is, holding temperature constant, there is a positive association between yield and rainfall. This seemingly paradoxical result, however, is not surprising. Since temperature X_3 affects both yield Y and rainfall X_2, in order to find out the net relationship between crop yield and rainfall, we need to remove the influence of the "nuisance" variable temperature. This example shows how one might be misled by the simple coefficient of correlation.

3. The terms $r_{12.3}$ and r_{12} (and similar comparisons) need not have the same sign.

4. In the two-variable case we have seen that r^2 lies between 0 and 1. The same property holds true of the squared partial correlation coefficients. Using this fact, the reader should verify that one can obtain the following expression from Eq. (7.11.1):

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \tag{7.11.4}$$

which gives the interrelationships among the three zero-order correlation coefficients. Similar expressions can be derived from Eqs. (7.11.2) and (7.11.3).

5. Suppose that $r_{13} = r_{23} = 0$. Does this mean that r_{12} is also zero? The answer is obvious from Eq. (7.11.4). The fact that Y and X_3 and X_2 and X_3 are uncorrelated does not mean that Y and X_2 are uncorrelated.

In passing, note that the expression $r_{12.3}^2$ may be called the **coefficient of partial determination** and may be interpreted as the proportion of the variation in Y not explained by the variable X_3 that has been explained by the inclusion of X_2 into the model (see Exercise 7.5). Conceptually it is similar to R^2.

Before moving on, note the following relationships between R^2, simple correlation coefficients, and partial correlation coefficients:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \tag{7.11.5}$$

$$R^2 = r_{12}^2 + (1 - r_{12}^2)\, r_{13.2}^2 \tag{7.11.6}$$

$$R^2 = r_{13}^2 + (1 - r_{13}^2)\, r_{12.3}^2 \tag{7.11.7}$$

In concluding this section, consider the following: It was stated previously that R^2 will not decrease if an additional explanatory variable is introduced into the model, which can be seen clearly from Eq. (7.11.6).

This equation states that the proportion of the variation in Y explained by X_2 and X_3 jointly is the sum of two parts: the part explained by X_2 alone ($= r_{12}^2$) and the part not explained by X_2 ($= 1 - r_{12}^2$) times the proportion that is explained by X_3 after holding the influence of X_2 constant. Now $R^2 > r_{12}^2$ so long as $r_{13.2}^2 > 0$. At worst, $r_{13.2}^2$ will be zero, in which case $R^2 = r_{12}^2$.

Summary and Conclusions

1. This chapter introduced the simplest possible multiple linear regression model, namely, the three-variable regression model. It is understood that the term *linear* refers to linearity in the parameters and not necessarily in the variables.
2. Although a three-variable regression model is in many ways an extension of the two-variable model, there are some new concepts involved, such as *partial regression coefficients, partial correlation coefficients, multiple correlation coefficient, adjusted and unadjusted (for degrees of freedom)* R^2, *multicollinearity*, and *specification bias*.
3. This chapter also considered the functional form of the multiple regression model, such as the *Cobb-Douglas production function* and the *polynomial regression model*.
4. Although R^2 and adjusted R^2 are overall measures of how the chosen model fits a given set of data, their importance should not be overplayed. What is critical is the underlying theoretical expectations about the model in terms of a priori signs of the coefficients of the variables entering the model and, as it is shown in the following chapter, their statistical significance.
5. The results presented in this chapter can be easily generalized to a multiple linear regression model involving any number of regressors. But the algebra becomes very tedious. This tedium can be avoided by resorting to matrix algebra. For the interested reader, the extension to the k-variable regression model using matrix algebra is presented in **Appendix C,** which is optional. But the general reader can read the remainder of the text without knowing much of matrix algebra.

Multiple Choice Questions

1. The simplest possible multiple regression model is a
   a. One variable model
   b. Two variable model
   c. Three variable model
   d. Multi-variable model
2. Multiple linear regression models
   a. are linear in parameter and linear in variables
   b. are linear in parameter and may not be linear in variables
   c. may not be linear in parameter but are linear in variables
   d. may not be linear in parameter and variables
3. $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ where $X_{1i} = 1$ for all i. This is an example of
   a. Three variable model
   b. X variable model
   c. Four variable model
   d. Three beta model

Dummy Variable

A. Defination:

Dummy variables are the qualitative variables. In regression analysis it frequently happens that the dependant variable is influenced by not only by variables which can be readily quantified on some well-defined scale (e.g., income, price, height etc.) but also by variables which are essentially qualitative in nature (e.g., sex, race, wars, religion etc.). Since qualitative variables usually indicate the presence or absence of "quality" or an attribute, such as male or female, black or white etc, One method of "qualifying" such attributes is by constructing artificial variables which take on values of '1' or '0'. '0' indicating the absence of an attribute and '1' indicate the presence of that attribute.

Since we can assign a value of 1 to the presence, 0 to the absence of the attribute in question, we may view it as a variable that is restricted to two values. such a variable is then called a "binary" or a "dummy" variable.

B. Some example where the dummy variable uses!

(a) Where there are two characteristics: (one dummy variable)

Let the equation is $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$,
X_i is dummy variable

$$X_i = \begin{cases} 1, & \text{if the output is obtained from machine A} \\ 0, & \text{if the output is obtained from machine B.} \end{cases}$$

Here two characteristics are machine A and machine B. We shall check that explanatory variables are not related to ε_i. ε_i satisfies all the basic assumptions.

Now we can estimate β_1 and β_2 using OLS method. What is the significance of β_1 and β_2?

$$E(Y_i / X_i = 0) = \beta_1$$

i.e, Expectation of Y_i given that $X_i = 0$

and $E(Y_i / X_i = 1) = \beta_1 + \beta_2$

β_1 is the measure of the output obtained from machine B.

$\beta_1 + \beta_2$ is the measure of output obtained from machine A.

$\beta_1 + \beta_2 - \beta_1 = \beta_2$ is the expected output obtained from machine A is statistically significance from machine B.

16 Friday April, '99

(b) case of three characteristics and two dummy variables:

Suppose the starting salaries of the high school teachers of english are normally distributed with variance σ^2, the mean depending on whether the highest degree attained by the candidate is a B.A, an M.A or a Ph.D. Let the mean starting salary for a B.A is equal to μ_A, that for an M.A. is μ_B and that for a Ph.D, μ_C. The appropriate regression equ^n can be represented by —

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

where Y_i is the salary of the ith candidate and

$X_{i2} = 1$, if the highest degree of the candidate is a Ph.D.

$= 0$, otherwise.

$X_{i3} = 1$, if the highest degree is an M.A.

$= 0$, otherwise.

When $X_{i2} = 1$, X_{i3} must be equal to zero and vice versa. X_{i2} and X_{i3} are not correlated with ε_i and ε_i satisfies all the basic assumption. So we can apply the OLS method to obtained β_1, β_2, and β_3.

The mean values of Y_i corresponding to different values of the regressors are

$E(Y_i/X_{i2}=0, X_{i3}=0) = \beta_1 = \mu_A$.

$E(Y_i/X_{i2}=0, X_{i3}=1) = \beta_1 + \beta_3 = \mu_B$.

$E(Y_i/X_{i2}=1, X_{i3}=0) = \beta_1 + \beta_2 = \mu_C$.

It follows that.

$\beta_1 = \mu_A$.

$\beta_2 = \mu_C - \mu_A$.

$\beta_3 = \mu_B - \mu_A$.

It means that, the intercept of the regression equation i.e., β_1 measure the mean salary of a B.A. The coefficient of X_{i2}, i.e. β_2, measures the difference between mean salary of a Ph.D and a B.A. The coefficient of X_{i3}, i.e., β_3, measures the difference between the mean salary of an M.A and a B.A. Again $\mu_B - \mu_C = \beta_3 - \beta_2$ measures the difference between the mean salary of an M.A. and a Ph.D.

Test of Hypothesis :

Now we shall see that there have any statistically significant difference between μ_C and μ_A or μ_B and μ_A or μ_B and μ_C

So we shall test the hypothesis that,

i) $H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$, to see the difference between μ_C and μ_A i.e., the difference between the mean value of a Ph.D $(\beta_1 + \beta_2)$ and a B.A. (β_1).

ii) $H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$, to see the difference between μ_B and μ_A i.e., $\beta_1 + \beta_3$ and β_1.

iii) $H_0 : \beta_2 = \beta_3$

$H_1 : \beta_2 \neq \beta_3$, to see the difference between the mean salary of an M.A (μ_B) and the mean salary of a Ph.D.

ⓒ Case of three characteristics and one dummy variable

Now if we use one variable with three values say 0 for a B.A, 1 for an M.A and 2 for a Ph.D. If we formed the regression model as

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

Where D_i is the dummy or explanatory variable with values 0, 1, and 2.

$D_i = 0$, when the person is a B.A.

$= 1$, when the person is an M.A.

$= 2$, when the person is a Ph.D.

Now, $E(Y_i/D_i=0) = \alpha = \mu_A$

$E(Y_i/D_i=1) = \alpha+\beta = \mu_B$

$E(Y_i/D_i=2) = \alpha+2\beta = \mu_C$

However, this implies that the difference between the mean salary of an M.A and a B.A is $\mu_B - \mu_A = \alpha+\beta-\alpha = \beta$, and the difference between the mean salary of a Ph.D and an M.A is

$$\mu_C - \mu_B = \alpha+2\beta-\alpha-\beta = \beta.$$

That is, by using one variable with values 0, 1 and 2 we are in fact assuming that the difference between the salary of a Ph.D and an M.A. is the same as that between the salary of an M.A. and a B.A. So in this case we are not justified in making such an assumption.

Evening

Memo

① Case of three characteristics and three dummy variables.

In the above example if we formed the regression model as

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

where $X_{i2} = 1$, if the highest degree is a Ph.D
$= 0$, otherwise
$X_{i3} = 1$, if the highest degree is an M.A.
$= 0$, otherwise.
$X_{i4} = 1$, if the highest degree is a B.A.
$= 0$, otherwise.

Y_i	X_{i2}	X_{i3}	X_{i4}
Ph.D.	1	0	0
Ph.D.	1	0	0
M.A.	0	1	0
M.A.	0	1	0
B.A.	0	0	1

The solution for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\hat{\beta}_4$ would be indeterminate. The reason is that

$$X_{i4} = 1 - X_{i2} - X_{i3}.$$

which implies that X_{i4} is the linear function of another two explanatory variables i.e., there is multicollinearity between the explanatory variables. Therefore $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\hat{\beta}_4$ can not be determined.

Now, we can conclude that the number of dummy variable will one short of the number of characteristics.

Auto-Correlation

Defination:

The term autocorrelation may be defined as correlation between members of series of observations ordered in time [as in time-series data] or space [as in cross-sectional data].

One of the basic assumptions in the regression model is that the value of the disturbance term in one period is independent of its value in any other period, so that

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

If this assumption is not satisfied, that is, if the value of u in any particular period is correlated with its own preceding value (or values) we say that there is autocorrelation or serial correlation of the random variable.

Symbolically $E(u_i u_j) \neq 0$ where $i \neq j$.

B. Auto-correlation in the disturbance term in such that first order auto-regressive scheme

or

Derive the auto-correlation co-efficient formula.

In the regression model where all the basic assumptions hold, each disturbance represents an independent drawing from a normal population with mean zero and variance σ_u^2. When the disturbances are first order auto-regressive, the drawings are no longer independent but are generated according to the following scheme:

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

i.e., ε_t is influenced by ε_{t-1}.

The regression equation is,

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

where $E(\varepsilon_t \varepsilon_s) \neq 0$, $t \neq S$. $(t > S)$

and $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$; $|\rho| < 1$

ρ is the auto-correlation co-efficient.

Here, $E(u_t \cdot \varepsilon_{t-1}) = 0$, for all t,

u_t is not auto-correlated. ε_t is auto-correlated.

$E(u_t \cdot u_s) = 0$ (for all $t \neq S$)

Now,
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$
$$= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t$$
$$= \rho^2\varepsilon_{t-2} + \rho u_{t-1} + u_t$$
$$= \rho^2(\rho\varepsilon_{t-3} + u_{t-2}) + \rho u_{t-1} + u_t$$
$$= \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t$$
$$= \ldots$$

$= \ldots\ldots$

$= \rho^s \varepsilon_{t-s} + \rho^{s-1} u_{t-s+1} + \rho^{s-2} u_{t-s+2} + \ldots\ldots$

$\ldots + \rho^2 u_{t-2} + \rho u_{t-1} + u_t$

since, $\rho^s \to 0$ as $s \to \infty$, we can write,

$$\varepsilon_t = \sum_{s=0}^{\infty} \rho^s u_{t-s} \quad , \text{ since } |\rho| < 1$$

This expression is called the moving average representation of ε_t. This expression is convenient for working out the variance and cov. of ε's.

Variance :-

Now, $var(\varepsilon_t) = V(u_t) + V(\rho u_{t-1}) + V(\rho^2 u_{t-2}) + \ldots\ldots$

$= \sigma_u^2 + \rho^2 \sigma_u^2 + \rho^4 \sigma_u^2 + \ldots$

$= \sigma_u^2 (1 + \rho^2 + \rho^4 + \ldots\ldots)$

$= \dfrac{\sigma_u^2}{1-\rho^2}$

$$\therefore var(\varepsilon_t) = \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}$$

or, $\sigma_u^2 = (1-\rho^2)\sigma^2$

Covariance :-

Now, $cov(\varepsilon_t, \varepsilon_{t-1}) = E(\varepsilon_t \varepsilon_{t-1})$

$= E[(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \ldots)(u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \ldots)]$

$= \rho \sigma_u^2 + \rho^3 \sigma_u^2 + \rho^5 \sigma_u^2 + \ldots\ldots$

$= \rho \sigma_u^2 (1 + \rho^2 + \rho^4 + \ldots\ldots)$

$= \dfrac{\rho \sigma_u^2}{1-\rho^2}$

$= \rho \sigma^2 \quad \left[\because \dfrac{\sigma_u^2}{1-\rho^2} = \sigma^2\right]$

$$\therefore Cov(\varepsilon_t \varepsilon_{t-1}) = \rho \sigma^2$$

Thus we get, $cov(\varepsilon_t \varepsilon_{t-1}) = \rho \sigma^2$

Similarly, $cov(\varepsilon_t \varepsilon_{t-2}) = \rho^2 \sigma^2$

$cov(\varepsilon_t \varepsilon_{t-3}) = \rho^3 \sigma^2$

$$\vdots$$

$$cov(\varepsilon_t \varepsilon_{t-s}) = \rho^s \sigma^2$$

Now we have seen earlier

$$cov(\varepsilon_t \varepsilon_{t-1}) = \rho \sigma^2$$

$$\therefore \rho = \frac{cov(\varepsilon_t \varepsilon_{t-1})}{\sigma^2}$$

$$\therefore \boxed{\rho = \frac{cov(\varepsilon_t \varepsilon_{t-1})}{\sqrt{Var(\varepsilon_t)} \cdot \sqrt{Var(\varepsilon_{t-1})}}}$$

This is the auto-correlation co-efficient formula.

This co-efficient measures the degree of relationship between the two random variables and its values range from −1 to +1. Positive values reflects the existence of a positive relationship and negative values presence of a negative relationship. Values close to +1 or −1 indicates a high degree of relationship and the coefficient whose values are zero, then

$$\varepsilon_t = u_t \quad or, \quad var(\varepsilon_t) = \sigma_u^2$$

Q. Determine whether the OLS estimators of the parameters (α, β) are unbiased or efficient or BLUES when the disturbance term is auto-regressive or auto-correlated.

Now examine the properties of the least squares estimators of α and β in $Y_t = \alpha + \beta X_t + \varepsilon_t$, $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$ when the disturbance term ε_t is autoregressive.

we know, $\hat{\beta} = \frac{\Sigma x_t' y}{\Sigma x_t'^2}$ where $x_t' = X_t - \bar{X}$

$$= \beta + \frac{\Sigma x_t' \varepsilon_t}{\Sigma x_t'^2}$$

$$\therefore E(\hat{\beta}) = \beta + \frac{\Sigma x_t'}{\Sigma x_t'^2} E(\varepsilon_t)$$

$$= \beta \qquad \left[\because E(\varepsilon_t) = 0\right]$$

$$E(\varepsilon_t) = \sum_{s=0}^{\infty} \rho^s E(u_{t-s}) = 0$$

The least square estimator of α is

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = (\alpha + \beta\bar{X} + \bar{\varepsilon}) - \hat{\beta}\bar{X}$$

$$E(\hat{\alpha}) = \alpha + \beta\bar{X} + E(\bar{\varepsilon}) - E(\hat{\beta})\bar{X}$$

$$= \alpha + \beta\bar{X} - \beta\bar{X}$$

$$= \alpha$$

This means that least square estimators are unbiased even when the disturbance term is auto-regressive or auto-correlation.

Next we determine whether the lest squares estimators are BLUES. by deriving the BLUE formulas for the autoregressive case and by comparing them with the least squares formulas. If the two sets of formulas differ, then the least squares estimators are not BLUEs'.

Now the original equation is

$$Y_t = \alpha + \beta X_t + \varepsilon_t \text{, where } \varepsilon_t = \rho\varepsilon_{t-1} + u_t \longrightarrow ①$$

Now multiplying each side by ρ and give one period lag we get, $\rho Y_{t-1} = \alpha\rho + \beta\rho X_{t-1} + \rho\varepsilon_{t-1} \longrightarrow ②$

Equn. ① − Equn. ②, we get

$$Y_t - \rho Y_{t-1} = (1-\rho)\alpha + \beta(X_t - \rho X_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

$$\text{or, } Y_t^* = \alpha' W_t^* + \beta X_t^* + u_t, \quad t = 2, 3, \dots, n.$$

where, $Y_t^* = Y_t - \rho Y_{t-1}$, $W_t^* = 1-\rho$ and $X_t^* = X_t - \rho X_{t-1}$

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

In this way we replaced the autoregressive ε_t by the classical disturbance u_t. This transformation is known as "Cochrance Orcutt transformation"

In this process we lost one observation pertaining to u_1. To get it we consider $Y_1 = \alpha + \beta X_1 + \varepsilon_1$ and the effect of the pre-sample disturbances into account is to specify the first sample disturbance as

$$\varepsilon_1 = \frac{u_1}{\sqrt{1-\rho^2}}$$

$\left[\because \varepsilon_t = \sum_{s=0}^{\infty} \rho^s u_{t-s}\right]$

$$\text{Now, } Y_1 = \alpha + \beta X_1 + \frac{u_1}{\sqrt{1-\rho^2}}$$

$$\text{or, } Y_1\sqrt{1-\rho^2} = \alpha\sqrt{1-\rho^2} + \beta X_1\sqrt{1-\rho^2} + u_1$$

Now using this relation for $t=1$, we get the complete transformation of the original equation

$$Y_t^* = \alpha W_t^* + \beta X_t^* + u_t \quad , (t=1,2,3,\ldots,n) \longrightarrow (3)$$

where for $t=1$,

$$Y_t^* = Y_t\sqrt{1-\rho^2}, \quad W_t^* = \sqrt{1-\rho^2}, \quad X_t^* = X_t\sqrt{1-\rho^2}$$

and for $t=2,3,\ldots,n$

$$Y_t^* = Y_t - \rho Y_{t-1}, \quad W_t^* = 1-\rho, \quad X_t^* = X_t - \rho X_{t-1}$$

In the transformed equation there are two nonstochastic explanatory variables W^* and X^*. W^* has the same value for all observation except the first one. This transformation is known as the "Prais-Winston transformation".

Since u_t satisfies all the basic assumptions, all pre-conditions for the equality of the LSEs and BLUEs of α and β are met. By applying the least squares method, we obtain the following "least squares normal equations".

$$\Sigma W_t^* Y_t^* = \tilde{\alpha}\, \Sigma W_t^{*2} + \tilde{\beta}\, \Sigma W_t^* X_t^* \longrightarrow (4)$$

$$\Sigma X_t^* Y_t^* = \tilde{\alpha}\, \Sigma W_t^* X_t^* + \tilde{\beta}\, \Sigma X_t^{*2} \longrightarrow (5)$$

where $\sum_{t=1}^{n} W_t^* Y_t^* = W_1^* Y_1^* + \sum_{t=2}^{n} W_t^* Y_t^*$

$$= Y_1(1-\rho^2) + (1-\rho)\sum_{t=2}^{n}(Y_t - \rho Y_{t-1})$$

$$\sum_{t=1}^{n} W_t^{*2} = W_1^{*2} + \sum_{t=2}^{n} W_t^{*2}$$

$$= (1-\rho^2) + (n-1)(1-\rho)^2$$

and $\sum_{t=1}^{n} w_t^* x_t^* = w_1^* x_1^* + \sum_{t=2}^{n} x_t^* w_t^*$

$$= x_1(1-\rho^2) + (1-\rho)\sum_{t=2}^{n}(x_t - \rho x_{t-1})$$

$$\sum_{t=1}^{n} x_t^* y_t^* = x_1 y_1 (1-\rho^2) + \sum_{t=2}^{n}(y_t - \rho y_{t-1})(x_t - \rho x_{t-1})$$

$$\sum_{t=1}^{n} x_t^{*2} = x_1^{*2} + \sum_{t=2}^{n} x_t^{*2}$$

$$= x_1^2(1-\rho^2) + \sum_{t=2}^{n}(x_t - \rho x_{t-1})^2$$

There are two equations (4) and (5) and three unknown $\tilde{\alpha}$, $\tilde{\beta}$ and ρ. The formulas for $\tilde{\alpha}$ and $\tilde{\beta}$ are somewhat complicated, but they will clearly involve the parameter ρ.

Since the ordinary least squares estimators do not involve ρ, they are not BLUE when the disturbance term is autoregressive. Therefore the least squares estimators do not have the smallest variance among all unbiased estimators and therefore, are not efficient.

Thus we can conclude that the OLS estimates of the parameters are not BLUEs when there are autocorrelation in the disturbance term.

Test the Presence of Autocorrelation

Durbin-Watson Test

To test the presence of Auto-correlation, the most widely used test is the Durbin-Watson Test.

Steps involved in this test :- As we know that the OLS method in the presence of auto-correlation is inefficient, then how we measure the auto-correlation? However if we do not know - or are not willing to assume - that the regression disturbance is not auto-correlated, we may have to turn to the sample for information. In the context of first-order autoregression, we may want to test the hypothesis of no auto-regression.

$H_0 : \rho = 0$, against a one side or two sides alternative.

To apply the Durbin-Watson test, at first we calculate the value of a statistic called the Durbin-Watson statistic (d).

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

where, e_t represent the OLS residual i.e., $e_t = Y_t - \hat{Y}_t$.

After the calculation of the value of 'd', we shall observe that it is negative or positive auto-correlation.

Hypothesis: Hypothesis is that

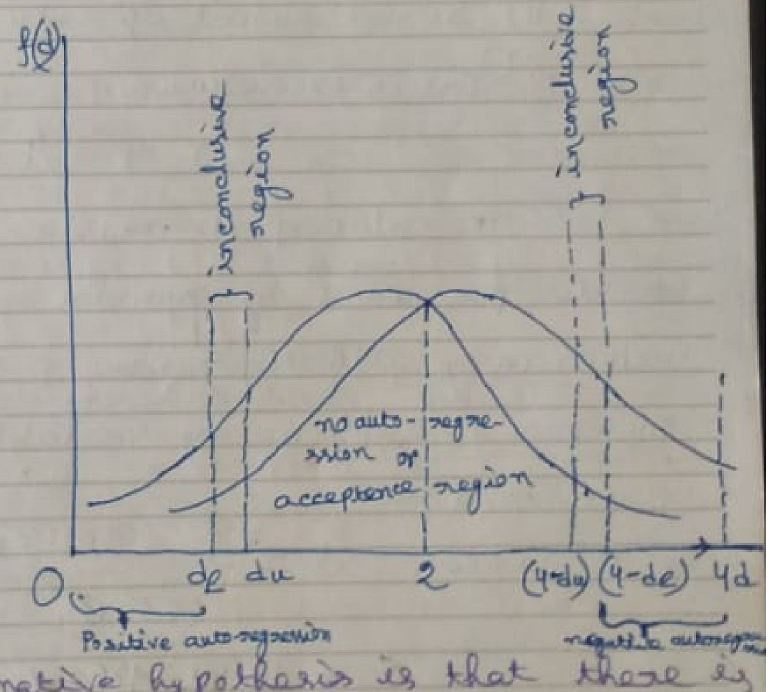
① $H_0 : \rho = 0$, $H_1 : \rho > 0$, ② $H_0 : \rho = 0$, $H_1 : \rho < 0$, ③ $H_0 : \rho = 0$, $H_1 : \rho \neq 0$ (two tailed hypothesis)

Decision rules about the acceptence or rejection of the null hypothesis:

The value of 'd' falling in between du and 4-du means that null hypothesis is accepted.

$d_L \Rightarrow$ Lower limit

$d_u \Rightarrow$ upper limit.



① If the alternative hypothesis is that there is positive autocorrelation, then the decision rules are
i) reject H_0 if $d < d_L$
ii) do not reject H_0 if $d_u < d < (4-d_u)$
~~iii) the test is inconclusive if $d_L \leq d \leq d_u$~~

(2) If the alternative hypothesis is that there is negative auto-correlation, then the decision rules are

i) reject H_0 if $d > 4 - d_L$

ii) Do not reject H_0 if $d_u < d < 4 - d_u$

iii) test is inconclusive if $4 - d_u \leq d \leq 4 - d_L$

(3) If the alternative hypothesis is two-tailed one then the decision rules are

i) Reject H_0 if $d < d_L$ or if $d > 4 - d_L$

ii) Do not reject if $d_u < d < 4 - d_u$

iii) Test is inconclusive if $d_L \leq d \leq d_u$

or, if $4 - d_u \leq d \leq 4 - d_L$

The values of d_L and d_u are given in the table provided by Durbin and watson. These values vary with the number of observations and the number of explanatory variables in the regression equation.

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

$$= \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2 \sum e_t e_{t-1}}{\sum e_t^2}$$

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

If 't' is infinitely large then there have no any differences between e_t and e_{t-1}

$$\therefore \text{plim } d = 1 + 1 - 2\hat{\rho} \qquad \left[\because \hat{\rho} = \frac{\Sigma e_t e_{t-1}}{\Sigma e_t^2}\right]$$
$$= 2(1 - \hat{\rho})$$

Therefore, values of 'd' close to 2 will lead to the acceptance of the null hypothesis, where as those close to zero or close to 4 will leads to its rejection.

Limitation of this test: The D-W test is not applicable when i) in the regression equation in which the place of the explanatory variables is taken by the lagged value of the dependent variable, and

ii) if there is no constant term in the equation.

iii) if the sample size is very small.

Modification of the D-W Test: (by sir P. S. Das)