## Econometrics:

Econometrics deals with the application of statistical tools to economic data. The first task an econometrician faces is that of formulating an economic relationship, which is necessarily a simplified model of the real-world process.

Estimation and testing of these models with observed data, and the use of the estimated models for prediction and policy analysis are the other two major goals of econometrics.

## Definition:

Econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observations, related by appropriate method of inference.

OR

Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena.

## Methodology of Econometrics:

Econometric methodology proceeds along the following lines;

1) Statement of theory or hypothesis.
2) Specification of the mathematical model of the theory.
3) Specification of the statistical, or econometric, model.
4) Obtaining the data.
5) Estimation of the parameters of the econometric model.

6) Hypothesis testing.
7) Forecasting or prediction.
8) Using the model for control or policy purpose.

To illustrate the preceding steps, let us consider the Keynesian theory of consumption.

1) Statement of theory or hypothesis

Keynes stated that the fundamental psychological law is that men [women] are disposed, as a rule and on average, to increase their consumption as their income increases, but not as much as the increase in their income.

In short, Keynes postulated that the marginal propensity to consume (MPC), the rate of change of consumption for a unit (say, a dollar) change in income, is greater than zero but less than 1.

2) Specification of mathematical model of consumption.

Although Keynes postulated a positive relationship between consumption and income, he did not specify the precise form of functional relationship between the two. For simplicity, a mathematical economist might suggest the following form of the Keynesian consumption function:

$$Y = \beta_1 + \beta_2 X \quad , 0 < \beta_2 < 1 \quad \cdots (1.1)$$

where, $Y$ = consumption expenditure,
$X$ = Income
$\beta_1$ and $\beta_2$ known as parameters of model are respectively, the intercept and slope coefficients.

The slope coefficient $B_2$ measures the mpc. Geometrically eqn (1.1) is as shown in fig. (1.1)

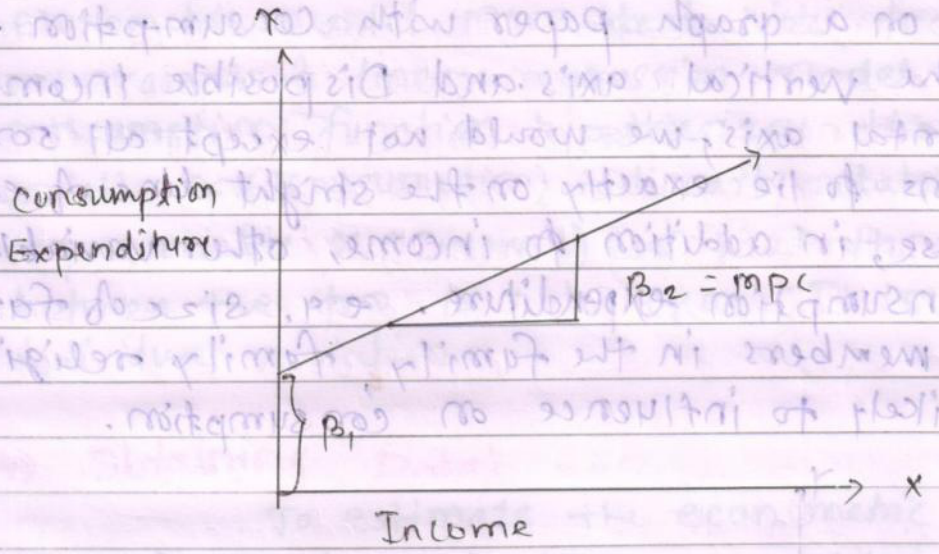



fig (1.1) Keynesian consumption function.

Eqn (1.1) state that consumption is linearly related to income, is an example of a mathematical model of the relationship between consumption and income. that is called the consumption function in economics. The model is simply a set of mathematical equation.

If the model has only one equation, as in the preceding example, it is called a single equation model, where as if it has more than one equation, it is known as multiple equation model.

### 3) Specification of the econometric model of consumption:

The purely mathematical model of the consumption function given in eqn (1.1) is of limited intrest to the econometrician, for it assumes that there is an exact or deterministic relationship between consumption and income. But relationship between economic variables are generally inexact.

Thus, If we were to obtain data on consumption expenditure and disposable income of a sample of, say, 500 American families and plot these data on a graph paper with consumption expenditure on the vertical axis and Disposible Income on horizontal axis, we would not except all 500 observations to lie exactly on the stright line of eqn (1.1) because, in addition to income, other variables affect consumption expenditure. e.g. size of family, ages members in the family, family religion etc, are likely to influence on consumption.

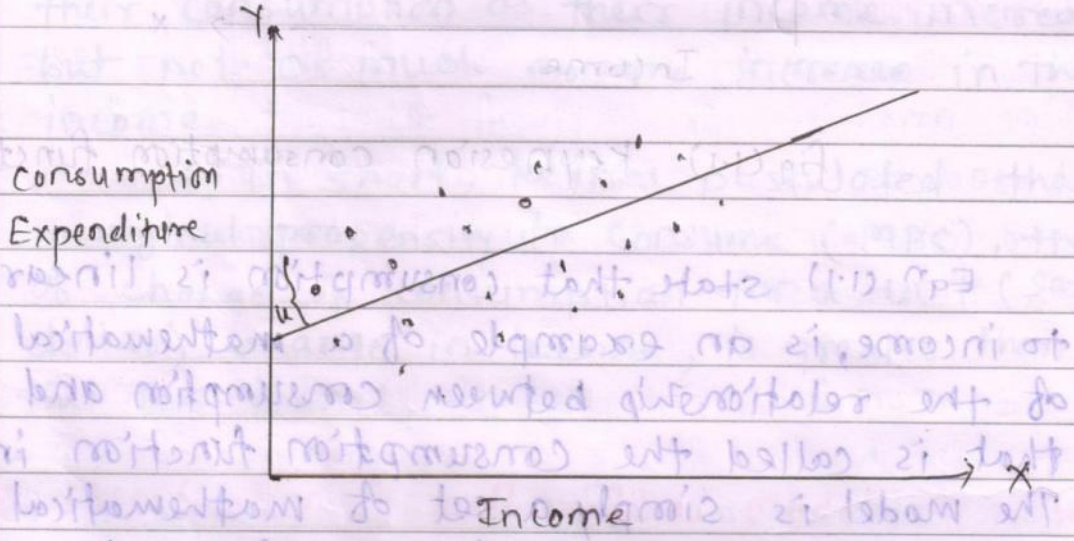



fig (1.2): Econometric model of the keynesian consumption function.

To allow for the inexact relationship between economic variables, the econometrician would modify the deterministic consumption function (1.1) as follows,

$$Y = \beta_1 + \beta_2 X + u \quad - - - - \quad (1.2)$$

where, u known as the disturbance, or error, is a random (stochastic) variable, That has well defined probabilistic properties. The disturbance term 'u' may well represent all those

factors that affect consumption but are not taken into account explicitly.

Eqn (1.2) is an example of an econometric model. more ~~technique~~ technically, it is an example of linear regression model. The econometric consumption function hypothesizes that the dependent variable Y (consumption) is linearly related to the explanatory variable X (Income) but that the relationship between the two is not exact; It is subject to individual variation.

## 4) Obtaining Data:

To estimate the econometric model given in equation (1.2), that is, to obtain the numerical values of $B_1$ and $B_2$, we need data. Let us look at the data given in table (1.1) which related to the U.S. economy for the period 1981 - 1996.

Table 1.1

Data on Y (Personal consumption expenditure) and X (Gross Domestic Product, 1982-1996). Both in 1992 Billions of Dollars.

| Year | Y | X |
|---|---|---|
| 1982 | 3081.5 | 4620.3 |
| 1983 | 3240.6 | 4803.7 |
| 1984 | 3407.6 | 5140.1 |
| 1985 | 3566.5 | 5323.5 |
| 1986 | 3708.7 | 5487.7 |
| 1987 | 3822.3 | 5649.5 |
| 1988 | 3972.7 | 5865.2 |
| 1989 | 4064.6 | 6062.0 |
| 1990 | 4132.2 | 6136.3 |
| 1991 | 4105.8 | 6079.4 |
| 1992 | 4219.8 | 6244.4 |
| 1993 | 4343.7 | 6389.6 |
| 1994 | 4486.0 | 6610.7 |
| 1995 | 4595.3 | 6742.1 |

The Y variable in this table is the aggregate (for the economy as a whole) personal consumption expenditure [PCE] and the variable X is Gross domestic product (GDP), a measure of aggregate income, both measured in billions of 1992 dollars. Therefore, the data are in "real" terms; that is, they are measured in constant (1992) prices. The data are plotted in the fig (1.3).

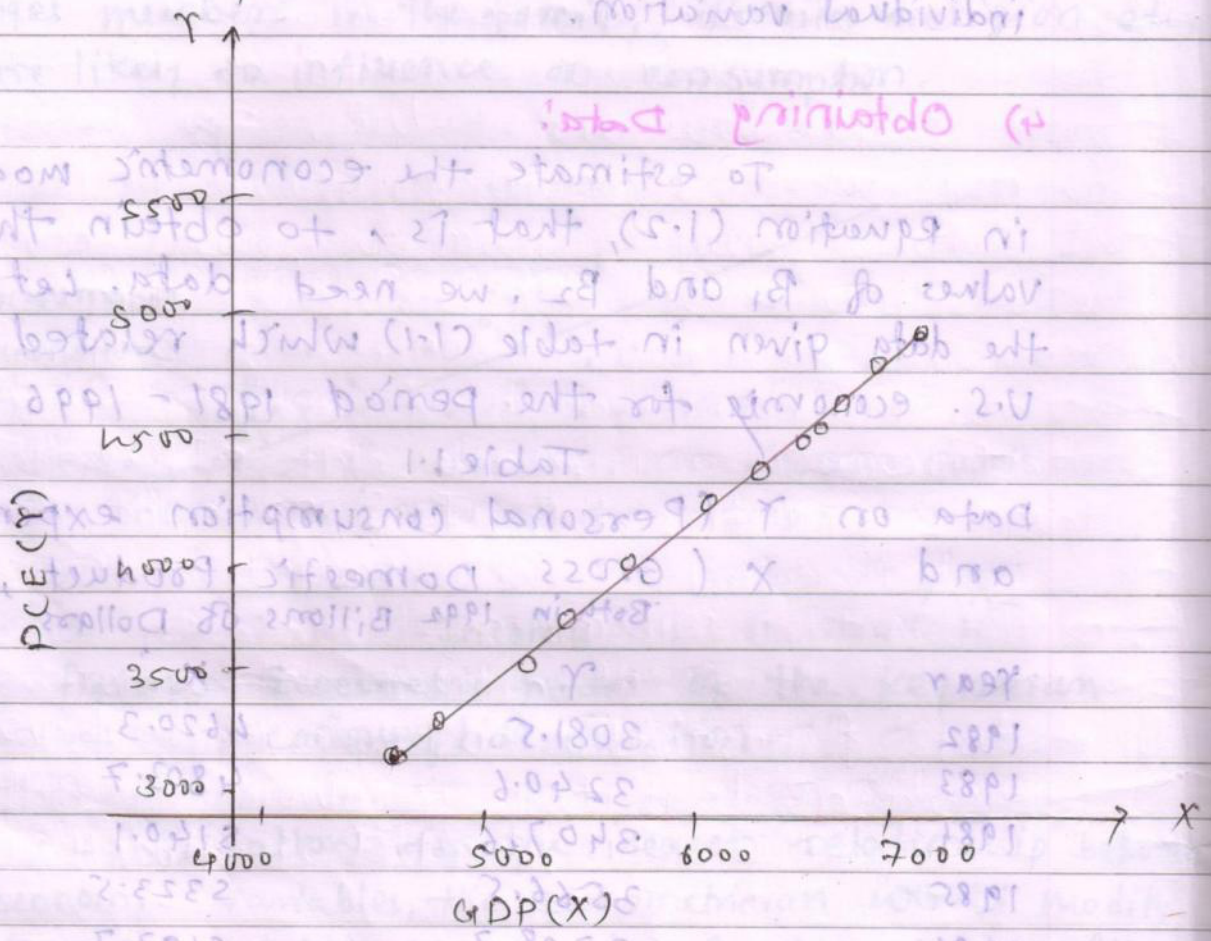



fig (1.3) Personal Consumption expenditure PCE (Y) in relation to GDP (X).

5) Estimation of Econometric Model:

Now that we have the data, our next task is to estimate the parameters of the consumption function. The numerical estimates

of the parameters give empirical content to the consumption function. For now, note that the statistical techniques of regression analysis is the main tool used to obtain the estimates. Using this techniques and the data given in the table 1.1, we obtain the following estimates $B_1$ and $B_2$, namely, $-184.08$ and $0.7064$, thus the estimated consumption function is,

$$\hat{Y} = -184.08 + 0.7064X_i \quad \cdots \quad (1.3)$$

The hat of the Y indicate that it is an estimate. The estimated consumption function (i.e. regression line) is shown in fig (1.3)

As fig (1.3) shows, the regression line fits the data quite well in that the data points are very close to the regression line. From this figure we see that for the period 1982–1996 the slope coefficient (i.e. MPC) was about 0.70, suggesting that for the sample period an increase in real income of one dollar, on average, to an average of about 70 cents in real consumption expenditure. We say on average because the relationship between consumption and income is inexact; as is clear from figure (1.3); not all the data points lie exactly on the regression line. In simple terms, we can say that, according to our data, the average, or mean, consumption expenditure went up by about 70 cents for a dollars increase in real income.

## 6) Hypothesis Testing:

Assuming that the fitted model is reasonably good approximation of reality, we have

to develop suitable criteria to find out whether the estimates obtained in, say, eqn (1.3) are in accord with the expectation of the theory that is being tested. According to "positive" economists like Milton Friedman, a theory or hypothesis that is not verifiable by appeal to empirical evidence may not be admissible as a part of scientific enquiry.

As noted earlier, Keynes expected the mpc to be positive but less than one. In our example we found the mpc to be about 0.70. But before we accept this finding as confirmation of keynesian consumption theory, we must enquire whether this estimate is sufficiently below unity to convince us that this is not a chance occurance or peculiarity of the particular data we have used. In other words, is 0.70 statistically less than one? If it is, it may support Keynes' theory.

Such confirmation or refutation of economic theories on the basis of sample evidence is based on a branch of statistical theory known as Statistical Inference. (Hypothesis testing).

7) Forcasting or Prediction:

If the choosen model does not refute the hypothesis or theory under consideration, we may use it to predict the future value (s) of the dependent, or forecast, variable (Y) on the basis of known or expected future value (s) of the explanatory, or predictor, variable X.

To illustrate, suppose we want to predict the mean consumption expenditure for 1997. The GDP value for 1997 was 7269.8 billion dollars. Putting this GDP figure on right hand side of eqn (1.3), we obtain,

$$\hat{Y}_{1997} = -184.0779 + 0.7064\,(7269.8)$$
$$= 4951.3167 \cong 4951 \text{ billion dollars.}$$

Thus, given the value of GDP, the mean, or average, forecast consumption expenditure is about 4951 billion dollars. The actual value of the consumption expenditure reported in 1997 was 4913.5 billion dollars. Estimated model (1.3) thus over predicted the actual consumption expenditure by about 37.82 billion dollars. We could say the forecast error is about 37.82 billion dollars, which is about 0.76 % of actual GDP value for 1997.

Suppose that, as a result of the proposed policy change, investment expenditure increase. What will be the effect on the economy? As macroeconomic theory shows, the change in income following, say, a dollar's worth of change in investment expenditure is given by the income multiplier M, which is defined as

$$M = \frac{1}{1 - MPC} \qquad \ldots\ldots (1.4)$$

If we use the MPC of 0.70 obtained in (1.3), this multiplier becomes about M = 3.33. That is, an increase (decrease) of a dollar in investment will eventually lead to more than a threefold increase (decrease) in income; note that it takes time for the multiplier to work.

The critical value in this computation is MPC, for the multiplier depends on it. And this estimate of the MPC can be obtained from regression models such as (1.3). Thus, a quantitative estimate of MPC provides valuable information for policy purposes. Knowing MPC, one can predict the future course of income, consumption expenditure, and employment following a change in the government's fiscal policies.

## 8. Use of the Model for Control or Policy Purposes.

Suppose we have the estimated consumption function given in (1.3). Suppose further the government believes that consumer expenditure of about 4900 (billions of 1992 dollars) will keep the unemployment rate at its current level of about 4.2 percent (early 2000). What level of income will guarantee the target amount of consumption expenditure?

If the regression results in (1.3) seem reasonable, simple arithmetic will show that

$$4900 = -184.0779 + 0.7064X \qquad (1.5)$$

which gives $X = 7197$, approximately. That is, an income level of about 7197 (billion) dollars, given an MPC of about 0.70, will produce an expenditure of about 4900 billion dollars.

As these calculations suggest, an estimated model may be used for control, or policy, purposes. By appropriate fiscal and monetary policy mix, the government can manipulate the control variable X to produce the desired level of the target variable Y.

## * Regression Analysis :

Regression analysis is a statistical tool for investigating the relationship between a dependent variable and one or more independent variables.

Regression analysis is widely used for prediction & forecasting.

Applications of Regression Analysis:

1) Economics
2) Management
3) Life and Biological Science
4) Physical and chemical science.
5) Engineering
6) Social Science.

## Regression Vs Correlation

The primary objective of correlation analysis is to measure the strength or degree of linear association between two variables.

e.g. i) Correlation between smoking and lung cancer
ii) Correlation between scores on statistics and mathematics examinations, etc.

In Regression Analysis, we are not primarily interested in such a measure. Instead, we try to estimate or predict the average value of one variable on the basis of the fixed values of other variable e.g. We may want to know whether we can predict the average score on a statistics examination by knowing a student's score on mathematics examinations.

In Regression analysis there is an asymmetry in the way the dependent and explanatory variables are treated. The dependent variable is assumed to be statistical, random or stochastic that is, to have a probability distribution. The explanatory variables, on the other hand, are assumed to have fixed values (in repeated sampling).

## Terminology and Notation:

The terms dependent variable and explanatory variable are described variously. A representative list is:

| | |
|---|---|
| Dependent variable | Explanatory variable |
| ⇕ | ⇕ |
| Explained variable | Independent variable |
| ⇕ | ⇕ |
| Predictand | Predictor |
| ⇕ | ⇕ |

| | |
|---|---|
| Regressand | Regressor |
| ⇕ | ⇕ |
| Response | Stimulus |
| ⇕ | ⇕ |
| Endogenous | Exogenous |
| ⇕ | ⇕ |
| Outcome | Covariate |
| ⇕ | ⇕ |
| Controlled variable | control variable. |

If we are studying the dependance of a variable on only a single explanatory variable, such as that of consumption expenditure on real income, such a study is known as simple or two-variable, regression analysis.

However, If we are studying the dependence of one variable on more than one explanatory variable, as in the crop-yield, rainfall, temperature sunshine and fertilizer examples, it is known as multiple regression analysis. In other words, in two-variable regression there is only one explanatory variable, whereas in multiple regression there is more than one explanatory variables.

The term random is a synonym for the term stochastic. A random or stochastic variable is a variable that can take on any set of values, positive or negative, with a given probability.

The letter Y will denote the dependent variable and X's ($X_1, X_2, \cdots X_k$) will denote the explanatory variables, $X_k$ being the $K^{th}$ explanatory variable. The

subscript i or t will denote the $i^{th}$ or the $t^{th}$ observation or value. $X_{Ki}$ (or $X_{Kt}$) will denote the ith (or $t^{th}$) observation on variable $X_K$. N (or T) will denote the total number of observations or values in the population, and n (or t) the total number of observations in a sample.

As a matter of convention, the observation subscript i will be used for cross sectional data (i.e. data collection at one point of time) and the subscript t will be used for time series data (i.e. data collected over a period of time). The nature of cross-sectional and time series data, as well as the important topic of the nature and sources of data for empirical analysis, is discussed in the following section.

## * Types of Data:

Three types of data may be available for empirical analysis; (i) Time series, (ii) Cross-sectional and (iii) Pooled (i.e. combination of time series and cross section) data.

### (i) Time series data:

A time series is a set of observations on the values that a variable taken at different times. Such data may be collected at regular time intervals, such as daily (e.g. stock prices, weather reports), weekly (e.g. money supply figures), monthly [e.g. unemployment rate; the consumer Price Index (CPI)], quarterly (e.g. GDP)

annually (e.g. government budgets), quinquennially that is, every 5 years (e.g. the censuses of manufactures) or decennially (e.g. the census of population).

(ii) Cross-section Data: Cross section data are data on one or more variables collected at the same point in time, such as the censuses of population conducted by the Census Bureau every 10 years, the surveys of consumer expenditures conducted by the University of Michigan etc.

(iii) Pooled Data:

In Pooled, or combined, data are elements of both time series and cross-section data.

## Classical Linear Regression Model (CLRM):

Consider a linear regression model involving one dependent variable and $(K-1)$ independent variables say $X_2, X_3, X_4, \cdots, X_K$ as regressors,

The CLRM is defined as,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK} + u_i \quad \cdots (1.1)$$

$i = 1, 2, \cdots, n$

Where $Y_i$ is dependent variable and $X_i$, $i = 2, \cdots, K$ is independent variables,

Where $\beta_j$ are parameters of the model and known as regresson coefficients, $\beta_1$ is intercept and $\beta_j$, $j = 2, \cdots K$ are often called partial regression coefficients and $u_i$ is the disturbance in the $i$th regressor model.

The above eqn (1.1) be written as matrix form,

$$Y = X\beta + u$$

$$Y = [Y_1, Y_2, \cdots, Y_n]'_{n \times 1}$$

$$X = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1K} \\ 1 & X_{22} & X_{23} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots & & \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nK} \end{bmatrix}_{n \times K}$$

= Data Matrix (Regression matrix)

= $n \times K$ matrix given n observations on k-1 variables $X_2$ to $X_K$, the first column of 1's representing the intercept term.

$\beta$ = $K \times 1$ column vector of the unknown parameters $\beta_1, \beta_2, \cdots \beta_K$

$$= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

$u$ = $n \times 1$ column vector of n disturbances $u_i$

$$= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

* Deviation form of model:

From eqn (1.1), we can write

$$\bar{Y} = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + \cdots + \beta_K \bar{X}_K + \bar{u} \quad \cdots (1.2)$$

Therefore, substracting eqn (1.2) from eqn (1.1), we get

$$Y_i - \bar{Y} = \beta_2 (X_{i2} - \bar{X}_2) + \beta_3 (X_{i3} - \bar{X}_3) + \cdots + \beta_K (X_{iK} - \bar{X}_K) + (u_i - \bar{u})$$

$$\therefore y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + u_i$$

$$y = x\beta + u' \quad \cdots (1.3)$$

Where, $\underline{y} = (y_1, y_2, \cdots, y_n)'$ ; $n \times 1$

$\underline{\beta} = (\beta_2, \beta_3, \cdots, \beta_k)'$ ; $(k-1) \times 1$

$$x = \begin{bmatrix} x_{12} & x_{13} & \cdots & x_{1k} \\ x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times (k-1)}$$

## Assumption of the Classical Linear Regression Model:

model: $Y = X\beta + u$

Assumption 1:

The expected value of the disturbance vector $u$ is zero. i.e. $E(u) = 0$

$$E(u) = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Assumption 2:

$$E(uu') = \sigma^2 I$$

Proof:

$$E(uu') = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} [u_1, u_2 \cdots u_n]$$

Where $u'$ is a transpose of the column vector $u$.

$$E(uu') = E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \cdots & \cdots & \cdots & \cdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

Applying the expectations operator E to each element of the preceding matrix, we get

$$\therefore E(uu') = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_2) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} \quad \cdots (1.4)$$

Because of the assumptions of homoscedasticity and no serial correlation, above matrix reduces to

$$\therefore E(uu') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$= \sigma^2 I \quad \cdots (1.5)$$

where I is $n \times n$ identity matrix.

The matrix (1.4) is called the variance-covariance matrix of the disturbances $u_i$; the elements on the main diagonal of this matrix gives the

variances and the elements off the main diagonal give the covariances.

Assumption-3:

The matrix X is non-stochastic that is it consists of fixed numbers. Regression analysis is conditional regression analysis, conditional upon the fixed values of the X variables.

Assumption-4:

The X matrix has full column rank equal to k, the number of columns in the matrix. This means that the columns of the X matrix are linearly independent; that is there is no exact linear linear relationship among the X variables.

In other words there is no multicollinearity. In scalar notation this is equivalent to saying that there exists no set of numbers $\lambda_1, \lambda_2, \cdots, \lambda_k$ not all zero such that

$$\lambda_1 X_{i1} + \lambda_2 X_{i2} + \cdots + \lambda_k X_{ik} = 0 \quad \cdots (1.6)$$

where $X_{i1} = 1$ for all i (to allow for the column of 1's in the matrix X). In matrix notation eqn (1.6) can be represented as,

$$\lambda' x = 0 \quad \cdots (1.7)$$

where $\lambda'$ is a $1 \times k$ row vector and $x$ is a $k \times 1$ column vector.

In exact linear relationship such as eqn (1.6) exists, the variable are said to be collinear. If, on the other hand, eqn (1.6) holds true only if $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$, then the

X variables are said to linearly independent.

(*)

## ✓ Ordinary Least square (OLS) Estimation:

To obtain the OLS estimate of $\beta$, Let us first write k-variable sample regression

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \cdots + \hat{\beta}_K X_{iK} + \hat{u}_i \qquad i = 1, 2, \cdots n$$

which can be written more compactly in matrix notation as

$$Y = X\hat{\beta} + \hat{u} \qquad \cdots\cdots (1.8)$$

and in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1K} \\ 1 & X_{22} & X_{23} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nK} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

$$\underset{n\times 1}{Y} = \underset{n\times K}{X} \quad \underset{K\times 1}{\hat{\beta}} + \underset{n\times 1}{\hat{u}}$$

where $\hat{\beta}$ is a K-element column vector of the OLS estimators of the regression coefficients and where $\hat{u}$ is $n\times 1$ column vector of n residuals

The K-variable case the OLS estimators are obtained by minimizing

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_K X_{iK})^2$$

$\cdots (1.9)$

Where $\sum \hat{u}_i^2$ is the residual sum of squares (RSS). In matrix notation, this amounts to minimizing $\hat{u}'\hat{u}$ since

$$\hat{u}'\hat{u} = [\hat{u}_1 \ \hat{u}_2 \cdots \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2$$

$$= \sum_{i=1}^{n} \hat{u}_i^2$$

$$\cdots (1.10)$$

From (1.8), we obtain

$$\hat{u} = (Y - X\hat{\beta}) \qquad \cdots (1.11)$$

Therefore,

$$\hat{u}'\hat{u} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$\hat{u}'\hat{u} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \qquad \cdots (1.12)$$

Where use is made of properties of the transpose of matrix namely, $(X\hat{\beta})' = \hat{\beta}'X'$ and since $\hat{\beta}'X'Y$ is a scalar (a real number) it is equal to its transpose $Y'X\hat{\beta}$.

Differentiating eqn (1.9) partially with respect to $\hat{\beta}_1, \hat{\beta}_2, \cdots \hat{\beta}_k$ and setting the resulting expression to zero. This process yields k simultaneous equations in k unknowns, these equations are as follows.

$$\hat{\beta}_1 + \hat{\beta}_2 \Sigma X_{i2} + \hat{\beta}_3 \Sigma X_{i3} + \cdots + \hat{\beta}_k \Sigma X_{ik} = \Sigma Y_i$$

$$\hat{\beta}_1 \Sigma X_{i2} + \hat{\beta}_2 \Sigma X_{i2}^2 + \hat{\beta}_3 \Sigma X_{i2} X_{i3} + \cdots + \hat{\beta}_k \Sigma X_{i2} X_{ik} = \Sigma X_{i2} Y_i$$

$$\hat{\beta}_1 \Sigma X_{i3} + \hat{\beta}_2 \Sigma X_{i3} X_{i2} + \hat{\beta}_3 \Sigma X_{i3}^2 + \cdots + \hat{\beta}_k \Sigma X_{i3} X_{ik} = \Sigma X_{i3} Y_i$$

$$\hat{\beta}_1 \Sigma X_{ik} + \hat{\beta}_2 \Sigma X_{ik} X_{i2} + \hat{\beta}_3 \Sigma X_{ik} X_{i3} + \cdots + \hat{\beta}_k \Sigma X_{ik}^2 = \Sigma X_{ik} Y_i \quad \cdots (1.13)$$

In matrix form eq$^n$ (1.13) can be represented as,

$$\begin{bmatrix} n & \Sigma X_{i2} & \Sigma X_{i3} & \cdots & \Sigma X_{ik} \\ \Sigma X_{i2} & \Sigma X_{i2}^2 & \Sigma X_{i2} X_{i3} & \cdots & \Sigma X_{i2} X_{ik} \\ \Sigma X_{i3} & \Sigma X_{i3} X_{i2} & \Sigma X_{i3}^2 & \cdots & \Sigma X_{i3} X_{ik} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma X_{ik} & \Sigma X_{ik} X_{i2} & \Sigma X_{ik} X_{i3} & \cdots & \Sigma X_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{n2} \\ X_{13} & X_{23} & \cdots & X_{n3} \\ \cdots & \cdots & \cdots & \cdots \\ X_{1k} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

$$(X'X) \quad \hat{\beta} \quad = X'Y$$

or, more compactly as,

$$(X'X)\hat{\beta} = X'Y \quad \cdots (1.14)$$

Now using matrix algebra, if the inverse of $(X'X)$ exists, say $(X'X)^{-1}$, then premultiplying both sides of eqn (1.14) by this inverse, we obtain

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

But since $(X'X)^{-1}(X'X) = I$, an identity matrix of order $K \times K$, we get

$$I\hat{\beta} = (X'X)^{-1}X'Y$$

or

$$\therefore \hat{\beta} = (X'X)^{-1}X'Y \quad \ldots \quad (1.15)$$

OR

Diff. eqn (1.12), w.r.t $\hat{\beta}$, we get.

$$\frac{\partial(\hat{u}'\hat{u})}{\partial\hat{\beta}} = -2X'Y + 2X'X\hat{\beta}$$

Setting the preceding eqn to zero,

$$(X'X)\hat{\beta} = X'Y$$

$$\therefore \hat{\beta} = (X'X)^{-1}X'Y$$

provided the inverse exists.

Eqn (1.15) is a fundamental result of the OLS theory.

Variance-Covariance matrix of $\hat{\beta}$:

From eqn (1.15), we have

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

Substituting $Y = X\beta + u$ in above eqn, we get

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

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

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \quad \cdots (1.16)$$

Now

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

$$= \beta + (X'X)^{-1}X'(0) \quad \because E(u) = 0 \text{ (Assumption-2)}$$

$$E(\hat{\beta}) = \beta$$

From eqn (1.16), we obtain

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

By definition, the variance-covariance matrix of $\hat{\beta}$ is

$$\text{Var-Cov}(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}$$

$$= E\{[\hat{\beta} - \beta][\hat{\beta} - \beta]'\}$$

$$= E\{[(X'X)^{-1}X'u][(X'X)^{-1}X'u]'\}$$

$$= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \quad \cdots (1.17)$$

Noting that the $x$'s are nonstochastic, on taking expectation of eqn (1.17), we get

$$\text{Var-Cov}(\hat{\beta}) = (X'X)^{-1}X' E(UU') X(X'X)^{-1}$$

$$= (X'X)^{-1} X' \sigma^2 I X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} \quad \text{---} (1.18)$$

Where $\sigma^2$ is the homoscedastic variance of $u_i$ and $(X'X)^{-1}$ is the inverse matrix.

## * Properties of OLS estimator:

### BLUE (Best linear unbiased estimator):

From eqn (1.15), we have

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

since $(X'X)^{-1}X'$ is a matrix of fixed numbers, $\hat{\beta}$ is linear function of $Y$. Hence by definition it is linear estimator.

Substituting $Y = X\beta + u$ into the preceding expression gives

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

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

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \quad \text{---} (1.19)$$

since $(X'X)^{-1}X'X = I$.

Taking expectation of eqn (1.19) we get.

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

$$= \beta + (X'X)^{-1}X'(0) \quad | \because E(u) = 0$$

$$\therefore E(\hat{\beta}) = \beta$$

Which shows that $\hat{\beta}$ is an unbiased estimator of $\beta$.

Let $\hat{\beta}^*$ be any other linear estimator of $\beta$, which can be written as

$$\hat{\beta}^* = [(X'X)^{-1}X' + C]Y \quad \cdots (1.20)$$

Where C is a matrix of constant.

Substituting $Y = X\beta + u$ in above eqn (1.20), we get

$$\hat{\beta}^* = [(X'X)^{-1}X' + C][X\beta + u]$$

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

$$\hat{\beta}^* = \beta + (X'X)^{-1}X'u + CX\beta + Cu \quad \cdots (1.21)$$

$$E(\hat{\beta}^*) = E(\beta) + (X'X)^{-1}X'E(u) + E(CX\beta) + CE(u)$$

$$E(\hat{\beta}^*) = \beta + CX\beta$$

If $\hat{\beta}^*$ is to be an unbiased estimator of $\beta$, we must have $CX = 0$

Using eqn (1.21) and $CX = 0$, we get.

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

By definition, the Var-Cov($\hat{\beta}^*$) is,

$$E[(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'] = E[((X'X)^{-1}X'u + Cu)((X'X)^{-1}X'u + Cu)']$$

$$= E[(X'X)^{-1}X'uu'X(X'X)^{-1} + (X'X)^{-1}X'uu'C' + Cuu'X(X'X)^{-1} + Cuu'C']$$

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

$$= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1} + (X'X)^{-1}X'\sigma^2 IC' + C\sigma^2 IX(X'X)^{-1} + C\sigma^2 IC'$$

$$= \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1}(CX)' + \sigma^2 CX(X'X)^{-1} + \sigma^2 CC'$$

$$= \sigma^2(X'X)^{-1} + \sigma^2 CC' \quad \because CX = 0$$

$$= \text{Var-Cov}(\hat{\beta}) + \sigma^2 CC' \quad \text{---} \quad (1.21)$$

Which shows that the variance-covariance matrix of the alternative unbiased linear estimator $\hat{\beta}^*$ is equal to the variance-covariance matrix of the OLS estimator $\hat{\beta}$ plus $\sigma^2$ times $CC'$, which is a positive semidefinite matrix. Hence the variance of a given element of

$\hat{\beta}^*$ must necessarily be equal to or greater than the corresponding element of $\hat{\beta}$, which shows that $\hat{\beta}$ is BLUE.

Of course, if C is a null matrix i.e. $c = 0$, then $\hat{\beta}^* = \hat{\beta}$, which is another way of saying that, if we have found a BLUE estimator, it must be the least squares estimator $\hat{\beta}$.

## Distribution of SS residual

From eqn (1.12), we write.

$$\hat{u}'\hat{u} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

where $\hat{\beta} = (X'X)^{-1}X'Y$ | from eqn (1.15)

$$\therefore \hat{u}'\hat{u} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X(X'X)^{-1}X'Y$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'Y$$

$$\hat{u}'\hat{u} = Y'Y - \hat{\beta}'X'Y$$

$$\therefore \hat{u}'\hat{u} = \Sigma\hat{u}_i^2 = Y'Y - \hat{\beta}'X'Y \quad \ldots (1.22)$$

$SS_{Res} = Y'Y - \hat{\beta}'X'Y$

We know that,

$u_i \sim N(0, \sigma^2)$

$$\frac{u_i}{\sigma} \sim N(0,1)$$

$$\frac{u_i^2}{\sigma^2} \sim \chi^2_{(1)}$$

$$\therefore \frac{\Sigma u_i^2}{\sigma^2} \sim \chi^2_{(n-k)}$$

$$\frac{SS_{Res}}{\sigma^2} \sim \chi^2_{n-k}$$

& $MS_{Res} = \frac{SS_{Res}}{n-k}$

We know that, $MS_{Res}$ is an unbiased estimator of $\sigma^2$.

i.e. $E(MS_{Res}) = \sigma^2$

$\therefore \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{(n-k)} = \frac{\sum \hat{u}_i^2}{n-k}$

Now $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$

$= \sum y_i^2 - n\bar{y}^2$ --- (1.23)

$SS_T$ has d.f. $(n-1)$

$\therefore$ From $SS_T$ and $SS_{Res}$ we can calculate $SS_{Reg}$

$SS_{Reg} = SS_T - SS_{Res}$

$= \sum y_i^2 - n\bar{y}^2 - (y'y - \hat{\beta}' x' y)$ | from (1.22)

$= y'y - n\bar{y}^2 - y'y + \hat{\beta}' x' y$

$SS_{Reg.} = \hat{\beta}' x' y - n\bar{y}^2$ --- (1.24)

## The coefficient of Determination $R^2$:

The coefficient of Determination $R^2$ has been defined as,

$$R^2 = \frac{SS_{Reg}}{SS_T}$$

$$R^2 = \frac{\hat{\beta}' x' y - n\bar{y}^2}{y' y - n\bar{y}^2}$$

The coefficient of Determination $R^2$ is a summary measure that tells how well the sample regression line fits the data.

If $R^2 = 0.8$, will suggest that, 80% of total variation is explained by regressors.

$R^2$ value always increases when we add new variable in the model, even if added variable is insignificant.

✓ Relation between F and $R^2$:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F = \frac{(n-k)}{(k-1)} \frac{R^2}{(1-R^2)}$$

## * Hypothesis testing about Individual Regression coefficients:

In Regression Analysis we usually assume that each $u_i$ follows normal distribution with mean zero and constant variance $\sigma^2$.

i.e. $u_i \sim N(0, \sigma^2)$ & In matrix notation, we have, $u \sim N(0, \sigma^2 I)$

Suppose we wish to test hypothesis;

$H_0: \beta_i = \beta_{i0}$ Vs $H_1: \beta_i \neq \beta_{i0}$

Since

@ We know that,

$\hat{\beta}_i \sim N(\beta_i, \sigma^2 a_{ii})$

Where $a_{ii}$ is the ith element of $(x'x)^{-1}$.

Test Statistics testing $H_0$ is,

$$\therefore t = \frac{\hat{\beta}_i - \beta_{i0}}{\sqrt{\hat{\sigma}^2 a_{ii}}}$$

$$\therefore t = \frac{\hat{\beta}_i - \beta_{i0}}{\sqrt{\frac{\hat{u}'\hat{u}}{n-k} a_{ii}}} \sim t_{(n-k)} \quad \text{Under } H_0$$

Reject $H_0$ if $|t| > t_{\alpha/2, n-k}$.

## * 100(1-α)% confidence interval for $\beta_i$:

100(1-α)% confidence interval for $\beta_i$ is given by

$$\hat{\beta}_i \pm t_{(\alpha/2, n-k)} \sqrt{\frac{\hat{u}'\hat{u}}{n-k} a_{ii}}$$

## * Testing the overall significance of Regression:

Assuming that the distribution of $u_i \sim N(0, \sigma^2)$

$H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$
(i.e. all slope coefficients are simultaneously zero)
Vs $H_1: \beta_j \neq 0$ for at least one $j$

(i.e. Not all slope coefficients are simultaneously zero)

We know that,

$$SS_T = SS_{Reg} + SS_{Res}$$

$$\frac{SS_{Reg}}{\sigma^2} \sim \chi^2_{(k-1)} \quad \& \quad \frac{SS_{Res}}{\sigma^2} \sim \chi^2_{(n-k)}$$

be independent.

By definition of F statistics,

$$F = \frac{\frac{SS_{Reg}}{\sigma^2}/(k-1)}{\frac{SS_{Res}}{\sigma^2}/(n-k)} = \frac{\frac{SS_{Reg}}{(k-1)}}{\frac{SS_{Res}}{(n-k)}} \sim F_{(k-1, n-k)}$$

In terms of $R^2$:

$$F = \frac{SS_{Reg}/(k-1)}{SS_{Res}/(n-k)} = \frac{(n-k)}{(k-1)} \frac{SS_{Reg}}{SS_{Res}}$$

$$= \frac{(n-k)}{(k-1)} \frac{SS_{Reg}}{SS_T - SS_{Reg}}$$

$$= \frac{(n-k)}{(k-1)} \frac{SS_{Reg} / SS_T}{(SS_T - SS_{Reg}) / SS_T}$$

$$= \frac{(n-k)}{(k-1)} \frac{R^2}{1-R^2}$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

When $R^2 = 0$, F is ipso facto.
When $R^2 = 1$, F is infinite.
The larger the $R^2$, the greater the F value. Thus the F test, which is measure of the overall significance of the estimated regression, is also a test of significance of $R^2$. In other words, testing the null hypothesis is equivalent to testing the null hypothesis that $R^2$ is zero.

ANOVA table

| Source of variation | SS | d.f | MSS |
|---|---|---|---|
| Due to regression (that is, $X_2, X_3, \ldots, X_k$) | $\hat{\beta}' X' y - n\bar{y}^2$ | $k-1$ | $\frac{\hat{\beta}' X' y - n\bar{y}^2}{k-1}$ |
| Due to residuals | $y'y - \hat{\beta}' X' y$ | $n-k$ | $\frac{y'y - \hat{\beta}' X' y}{n-k}$ |
| Total | $y'y - n\bar{y}^2$ | $n-1$ | |

We reject $H_0$, if

$$F > F_{\alpha, (k-1, n-k)}$$

ANOVA table in terms of $R^2$.

| Source of Variation | SS | d.f. | MSS |
|---|---|---|---|
| Due to regression ($i.e.$ due to $X_2, X_3, \cdots, X_k$) | $R^2(Y'Y - n\bar{Y}^2)$ | $(k-1)$ | $\frac{R^2(Y'Y - n\bar{Y}^2)}{(k-1)}$ |
| Due to residuals | $(1-R^2)(Y'Y - n\bar{Y}^2)$ | $(n-k)$ | $\frac{(1-R^2)(Y'Y - n\bar{Y}^2)}{(n-k)}$ |
| Total | $Y'Y - n\bar{Y}^2$ | $(n-1)$ | |

We Reject $H_0$, if $F > F_\alpha(k-1, n-k)$

i.e. $R^2$ is not zero.

## Prediction Using Multiple Regression:

How the estimated multiple regression can be used for predicting (1) the mean (2) individual value of $Y$, given the values of $X$ regressors.

(1) Mean Prediction:

Let $$X_0 = \begin{bmatrix} 1 \\ X_{02} \\ X_{03} \\ \vdots \\ X_{0k} \end{bmatrix} \quad \cdots (1.25)$$

be the vector of values of the $X$ variables for which we wish to predict $\hat{Y}_0$, the mean prediction of $Y$.

Now the estimated multiple regression, in scalar form is,

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \cdots + \hat{\beta}_k X_{ik} \quad \cdots (1.26)$$

which in matrix notation can be written as

$$\hat{Y}_i = x_i' \hat{\beta} \quad \cdots (1.27)$$

Where $x_i' = [1 \quad x_{i2} \quad x_{i3} \quad \cdots \quad x_{ik}]$

and $$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

Equation (1.26) or (1.27) is of course the mean prediction of $Y_i$ corresponding to given $x_i'$.

If $x_i'$ is as given in (1.25), (7.27) becomes

$$(\hat{Y}_i / x_0') = x_0' \hat{\beta} \quad \cdots (1.28)$$

where, of course, the values of $x_0$ are specified. Note that eq$^n$ (1.28) gives an unbiased prediction of $E(Y_i / x_0')$. Since $E(x_0' \hat{\beta}) = x_0' \hat{\beta}$.

Variance of Mean Prediction:

The formula to estimate the variance of $(\hat{Y}_0 / x_0')$ is as follows:

$$\text{Var}(\hat{Y}_0 / x_0') = \sigma^2 x_0' (X'X)^{-1} x_0 \quad \cdots (1.29)$$

Where $\sigma^2$ is variance of $u_i$, $x_0'$ are the given values of the X variables for which we wish to predict, and $(X'X)$ is the matrix given in eq$^n$ (1.14). In practice, we replace $\sigma^2$ by its unbiased estimator $\hat{\sigma}^2$.

## Generalized Least Squares:

Let us retain all the previous assumption of CLRM and $E(u)=0$ but now we add to it that $E(uu') = \sigma^2 V$ where $\sigma^2$ is unknown and $V$ is $n \times n$ matrix (positive definite matrix)

To see this return to eqn (1.14).

To take into account heteroscedastic variances [The element on the main diagonal of eqn (1.4)] and autocorrelations in the error terms [the elements off the main diagonal of eqn (1.4)] assume that $E(uu') = \sigma^2 V$

Therefore, our model is,

$$Y = X\beta + u \qquad \text{---(1.30)}$$

As $V$ is a positive definite, there exist a non-singular matrix $P$ such that $V = PP'$ where $P$ is $n \times n$ matrix. Premultiplying eqn (1.30) by $P^{-1}$, we get

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u$$

$$Y^* = X^*\beta + u^* \qquad \text{---(1.31)}$$

Where $Y^* = P^{-1}Y$, $X^* = P^{-1}X$ & $u^* = P^{-1}u$.

It is easy to see that,

$$E(u^*) = E[P^{-1}u]$$
$$= P^{-1}E(u)$$
$$= P^{-1}(0)$$
$$= 0$$

$$E(u^* u^{*\prime}) = \text{Var}(u^*)$$
$$= \text{var}(P^{-1}u)$$
$$= P^{-1}\,\text{Var}(u)\,(P^{-1})'$$
$$= P^{-1}\sigma^2 V (P^{-1})' \quad \because E(uu') = \sigma^2 V$$
$$= P^{-1}\sigma^2 PP'(P^{-1})'$$
$$= \sigma^2 P^{-1}P(P^{-1}P)'$$
$$= \sigma^2 I$$

So that eqn (1.31) satisfy all the assumption required for OLS method

The estimate of parameter $\hat{\beta}$ is $\hat{\beta}^{gls}$

$$\hat{\beta}^{gls} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$$
$$= \left((P^{-1}X)'P^{-1}X\right)^{-1}(P^{-1}X)'P^{-1}Y$$
$$= \left(X'(P^{-1})'P^{-1}X\right)^{-1}X'(P^{-1})'P^{-1}Y$$
$$= \left(X'(PP')^{-1}X\right)^{-1}X'(PP')^{-1}Y$$

$$\hat{\beta}^{gls} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad \text{---(1.32)}$$

* $\hat{\beta}^{gls}$ is unbiased estimator of $\beta$.

Now

$$\hat{\beta}^{gls} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

Substituting $Y = X\beta + u$ into preceding expression gives,

$$\hat{\beta}^{gls} = (X'V^{-1}X)^{-1}X'V^{-1}(X\beta+U)$$

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

$$\hat{\beta}^{gls} = \beta + (X'V^{-1}X)^{-1}X'V^{-1}U \quad \ldots (1\text{-}33)$$

Taking expectation of eqn (1-33), we get

$$E(\hat{\beta}^{gls}) = E(\beta) + (X'V^{-1}X)^{-1}X'V^{-1}E(U)$$

$$= E(\beta) + (X'V^{-1}X)^{-1}X'V^{-1}(0)$$

$$E(\hat{\beta}^{gls}) = \beta \quad \ldots (1\text{-}34)$$

Hence $\hat{\beta}^{gls}$ is unbiased estimator of $\beta$.

## Variance-covariance matrix of $\hat{\beta}^{gls}$:

From eqn (1.30), our model is,

$$Y = X\beta + U$$

& converted model is, from eqn (1-31)

$$Y^* = X^*\beta + U^*$$

where $U^* \sim N(0, \sigma^2 I)$

$$V(\hat{\beta}) = \sigma^2(X^{*'}X^*)^{-1}$$

$$= \sigma^2\left[(P^{-1}X)'(P^{-1}X)\right]^{-1} \quad \left| \because V = PP' \right.$$

$$= \sigma^2(X'(P^{-1})'P^{-1}X)^{-1}$$

$$= \sigma^2(X'(PP')^{-1}X)^{-1} = \sigma^2(X'V^{-1}X)^{-1}$$

## Heteroscedasticity:

As noted in previous chapter, one of the important assumptions of the classical linear regression model is that the variance of each disturbance term $u_i$, conditional on the chosen values of the explanatory variables, is some constant number equal to $\sigma^2$. This is the assumption of homosce-dasticity, or equal (homo) spread (scedasticity), that is, equal variance,

symbolically,

$$E(u_i^2) = \sigma^2 \quad , \quad i = 1, 2, \cdots, n \qquad \cdots (1.35)$$

Diagrammatically, in the two-variable regression model homoscedasticity can be shown as in fig. (1.4), ~~which, for convenience is reproduced as~~

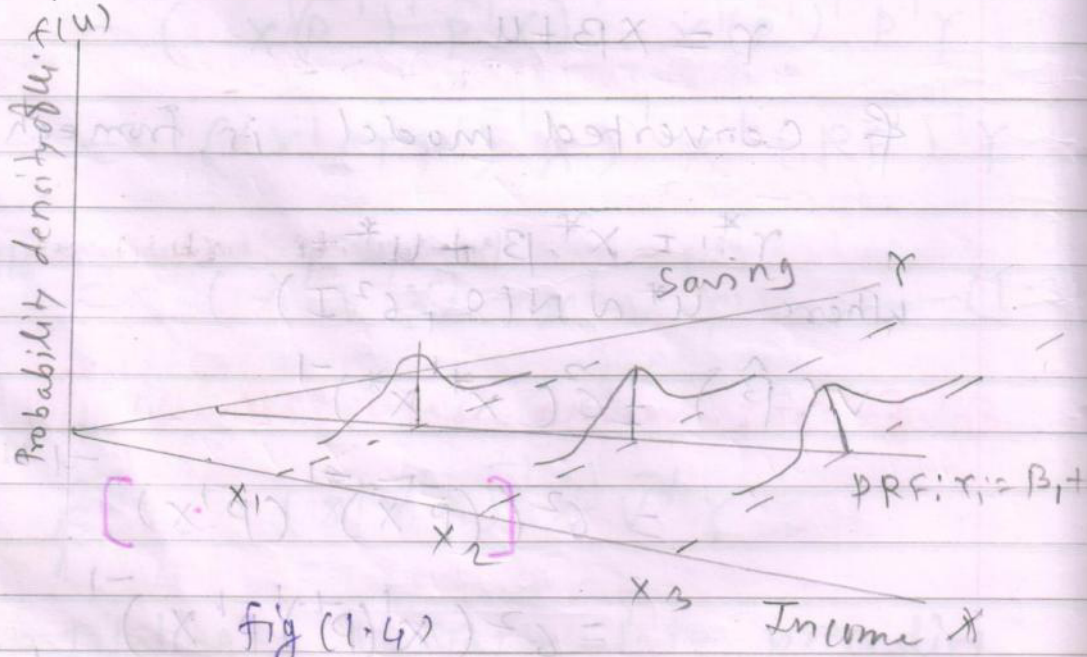



fig (1.4)
Homoscedasticity

In constrast, consider figure (1.5), which shows that the conditional variance of $Y_i$ increases as X increases. Here the variance of $Y_i$ are not the same. Hence there is heteroscedasticity,

Symbolically,

$$E(u_i^2) = \sigma_i^2 \qquad (1.36)$$

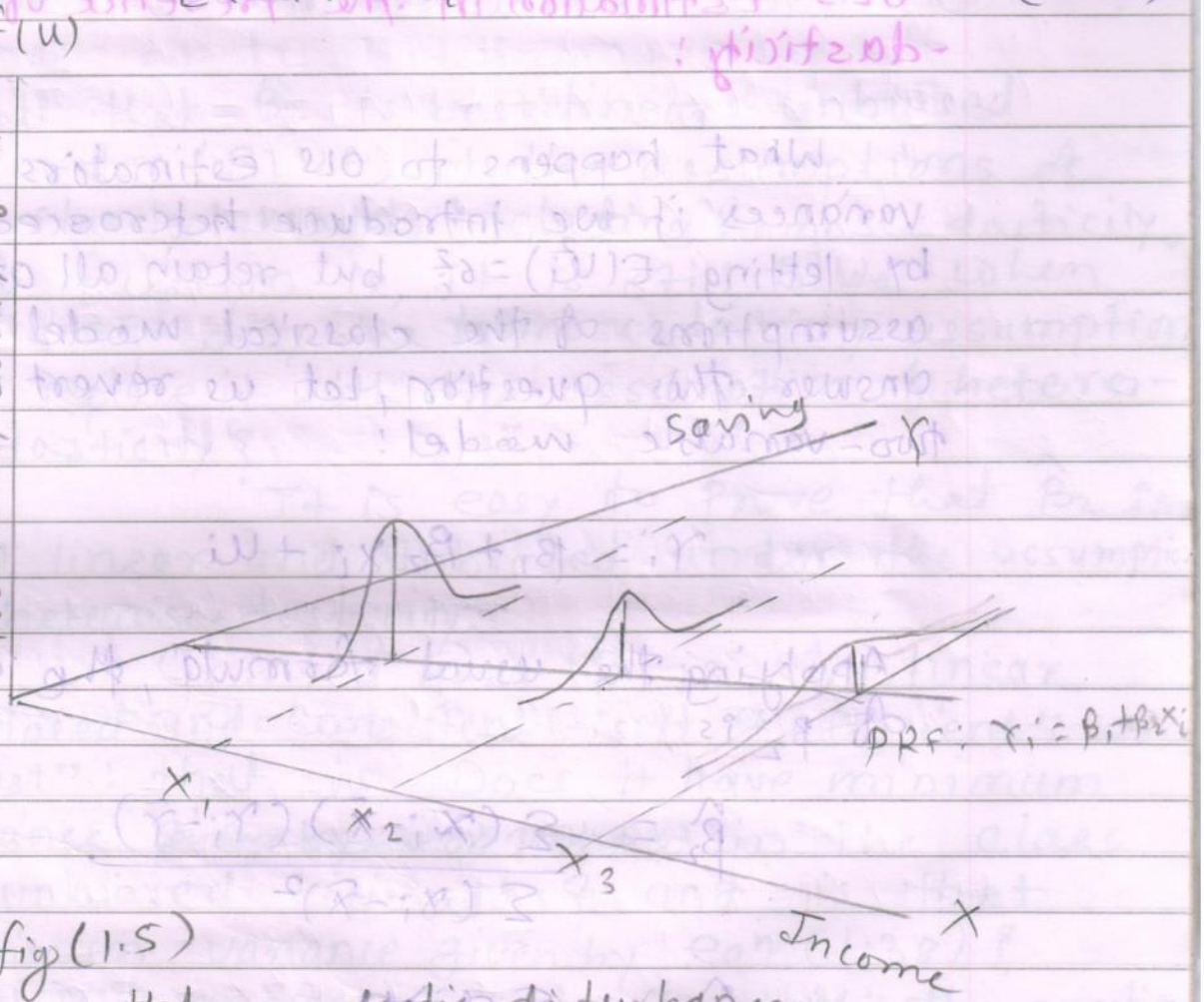



fig (1.5)
Heteroscedastic disturbances.

To make the difference between homoscedasticity and heteroscedasticity clear, assume that in the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y represents saving and X represents income, fig. (1.4) and (1.5) show that as income increases, savings on the average also increases. But in

figure (1.4) the variance of savings remains the same at all levels of income, whereas in fig. (1.5) it increases with income. It seems that in fig (1.5) the higher income families on the average save more than the lower-income families, but there is also more variability in their savings.

## OLS Estimation in the Presence of Heteroscedasticity:

What happens to OLS estimators and their variances if we introduce heteroscedasticity by letting $E(u_i^2) = \sigma_i^2$ but retain all other assumptions of the classical model? To answer this question, let us revert to the two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Applying the usual formula, the OLS estimator of $\beta_2$ is,

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \qquad \text{---} \quad (1.37)$$

but its variance is now given by the following expression

$$V(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \qquad \text{---} \quad (1.38)$$

which is obviously different from the usual variance formula obtained under the assumption of homoscedasticity,

$$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad \ldots\ldots (1.39)$$

of course, if $\sigma_i^2 = \sigma^2$ for each $i$, the above two formulas will be identical.

Recall that $\hat{\beta}_2$ is best linear unbiased estimator (BLUE) if the assumptions of the classical model, including homoscedasticity. Now question is, It is still BLUE when we drop only the homoscedasticity assumption and replace it with the assumption of heteroscedasticity?

It is easy to prove that $\hat{\beta}_2$ is still linear and unbiased under the assumption of heteroscedasticity.

Granted that $\hat{\beta}_2$ is still linear unbiased and consistent, is it "efficient" or "best": that is, Does it have minimum variance ~~given by eqn (1.38)~~ in the class of unbiased estimators? and is that minimum variance given by eqn (1.38)?

The answer is "no" to both the questions. $\hat{\beta}_2$ is no longer best and the minimum variance is not given by eqn (1.38). Then what is BLUE in the presence of heteroscedasticity?

## * The method of Generalized least squares (GLS)

Why is the usual OLS estimator of $\beta_2$ given in the equation (1.37) not best, although it is still unbiased?

The method of estimation, known as generalized least squares (GLS) capable of producing ~~information into account explicitly~~ estimaters that are BLUE. To see how this is accomplished, Let us continue with the now-familiar two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{---} (1.40)$$

Which for ease of algebraic manipulation we write as

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \quad \text{---} (1.41)$$

Where $X_{0i} = 1$ for each $i$.

Now assume that the heteroscedastic variances $\sigma_i^2$ are known. Divide eqn (1.41) by $\sigma_i$:

$$\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \quad \text{---} (1.42)$$

Which for ease of exposition we write as

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \quad \text{---} (1.43)$$

Where the starred, or transformed, variables are the ~~the~~ original variables divided by the (known) $\sigma_i$. We use the notation $\beta_1^*$ and $\beta_2^*$,

(GLS)

the parameters of the transformed model, to distinguish them from the usual OLS parameters $\beta_1$ and $\beta_2$.

What is the purpose of transforming the original model? To see this the following feature of the transformed error term $u_i^*$:

$$\text{var}(u_i^*) = E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2$$

$$= \frac{1}{\sigma_i^2} E(u_i^2) \qquad \text{Since } \sigma_i^2 \text{ is known}$$

$$= \frac{1}{\sigma_i^2}(\sigma_i^2) \qquad \text{Since } E(u_i^2) = \sigma_i^2$$

$$= 1$$

Which is a constant. That is variance of the transformed disturbance term $u_i^*$ is now homoscedastic. Since we are still retaining the other assumptions of the classical model, the finding that it is $u^*$ that is homoscedastic suggests that if we apply OLS to the transformed model (11.42), it will produce estimators that are BLUE. In short, the estimated $\beta_1^*$ and $\beta_2^*$ are now BLUE and not the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

The actual mechanics of estimating $\beta_1^*$ and $\beta_2^*$ are as follows.

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^*$$

Now, to obtain the GLS estimators, we minimize

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2 \quad \text{---(1.44)}$$

that is

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right)\right]^2 \quad \text{---(1.45)}$$

The GLS estimator of $\beta_2^*$ is,

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \quad \text{---(1.46)}$$

and its variance is given by,

$$V(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \quad \text{---(1.47)}$$

where $w_i = 1/\sigma_i^2$

**Difference between OLS and GLS:**

In OLS we minimize

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{B}_1 - \hat{B}_2 X_i)^2 \quad \text{---(1.48)}$$

But in GLS, we minimize,

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2$$

where $w_i = \frac{1}{\sigma_i^2}$

Thus in GLS we minimize a weighted sum of residual squares with $w_i = 1/\sigma_i^2$ acting as the weights, but in OLS we minimize an unweighted or equally weighted RSS. As eqn (1.4 shows in GLS the weight assigned to each observation is inversely proportional to it's $\sigma_i$, that is, observations coming from a population with larger $\sigma_i$ will get relatively smaller weight and those from population with smaller $\sigma_i$ will get proportionately larger weights in minimizing the RSS (1.49)

Since (1.49) minimizes a weighted RSS, it is appropriately known as weighted least squares (WLS) and the estimators thus obtained and given in eqn (1.46) & (1.47) are known as WLS estimators. But WLS is just a special case of the more general estimating technique, GLS. In the context of heteroscedasticity, one can treat the two terms WLS and GLS interchangeably.

Note that if $w_i = w$, a constant for all $\hat{\beta}_2^*$ is identical with $\hat{\beta}_2$ and $V(\hat{\beta}_2^*)$ is identical with the usual (i.e. homoscedastic) var $(\hat{\beta}_2)$.

## Detection of Heteroscedasticity:

How does one know that heteroscedasticity is present in a specific situation? There are no hard-and-fast rules for detecting heteroscedasticity, only a few rules of thumb. There are two methods commonly used researchers, namely, informal method and formal method for detection of Heteroscedasticity

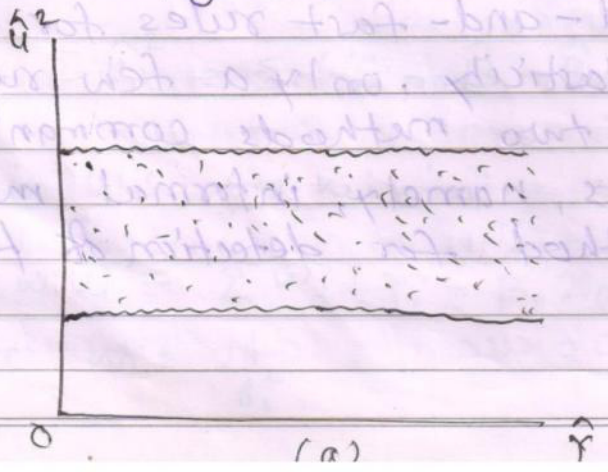
## * (i) Informal Method:

### (i) Nature of the Problem:

Very often the nature of the Problem under consideration suggests whether heteroscedasticity is likely to be encountered. For example, following the pioneering work of Prais and Houthakker on family budget studies, where they found that the residual variance around the regression of consumption on income increased with income, one now generally assumes that in similar surveys one can expect unequal variances among the disturbances

### (ii) Graphical Method:

If there is no a priori or empirical information about the nature of heteroscedasticity, in practice one can do the regression analysis on the assumption that there is no heteroscedasticity and then do a postmortem examination of the residual squared $\hat{u}_i^2$ to see if they exhibit any systematic pattern. Although $\hat{u}^2$ are not the same thing as $u_i^2$, they can be used as proxies especially if the sample size is sufficiently large. An examination of the $\hat{u}_i^2$ may reveal patterns such as those shown in fig (1.6).




(a)

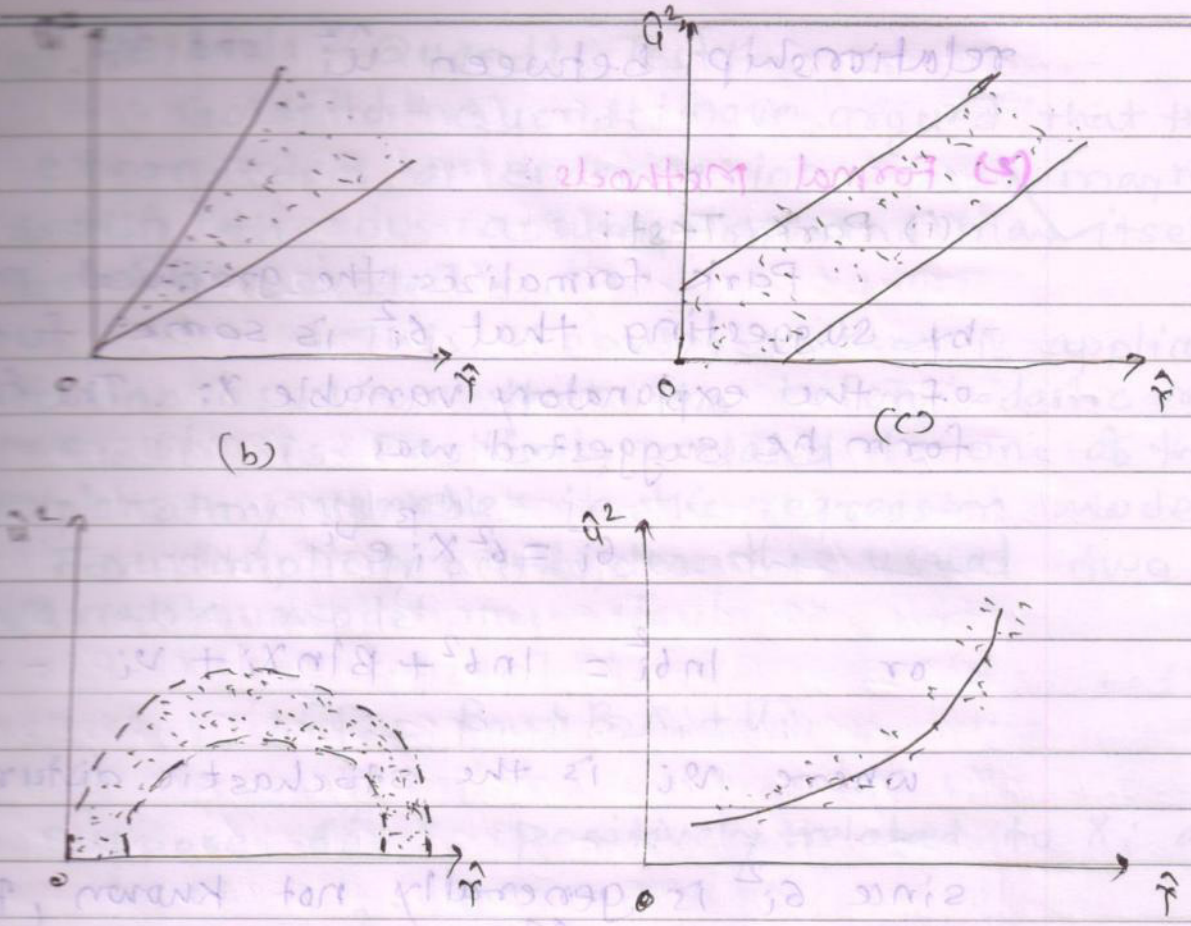



fig (1.6) Hypothetical patterns of estimated squared residuals.

In figure (1.6), $\hat{u}_i^2$ are plotted against $\hat{Y}_i$, the estimated $Y_i$ from the regression line, the idea being to find out whether the estimated mean value of Y is systematically related to the squared residual. In fig. 1.6(a) we see that there is no systematic pattern between the two variables, suggesting that perhaps no heteroscedasticity is present in the data. Figure 1.6(b) to (e), however, exhibits definite patterns. For instance, fig 1.6 (c) suggests a linear relationship, whereas figure 1.6(d) and (e) indicates a quadratic

relationship between $\hat{u}_i^2$ and $\hat{Y}_i$.

(*) Formal methods:

(i) Park Test:

Park formalizes the graphical method by suggesting that $\sigma_i^2$ is some function of the explanatory variable $X_i$. The functional form he suggested was

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$$

or $$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad \text{---} (1.50)$$

where $v_i$ is the stochastic disturbance term

since $\sigma_i^2$ is generally not known, Park suggests using $\hat{u}_i^2$ as a proxy and running the following regression:

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad \text{---} (1.51)$$

$$= \alpha + \beta \ln X_i + v_i \quad \text{---} (1.52)$$

If $\beta$ turns out to be statistically significant it would suggest that heteroscedasticity is present in the data. If it turns out to be insignificant we may accept the assumption of homoscedasticity. The Park test is thus a two stage procedure. In the first stage we run the OLS regression disregarding the heteroscedasticity question. We obtain $\hat{u}_i$ from this regression, and then in second stage we run the regression (1.52).

(ii) Goldfeld - Quandt Test:

Goldfeld - Quandt have argued that the error term $u_i$ entering into (1.52) may not satisfy the OLS assumptions and may itself be heteroscedastic.

This popular method is applicable if one assumes that the heteroscedastic variance, $\sigma_i^2$, is positively related to one of the explanatory variables in the regression model. For simplicity, consider the usual two variable model,

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

suppose $\sigma_i^2$ is positively related to $X_i$ as

$$\sigma_i^2 = \sigma^2 X_i^2 \quad \text{---} \quad (1.53)$$

where $\sigma^2$ is a constant.

Assumption (1.53) postulates that $\sigma_i^2$ is proportional to the square of the X variable. If (1.53) is appropriate, it would mean $\sigma_i^2$ would be larger the values of $X_i$. If that turns out to be the case, heteroscedasticity is most likely to be present in the model. To test this explicitly, Goldfeld and Quandt suggest the following steps:

step 1:

Order or rank the observations according to the values of $X_i$, beginning with the lowest X value.

(ii) Goldfeld - Quandt Test:

Goldfeld - Quandt have argued that the error term $u_i$ entering into (1.52) may not satisfy the OLS assumptions and may itself be heteroscedastic.

This popular method is applicable if one assumes that the heteroscedastic variance, $\sigma_i^2$, is positively related to one of the explanatory variables in the regression model. For simplicity, consider the usual two variable model,

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Suppose $\sigma_i^2$ is positively related to $X_i$ as

$$\sigma_i^2 = \sigma^2 X_i^2 \quad \text{---} (1.53)$$

where $\sigma^2$ is a constant.

Assumption (1.53) postulates that $\sigma_i^2$ is proportional to the square of the X variable. If (1.53) is appropriate, it would mean $\sigma_i^2$ would be larger the values of $X_i$. If that turns out to be the case, heteroscedasticity is most likely to be present in the model. To test this explicitly, Goldfeld and Quandt suggest the following steps:

step 1:

Order or rank the observations according to the values of $X_i$, beginning with the lowest X value.

Step 2:
Omit c central observations, where c is specified a priori, and divide the remaining $(n-c)$ observations into two groups each of $(n-c)/2$ observations.

Step 3:
Fit separate OLS regressions to the first $(n-c)/2$ observations and the last $(n-c)/2$ observations and obtain the respective residual sums of squares $RSS_1$ and $RSS_2$, $RSS_1$ representing the RSS from the regression corresponding to the smaller $X_i$ values (the smaller variance group) and $RSS_2$ that from the larger $X_i$ values (the larger variance group).

These RSS each have,

$$\frac{(n-c)}{2} - k \quad \text{or} \quad \left(\frac{n-c-2k}{2}\right) df$$

where k is the number of parameters to be estimated, including the intercept.

Step 4:
compute the ratio,

$$\lambda = \frac{RSS_2/df}{RSS_1/df} \quad \ldots (1.54)$$

If $u_i$ are assumed to be normally distributed (which we usually do), and if the assumption of homoscedasticity is valid, then it can be shown that $\lambda$ of eqn (1.54) follows the F-distribution with numerator and denominator df each $\frac{(n-c-2k)}{2}$.

If $F > F\left(\frac{n-c}{2}-k, \frac{n-c}{2}-k\right)$,

we reject the hypothesis of Homoscedasticity.

The ability of the Goldfeld-Quandt test to do this successfully depends on how c is choosen. For the two variable model the Monte Carlo experiments done by Goldfeld and Quandt suggest that c is about 8 if the sample size is about 30, and it is about 16 if the sample size is about 60. But Judge et al. note that c=4 if n=30 and c=10 if n is about 60 have been found satisfactory in practice.

(iii) White's General Heteroscedasticity Test:

Unlike the Goldfeld-Quandt test, which requires recordering the observations with resp. to the X variable that supposedly caused heteroscedasticity, The general test of heteroscedasticity proposed by White does not rely on the normality assumption and is easy to implement. As an illustration of the basic idea, consider the following three-variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i \quad \cdots (1.55)$$

The White test proceds as follows:

Step 1:

Given the data, we estimate eqn (1.55) and obtain the residuals, $\hat{u}_i$.

Step 2:

We then run the following regression

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \alpha_4 X_{i2}^2 + \alpha_5 X_{i3}^2 + \alpha_6 X_{i2} X_{i3} + v_i \quad \text{---(1.56)}$$

That is, the squared residuals from the original regression are regressed on the original X variables or regressors, their squared values, and the cross product(s) of the regressors. Obtain the $R^2$ from this regression.

Step 3:

Under the null Hypothesis that there is no heteroscedasticity, it can be shown that sample size (n) times the $R^2$ asymptotically follows the chi-square distribution with df equal to the number of regressors in the auxiliary regression

i.e. $$n \cdot R^2 \underset{asy}{\sim} \chi^2_{df} \quad \text{(1.57)}$$

Where df is defined previously. In our example, there are 5 df since there are 5 regressors in the auxiliary regression.

Step 4:

If the chi-square value exceeds the critical chi-square value at the choosen level of significance, the conclusion is that there is heteroscedasticity. If it does not exceed the critical chi-square value, there is no heteroscedasticity, which is to say that in the auxiliary regression (1.55), $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$.

# Unit-2

## Autocorrelation:

If the assumption of classical linear regression model, the disturbances $u_t$ are uncorrelated is violated then problem of autocorrelation or serial correlation arises.

Autocorrelation can arise due to several reasons, namely, sluggishness of economic time series, specification Bias resulting from excluded variable from the model or specification bias from Incorrect functional form, data manipulation etc.

## The nature of the problem:

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

In the regression context, the classical linear regression model assumes that such autocorrelation does not exist in the disturbances $u_i$.

Symbolically,

$$E(u_i, u_j) = 0 \quad , \quad i \neq j \qquad \ldots (2.1)$$

put simply, the classical model assumes that the disturbance term relating to any observation is not influenced by the disturbance term relating to any other observation.

However, if there is such a dependence, we have autocorrelation

Symbolically, $E(u_i, u_j) \neq 0 \quad , \quad i \neq j$ ... (2.2)

It should be noted also that autocorrelation can be positive as well as negative, although most economic time series generally exhibit positive autocorrelation because most of them ether move upward or downward over extended time periods and donot exhibit a constant up and down movement. ~~such as that~~

## * Consequences of Autocorrelation:

What happens to the OLS estimators and their variances if we introduce autocorrelation in the disturbances by assuming that

$$E(u_t, u_{t+s}) \neq 0, \text{ where } s \neq 0$$

but retain all the other assumptions of the classical linear regression model.

Note again that we are now using the subscript t on the disturbances to emphasize that we are dealing with time series data.

Consider two variable regression model to explain the basic ideas involved

$$Y_t = B_1 + \beta_2 X_t + u_t \quad \ldots(2.3)$$

To make any headway, we must assume the mechanism that generates $u_t$, for $E(u_t, u_{t+s}) \neq 0$ $(s \neq 0)$ is too general an assumption to be of any practical use. As a starting point, or first approximation, one can assume that the disturbance, or error, terms are generated by the following mechanism,

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

(2.5)

where $\rho$ is known as the coefficient of autocovariance and where $\epsilon_t$ is the stochastic disturbance term such that it satisfied the standard OLS assumption namely,

$$E(\epsilon_t) = 0$$
$$Var(\epsilon_t) = \sigma_\epsilon^2$$
$$cov(\epsilon_t, \epsilon_{t+s}) = 0, \quad s \neq 0 \qquad \cdots\cdots (2.5)$$

The scheme (2.4) is known as Markov first order autoregressive scheme, or simply a first order autoregressive scheme, usually denoted by AR(1).

The name autoregressive is appropriate because (2.4) can be interpreted as the regression of $u_t$ on itself lagged one period. It is first order because $u_t$ and its immediate past value are involved; that is, the maximum lag is 1. If the model were $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t$, it would be an AR(2), or second-order, autoregressive scheme.

We have,

$$u_t = \rho u_{t-1} + \epsilon_t$$
$$= \rho [\rho u_{t-2} + \epsilon_{t-1}] + \epsilon_t$$
$$= \rho^2 u_{t-2} + \epsilon_t + \rho \epsilon_{t-1}$$
$$= \rho^2 [\rho u_{t-3} + \epsilon_{t-2}] + \epsilon_t + \rho \epsilon_{t-1}$$
$$= \rho^3 u_{t-3} + \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2}$$
$$= \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \rho^3 \epsilon_{t-3} + \cdots$$

$$u_t = \sum_{r=0}^{\infty} \rho^r \epsilon_{t-r} \quad \ldots (2.6)$$

$$E(u_t) = E\left(\sum_{r=0}^{\infty} \rho^r \epsilon_{t-r}\right)$$

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

$$= 0 \qquad \because E(\epsilon_{t-r}) = 0 \ \forall t.$$

$$\text{Var}(u_t) = \text{var}\left(\sum_{r=0}^{\infty} \rho^r \epsilon_{t-r}\right)$$

$$= \text{var}(\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \rho^3\epsilon_{t-3} + \ldots)$$

$$= \text{var}(\epsilon_t) + \rho^2 \text{var}(\epsilon_{t-1}) + \rho^4 \text{var}(\epsilon_{t-2}) + \rho^6 \text{var}(\epsilon_{t-3}) + \ldots$$

$$(\because \text{cov}(\epsilon_t, \epsilon_{t+s}) = 0 \text{ if } s \neq 0)$$

$$= \sigma_\epsilon^2 + \rho^2\sigma_\epsilon^2 + \rho^4\sigma_\epsilon^2 + \rho^6\sigma_\epsilon^2 + \ldots$$

$$= \sigma_\epsilon^2 (1 + \rho^2 + \rho^4 + \rho^6 + \ldots)$$

$$\text{Var}(u_t) = \frac{\sigma_\epsilon^2}{1-\rho^2} = \sigma_u^2 \quad \ldots (2.7)$$

Now

$$\text{cov}(u_t, u_{t-1}) = E(u_t \cdot u_{t-1}) - E(u_t)E(u_{t-1})$$

$$= E(u_t \cdot u_{t-1}) - (0)(0)$$

$$= E(u_t \cdot u_{t-1})$$

$$\therefore \text{Cov}(u_t, u_{t-1}) = E\left[\left(\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \rho^3\epsilon_{t-3} + \cdots\right)\right.$$
$$\left.\left(\epsilon_{t-1} + \rho\epsilon_{t-2} + \rho^2\epsilon_{t-3} + \rho^3\epsilon_{t-4} + \cdots\right)\right]$$

$$= E\left(\rho\epsilon_{t-1}^2 + \rho^3\epsilon_{t-2}^2 + \rho^5\epsilon_{t-3}^2 + \cdots\right)$$

$$= \rho E(\epsilon_{t-1}^2) + \rho^3 E(\epsilon_{t-2}^2) + \rho^5 E(\epsilon_{t-3}^2 + \cdots$$

$$= \rho\sigma_\epsilon^2 + \rho^3\sigma_\epsilon^2 + \rho^5\sigma_\epsilon^2 + \cdots$$

$$= \sigma_\epsilon^2\left(\rho + \rho^3 + \rho^5 + \cdots\right)$$

$$= \sigma_\epsilon^2\rho\left(1 + \rho^2 + \rho^4 + \cdots\right)$$

$$= \sigma_\epsilon^2 \frac{\rho}{1-\rho^2}$$

$$= \rho\frac{\sigma_\epsilon^2}{1-\rho^2}$$

$$= \rho\sigma_u^2 \quad \cdots (2.8)$$

Similarly

$$\text{Cov}(u_t, u_{t-2}) = \rho^2\sigma_u^2$$

$$\text{Cov}(u_t, u_{t-s}) = \rho^s\sigma_u^2 \quad \cdots (2.9)$$

Now

$$\text{Cor}(u_t, u_{t-s}) = \frac{\text{Cov}(u_t, u_{t-s})}{\sqrt{\text{Var}(u_t)\cdot\text{Var}(u_{t-s})}}$$

$$\text{Cor}(u_t, u_{t-s}) = \frac{\rho^s \sigma_u^2}{\sqrt{\sigma_u^2 \cdot \sigma_u^2}}$$

$$= \frac{\rho^s \sigma_u^2}{\sigma_u^2}$$

$$\text{Cor}(u_t, u_{t-s}) = \rho^s$$

Note that because of symmetry property of covariances and correlations, $\text{cov}(u_t, u_{t-s}) = \text{cov}(u_t, u_{t+s})$ and $\text{cor}(u_t, u_{t-s}) = \text{cor}(u_t, u_{t+s})$

Since $\rho$ is a constant between $-1$ and $+1$ shows that [eqn (2.7)] under the AR(1) scheme the variance of $u_t$ is still homoscedastic, but is correlated not only with immediate past value but its values several periods in the past.

Now, From eqn (2.9) and symmetry property of covariances, we can write.

$$\therefore \rho^s = \frac{\text{cov}(u_t, u_{t+s})}{\sigma_u^2} = \frac{\text{cov}(u_t, u_{t-s})}{\sigma_u^2}$$

∴ Take $s = 0$

$$\rho^0 = \frac{\text{cov}(u_t, u_t)}{\sigma_u^2} = \frac{\text{Var}(u_t)}{\sigma_u^2} = \frac{\sigma_u^2}{\sigma_u^2} = 1$$

∴ Now Take $s = 1$

$$\rho = \frac{cov(u_t, u_{t+1})}{\sigma_u^2}$$

It is called 1st order autocorrelation.

Now, collecting term together, we obtain,

$$E(\underline{u}\underline{u}') = V = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \quad \ldots (2.10)$$

## * * OLS Estimation in presence of Autocorrelation:

Consider two-variable regression model ~~in deviated form~~,

$$~~Y_t = \beta x~~ \quad Y_t = \beta_1 + \beta_2 X_t + u_t.$$

The OLS estimator of slope coefficient is

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \quad \ldots (2.11)$$

& its variance is given by

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum x_t^2} \quad \ldots (2.12)$$

Where the small letters as usual denote deviation from the mean values.

$$\text{Var-cov}(\hat{\beta}) = (X'X)^{-1} X' \sigma^2 I X (X'X)^{-1}$$

Now under the AR(1) scheme, it can be shown that the variance of this estimate is,

$$\underset{AR1}{\text{Var}}(\hat{\beta}_2) = \frac{\sigma_u^2}{(\Sigma x_t^2)^2} [x_1\ x_2\ x_3 \cdots x_n] \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & & & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \frac{\sigma_u^2}{(\Sigma x_t^2)^2} [x_1 + \rho x_2 + \rho^2 x_3 + \cdots + \rho^{n-1} x_n \quad \rho x_1 + x_2 + \rho x_3 + \cdots + \rho^{n-2} x_n \quad \rho^2 x_1 + \rho x_2 + x_3 + \cdots$$
$$\cdots \quad \rho^{n-1} x_1 + \rho^{n-2} x_2 + \rho^{n-3} x_3 + \cdots + x_n] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \frac{\sigma_u^2}{(\Sigma x_t^2)^2} [x_1^2 + \rho x_2 x_1 + \rho^2 x_3 x_1 + \cdots + \rho^{n-1} x_n x_1 +$$
$$+ \rho x_1 x_2 + x_2^2 + \rho x_3 x_2 + \cdots + \rho^{n-2} x_n x_2 +$$
$$+ \rho^2 x_1 x_3 + \rho x_2 x_3 + x_3^2 + \cdots + \rho^{n-3} x_n x_3 +$$
$$\vdots$$
$$+ \rho^{n-1} x_1 x_n + \rho^{n-2} x_2 x_n + \rho^{n-3} x_3 x_n + \cdots + x_n^2]$$

$$= \frac{\sigma_u^2}{(\Sigma x_t^2)^2} \left[ \sum_{t=1}^{n} x_t^2 + 2\rho \sum_{t=1}^{n-1} x_t x_{t-1} + 2\rho^2 \sum_{t=1}^{n-2} x_t x_{t-2} + \cdots + 2\rho^{n-1} x_1 x_n \right]$$

$$= \frac{\sigma_u^2}{(\Sigma x_t^2)^2} \left[ 1 + 2\rho \frac{\sum_{t=1}^{n-1} x_t x_{t-1}}{\Sigma x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{n-2} x_t x_{t-2}}{\Sigma x_t^2} + 2\rho^{n-1} \frac{x_1 x_n}{\Sigma x_t^2} \right]$$

Where $\text{Var}(\hat{\beta}_2)_{AR_1}$ means the variance of $\hat{\beta}_2$ under first order autoregressive scheme.

A comparision of (2.13) with (2.12) shows the former is equal to the latter times a term that depends on $\rho$ as well as the sample autocorrelations between the values taken by the regressor X at various lages.

If $\rho$ is zero, then two formulas will coincide.

To give some idea about the difference between the variances given in (2.12) and (2.13), assume that the regressor X also follows the first order autoregressive scheme with a coefficient of autocorrelation of r. Then eqn (2.13) reduces to:

$$\text{Var}(\hat{\beta}_2)_{AR_1} = \frac{\sigma_u^2}{(\Sigma x_t^2)}\left[1 + 2\rho r + 2\rho^2 r^2 + \cdots + 2\rho^{n-1} r^{n-1}\right]$$

$$= \frac{\sigma_u^2}{\Sigma x_t^2}\left[1 + 2\rho r\left(1 + \rho r + \rho^2 r^2 + \cdots + \rho^{n-2} r^{n-2}\right)\right]$$

As $n \to \infty$,

$$= \frac{\sigma_u^2}{\Sigma x_t^2}\left[1 + \frac{2\rho r}{1 - \rho r}\right]$$

$$= \frac{\sigma_u^2}{\Sigma x_t^2}\left[\frac{1 + \rho r + 2\rho r}{1 - \rho r}\right]$$

$$\therefore \text{Var}(\hat{\beta}_2)_{AR_1} = \frac{\sigma_u^2}{\sum x_t^2}\left[\frac{1+\rho\delta}{1-\rho\delta}\right] \quad \ldots (2.14)$$

$$\text{i.e. } \text{Var}(\hat{\beta}_2)_{AR_1} = \text{Var}(\hat{\beta}_2)_{OLS}\left(\frac{1+\rho\delta}{1-\rho\delta}\right)$$

$$\ldots (2.15)$$

If for example, $\rho = 0.6$ & $\delta = 0.8$, Using eqn (2.14), we can check that,

$$\text{Var}(\hat{\beta}_2)_{AR_1} = 2.8461\, \text{Var}(\hat{\beta}_2)_{OLS}$$

$$\text{i.e. } \frac{1}{2.8461}\text{Var}(\hat{\beta}_2)_{AR_1} = \text{Var}(\hat{\beta}_2)_{OLS}$$

$$0.3513\, \text{Var}(\hat{\beta}_2)_{AR_1} = \text{Var}(\hat{\beta}_2)_{OLS}$$

That is usual OLS formula will underestimate the variance of $(\hat{\beta}_2)_{AR_1}$ by amount 65 percent As you will realize, this answer is specific for the given values of $\rho$ and $\delta$.

In presence of autocorrelation the OLS estimator $\hat{\beta}_2$ is still linear and unbiased but not BLUE. In short, $\hat{\beta}_2$ although linear-unbiased, is not efficient (i.e. it does not have minimum variance).

In the case of autocorrelation we find an estimator, that is BLUE in the following section.

## The BLUE estimator in presence of Autocorrelation:

Countinuing with the two-variable model and assuming the AR(1) process, we can show that the BLUE estimator of $\beta_2$ is given by the following expression,

$$\hat{\beta}_2^{GLS} = \frac{\sum_{t=2}^{n} (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n} (x_t - \rho x_{t-1})^2} + c \quad \cdots (2.16)$$

Where c is a correction factor that may be disregarded in practice. Note that the subscript t now runs from t=2 to t=n. and its variance is given by,

$$Var(\beta_2)^{GLS} = \frac{\sigma_u^2}{\sum_{t=2}^{n} (x_t - \rho x_{t-1})^2} + D. \quad \cdots (2.17)$$

Where D is a correction factor that may also be disregarded in practice.

The GLS estimator of $\beta_2$ given in eqn (2.16) incorporates the autocorrelation parameter $\rho$ in the estimating formula, whereas the OLS formula given in eqn (2.11) simply neglects it. Intuitively, this is the reason why the GLS estimator is BLUE and not the OLS Estimator.

The GLS estimator makes the most use of the available information. If $\rho = 0$, there is no additional information to be considered and hence both GLS and OLS estimators are identical.

In short, under autocorrelation the GLS estimator given in eqn (2.15) that is BLUE and minimum variance is now given by eqn (2.17).

## Detection of Autocorrelation:

### (i) Graphical Method:

The importance of producing and analyzing plots of residuals as a standard part of statistical analysis can not be overemphasized. Besides occasionally providing an easy to understand summary of a complex problem, they allow the simultaneous examination of the data as an aggregate while clearly displaying the behaviour of individual cases.

There are various way of examining the residual. We can simply plot them against time, the time sequence plot.

Alternatively, we can plot the standardized residuals against time. The standardized residuals are simply the residuals ($\hat{u}_t$) divided by the standard error of the regression ($\sqrt{\hat{\sigma}^2}$), that is, they are $(\hat{u}_t / \hat{\sigma})$. Notice that $\hat{u}_t$ and $\hat{\sigma}$ are measured in the units in which the regressand $Y$ is measured. The values of standardized residuals

will therefore the pure numbers and can be compared with the standardized residual of other regressions.

We can plot $\hat{u}_t$ against $\hat{u}_{t-1}$, that is, plot the residuals at time t. against their value at time $(t-1)$, a kind of empirical test of the AR(1) scheme.

The graphical method are powerful and suggestive, is subjective or qualitative in nature. But there are several quantitative tests that one can use to supplement the purely qualitative approach. We now consider some of these tests.

### * (i) The Runs Test:

If we plot time sequence plot, we notice a peculiar feature: ~~Initially~~, we have several residuals that are negative and positive. If these residuals were purely random, could we observe such a pattern? Intuitively, it seems unlikely. This intuition can be checked by the so-called runs test, sometimes also known as the Geary test, a nonparametric test.

(−−−−−−−−−)(+++++++++++++++++++++)(−−−−−−−

... (12.18)

Thus there are 9 negative residuals, followed by 21 positive residuals, followed by 10 negative residuals, for a total of 40 observations.

We now define a run as an uninterrupted sequence of one symbol or attribute, such as + or -. We further define the length of a run as the number of elements in it. In the sequence shown eqn (2.18) there are 3 runs; a run of 9 minuses (i.e. of length 9), a run of 21 pluses (i.e. of length 21) and a run of 10 minuses (i.e. of length 10).

By examining how runs behave in a strictly random sequence of observations, one can derive a test of randomness of runs. If there are too many runs, it would mean that in our example the residuals change sign frequently, thus indicating negative serial correlation. similarly, if there are too few runs, we may suggest positive autocorrelation. In ~~fig~~ eqn (2.18) would indicate positive correlation in the residual.

Now Let,

N = total number of observations
i.e. $N = N_1 + N_2$

$N_1$ = Number of positive symbols (i.e. + residuals)

$N_2$ = Number of Negative symbols (i.e. - residuals)

R = Number of runs.

$H_0$: Successive outcomes (Residuals) are independent.

$H_1$: Successive outcomes (Residuals) are dependent.

Assuming that $N_1 > 10$ and $N_2 > 10$, the number of runs is (asymptotically) normally distributed with

Mean: $$E(R) = \frac{2N_1N_2}{N} + 1$$

& Variance: $$V(R) = \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}$$

... (2.19)

If the null Hypothesis of randomness is sustainable, following the properties of normal distribution, we should expect that.

$$P[E(R) - 1.96\sigma_R \leq R \leq E(R) + 1.96\sigma_R] = 0.95$$

... (2.20)

i.e. The Probability is 95% that the preceding interval will include R. Therefore, we have this Rule:

Decision Rule:

Do not Reject the $H_0$ of randomness with 95% confidence if R the number of runs, lies in the preceding confidence interval; Reject the $H_0$ if R lies outside these limits.

In our example, $N_1 = 21$ & $N_2 = 19$, $N = 40$, & $R = 3$.

$$E(R) = \frac{2N_1N_2}{N} + 1 = \frac{2(21)(19)}{40} + 1$$

$$E(R) = 10.975 \quad \ldots (2.21)$$

$$\sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}$$

$$= \frac{2(21)(19)(2(21)(19) - 40)}{(40)^2(40-1)}$$

$$\sigma_R^2 = 9.6936$$

$$\therefore \sigma_R = 3.1134 \quad \ldots (2.22)$$

The 95% confidence interval for R is,

$$[10.975 \pm 1.96(3.1134)] = [4.8728, 17.0722]$$

Hence by using decision rule, we can conclude that, this interval does not include 3. Thus, we can reject the $H_0$ i.e. successive residuals are dependent i.e. The residuals in regression are random with 95% confidence.]

In other words, Accept $H_1$. i.e. the successive residuals are dependent. (i.e. the residuals exhibit autocorrelation.

**(iii) Durbin - Watson d Test:

The most celebrated test for detecting serial correlation is that developed by statisticians Durbin and Watson. It is popularly known as Durbin - Watson d statistics, which is defined as,

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \quad \cdots (2.23)$$

Which is simply the ratio of the sum of squared differences in successive residuals to the RSS. Note that in the numerator of the d statistic the number of observations is n-1 because one observation is lost in taking successive differences.

A great advantage of the d statistic is that it is based on the estimated residuals, which are computed in regression analysis. Because of this advantage, it is now a common practice to report the Durbin-Watson d along with summary measures, such as $r^2$, adjusted $R^2$, t, and F.

Assumptions of d-statistics:

(i) The regression model includes the intercept term. If it is not present, as in the case of the regression through the origin, it is essential to rerun the regression including the intercept term to obtain the RSS.

(ii) The explanatory variables, the X's, are non-stochastic, or fixed in repeated sampling.

(iii) The disturbances $u_t$ are generated

by the first-order autoregressive scheme, $u_t = \rho u_{t-1} + \varepsilon_t$. Therefore, it can not be used to detect higher-order autoregressive scheme.

(iv) The error term $u_t$ is assumed to be normally distributed.

(v) The regression model does not include the lagged value(s) of the dependent variable as one of the explanatory variable.

Thus, the test is inapplicable in models of the following type,

$$Y_t = B_1 + B_2 X_{2t} + B_3 X_{3t} + \cdots + B_K X_{Kt} + Y_{t-1} \quad \cdots (2.24)$$

where $Y_{t-1}$ is the one period lagged value of $Y$.

(vi) There are no missing observations in the data.

Test Procedure:

From eqn (2.23),

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \quad \cdots (2.25)$$

since $\sum \hat{u}_t^2$ and $\sum \hat{u}_{t-1}^2$ differ in only one observation, they are approximately equal.

Therefore, setting $\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_{t-1}^2$, Eq$^n$ (2.25) may be written as,

$$d \approx \frac{\sum \hat{u}_t^2 + \sum \hat{u}_t^2 - 2 \sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

$$\approx \frac{2 \sum \hat{u}_t^2 - 2 \sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

$$\approx 2 \left( \frac{\sum \hat{u}_t^2 - \sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \right)$$

$$d \approx 2 \left( 1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \right) \quad \text{---- (2.26)}$$

Now, let us define

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \quad \text{---- (2.27)}$$

as the sample first-order coefficient of autocorrelation, an estimator of $\rho$.

Using (2.27), we can express eq$^n$ (2.26) as

$$d \approx 2(1 - \hat{\rho}) \quad \text{---- (2.28)}$$

Since $-1 \leq \rho \leq 1$, eq$^n$ (2.28) implies that

$$0 \leq d \leq 4. \quad \text{---- (2.29)}$$

These are bounds of $d$, any estimated $d$ value must lie within these limits.

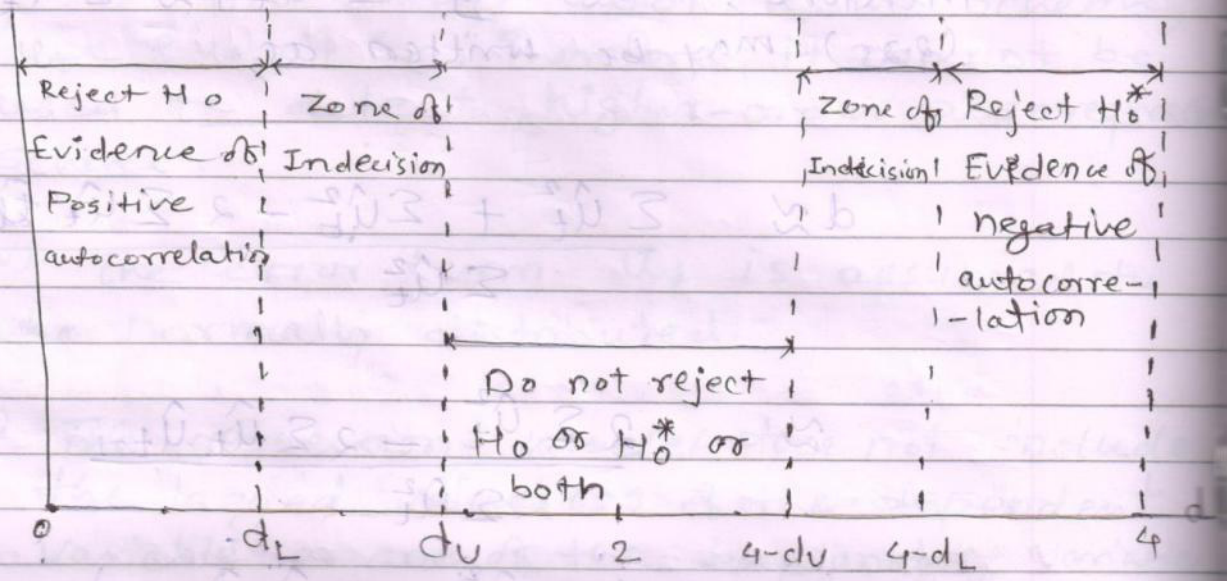



Figure 2.1. Durbin - Watson d statistic

$H_0$: No positive autocorrelation.

$H_0^*$: No negative autocorrelation.

It is apparent from eqn (2.28) that if $\hat{\rho} = 0$, $d = 2$; that is, if there is no serial correlation (of the first-order), d is expected to be about 2.

Therefore, as a rule of thumb, if d is found to be 2 in an application, one may assume that there is no first order autocorrelation, either positive or negative.

If $\hat{\rho} = \pm 1$, indicating perfect positive correlation in the residuals, $d \approx 0$. Therefore, the closer d is to 0, the greater the evidence of positive serial correlation.

If $\hat{\rho} = -1$, that is, there is perfect negative correlation among the successive residuals $d \approx 4$. Hence, the closer d is to 4, the greater

the evidence of negative serial correlation.

∴ Mechanics of Durbin-Watson test;

However, Durbin and watson were successful in deriving a lower bound $d_L$ and an upper bound $d_U$ such that if the computed d from eqn (2-23) lies outside these critical values, a decision can be made regarding the presence of positive or negative serial correlation. Moreover, these limits depends only on the number of observations n and the number of explanatory variables and do not depend on the values taken by these explanatory variables. These limits, for n going from 6 to 200 and up to 20 explanatory variables, have been seen in Durbin-watson d statistic table.

Mechanics of Durbin Watson test:

The mechanics of the durbin-Watson test are as follows, assuming that the assumptions underlying the test are fulfilled.

1) Run the OLS regression and obtain the residuals.
2) Compute d.
3) For the given sample size and given no. of explanatory variables, find out the critical $d_L$ and $d_U$ values.
4) Now follow the decision rules given in table (2.1). For ease of reference, these decision rules are also depicted in fig(2.1)

~~fig (2.1)~~

Table (2.1) Durbin-Watson d ~~statistic~~ 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 | $d_U < d < 4-d_U$ |

## Remedial Measure:

(What to do when you find autocorrelation)

If after applying one or more diagnostic tests of autocorrelation, we find that is autocorrelation then we have four options,

(i) Try to find out if the autocorrelation is pure autocorrelation and not the result of mis-specification of the model. i.e. it has excluded some important variable or because its functional form is incorrect.

(ii) If it is pure autocorrelation, one can use appropriate transformation of the original model so that in the transformated model we do not have the problem of (pure) autocorrelation.

(iii) In large sample, we can use the Newey-West method to obtain standard errors of OLS estimators that are corrected for

autocorrelation.

(iv) In some situations we can continue to use the OLS method.

## Correcting for (pure) autocorrelation!

Knowing the consequences of autocorrelation especially the lack of efficiency of OLS estimators, we may need to remedy the problem. The remedy depends on the knowledge one has about the nature of interdependence among the disturbances, that is, knowledge about the structure of autocorrelation.

Consider two variable model;

$$Y_t = B_1 + B_2 X_t + u_t \quad \text{---(2.30)}$$

and assume that the error term follows AR(1) scheme, namely,

$$u_t = \rho u_{t-1} + \epsilon_t, \quad -1 < \rho < 1 \quad \text{---(2.31)}$$

Now consider two cases (1) $\rho$ is known, (2) $\rho$ is not known but has to be estimated.

(i) when $\rho$ is known:

If the coefficient of first-order autocorrelation is known, then Problem of autocorrelation can be easily solved. If eqn (2.30) holds true at time $t$, it also holds true at time $(t-1)$. Hence

$$Y_{t-1} = B_1 + B_2 X_{t-1} + u_{t-1} \quad \text{---(2.32)}$$

Multiplying eqn (2.32) by $\rho$ on both sides, we obtain

$$\rho Y_{t-1} = \rho B_1 + \rho B_2 X_{t-1} + \rho u_{t-1} \quad \text{---(2.33)}$$

subtracting (2.33) from (2.30), we get

$$(Y_t - \rho Y_{t-1}) = B_1(1-\rho) + B_2(X_t - \rho X_{t-1}) + \epsilon_t \quad \text{---(2.34)}$$

Where $\epsilon_t = (u_t - \rho u_{t-1})$

we can express eqn (2.34) as,

$$Y_t^* = B_1^* + B_2^* X_t^* + \epsilon_t \quad \text{---(2.35)}$$

Where,

$$Y_t^* = (Y_t - \rho Y_{t-1})$$

$$B_1^* = B_1(1-\rho)$$

$$X_t^* = (X_t - \rho X_{t-1})$$

& $B_2^* = B_2$

Since the error term in eqn (2.35) satisfies the usual OLS assumptions, we can apply OLS to transformed variables $Y^*$ and $X^*$ and obtain estimators with all the optimum properties, namely, BLUE.

In effect, running eqn (2.35) is to using generalized Least squares (GLS) where GLS is nothing but OLS applied to the

transformed model that satisfies the classical assumptions.

Regression (2.34) is known as the generalized or quasi, difference equation. It involves regressing Y on X, not in the original form, but in the difference form, which is obtained by subtracting a proportion ($= \rho$) of the value of a variable in the previous time period from its value in the current time period.

(2) When $\rho$ is not known:

(#) The First difference method:

Since $\rho$ lies between 0 and $\pm 1$, one could start from two extreme positions. At one extreme, one could assume that $\rho = 0$, i.e. no (first order) serial correlation and at other extreme we could let $\rho = \pm 1$, that is, perfect positive or negative correlation.

As a matter of fact, when a regression is run, one generally assumes that there is no autocorrelation and then lets the Durbin–Watson or other test show whether this assumption is justified.

If, however, $\rho = +1$, the generalized difference equation (2.34) reduces to the first-difference equation.

$$Y_t - Y_{t-1} = \beta_2 (X_t - X_{t-1}) + (u_t - u_{t-1})$$

or

$$\Delta Y_t = \beta_2 \Delta X_t + \epsilon_t \quad \text{---(2.36)}$$

where $\Delta$ is the first difference operator.

since the error term in eqn (2.36) is free from serial correlation. To run the regression eqn (2.36) all one has to do is form the first differences of both the regressand and regressor(s) and run the regression on these first differences.

The first difference transformation may be appropriate if the coefficient of autocorrelation is very high, say in excess of 0.8, or the Durbin-Watson d is quite low.

Maddala has proposed this rough rule of thumb: Use the first difference form whenever $d < R^2$.

(iii) **Based on Durbin-Watson d Statistics:**

If we cannot use the first difference transformation because $\rho$ is not sufficiently close to unity, we have any easy method of estimating it from the relationship between d and $\rho$ established previously in eqn (2.28), from which we can estimate $\rho$ as follows:

$$\hat{\rho} \approx 1 - \frac{d}{2} \qquad \text{(2.37)}$$

Thus, in reasonably large samples one can obtain $\rho$ from eqn (2.37) and use it to transform the data as shown in the generalized difference equation (2.34).

Keep in mind that the relationship between $\rho$ and d given in eqn (2.37)

may not hold true in small samples, for which Theil and Nagar have proposed a modification, which is given below,

$$\hat{\rho} = \frac{n^2(1-\frac{d}{2}) + k^2}{n^2 - k^2} \quad \ldots (2.38)$$

Where $n$ = total number of observations, $d$ = Durbin-Watson d, and $k$ = no. of coefficients including intercept to be estimated.

### (iii) $\rho$ Estimated from residuals:

If the AR(1) scheme $u_t = \rho u_{t-1} + \varepsilon_t$ is valid, a simple way to estimate $\rho$ is to regress the residuals $\hat{u}_t$ on $\hat{u}_{t-1}$, for the $\hat{u}_t$ are consistent estimators of the true $u_t$, as noted previously. i.e. we run the following regression,

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t \quad \ldots (2.39)$$

Where $\hat{u}_t$ are the residuals obtained from the original regression.

### (iv) $\rho$ estimation from Cochrane-Orcutt iterative procedure:

It is an alternative method, for estimating $\rho$ from Durbin-Watson d-statistic. This method is frequently used in practice and it uses the estimated residual $u_t$.

To obtain information about unknown $\rho$. explain this method consider a two variable model,

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad \text{and}$$

Assume that $u_t$ generated by AR(1) scheme

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

Cochrane and Orcutt recommended following step to estimate $\rho$.-

Step-1:

Estimate two variable model by standard OLS method and obtain residual $u_t$.

Step-2:

Using the estimated residual run the following regression,

$$\hat{u}_t = \hat{\rho} \hat{u}_{t-1} + v_t$$

step-3:

Using $\hat{\rho}$ thus obtained, Run OLS to,

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + u_t^*$$

Where $\beta_1^* = \beta_1 (1 - \hat{\rho})$

$$Y_t^* = (Y_t - \hat{\rho} Y_{t-1})$$

$$\beta_2^* = \beta_2$$

$$X_t^* = (X_t - \hat{\rho} X_{t-1})$$

$$u_t^* = (u_t - \hat{\rho} u_{t-1})$$

Step-4:

Since apriory it is known that $\hat{\rho}$ is the best estimator of $\rho$, substitute the value of $\beta_1^*$ and $\beta_2^*$ in original regression and obtain the new residuals say $u_t'^*$

$$u_t'^* = Y_t - \hat{\beta}_1^* - \beta_2^* X_t \quad \cdots (2.40)$$

fix that, $e_t'^* = \hat{\rho}\, e_{t-1}'^* + u_t^*$

Thus, $\hat{\rho}$ is 2nd round estimate of $\rho$. Since we do not know, whether 2nd round estimate of $\rho$ is the best estimator of $\rho$, we can go for 3rd round estimate of $\rho$ and so on. In general we stop the iterative procedure when the sucresive estimator of $\rho$ differ by a very small quantity.

(v) $\rho$ estimating from Cocharane-Orcutt two step procedure:

In step-1 we estimate $\rho$ from the 1st iteration and in step-2 we use testimater of $\rho$ to run the generalize difference equation. Some time in practice these two step method gives result quit similar to those obtained from the iterative procedure.

(vi) $\rho$ estimation from Durbin two-step procedure:

To Illustrate this procedure let us write the generalize difference eq$^n$.

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

$$\therefore Y_t = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \rho Y_{t-1} + v_t \quad \text{---- (2.4)}$$

**Step-I:**

Let eqn (2.4) as multiple regression model, Regression $Y_t$ on $X_t$, $X_{t-1}$ & $Y_{t-1}$ and treat estimated value of the regression coefficient $Y_{t-1}$ as estimate of $\rho$. Although this estimator is biase but it provides the consistent estimator of $\rho$.

**Step-II:**

One obtain $\hat{\rho}$, transform the variables as,

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1}$$

$$X_t^* = X_t - \hat{\rho} X_{t-1}$$

And Run the OLS Regression on the transform variable.

**Comparision of Methods:**

Which method of estimating of $\rho$ one should use in practice to run generalized difference equation. If we are deling with large sample ($n > 60$), all these methods generally yield similar estimates of $\rho$ but in small samples they perform differently. In practice the Cochrane-Orcutt method become quite popular

## * Multicollinearity:

Definition:

"The existence of a "perfect" or exact linear relationship among some or all explanatory variables of a regression model is called multicollinearity."

### Types of Multicollinearity:

(i) Exact or Perfect multicollinearity:

For the k- variable regression involving explanatory variable $X_1, X_2, \cdots, X_k$ (where $X_1 = 1$ for all observations to allow for the intercept term), an exact linear relationship is said to exist if the following condition is satisfied;

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0 \quad \cdots (2.42)$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are constants such that not all of them zero simultaneously.

(ii) less extreme or less than perfect Multico-linearity:

For a k-variable regression involving explanatory variable $X_1, X_2, \cdots, X_k$, an less perfect relationship is said to exist if the following condition is satisfied:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k + v_i = 0 \quad \cdots (2.43)$$

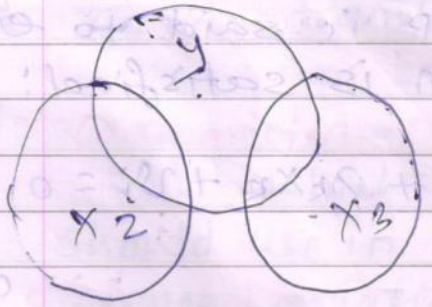
where $v_i$ is a stochastic error term.

As an numerical example, consider a following hypothetical data:

| $X_2$ | $X_3$ | $X_3^*$ |
|---|---|---|
| 10 | 50 | 52 |
| 15 | 75 | 75 |
| 18 | 90 | 97 |
| 24 | 120 | 129 |
| 30 | 150 | 152 |

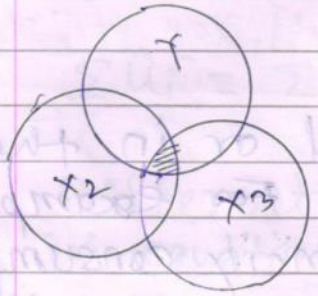
It is apparent that, $X_{3i} = 5X_{2i}$. Therefore, there is perfect collinearity bet$^n$ $X_2$ and $X_3$. The variable $X_3^*$ was created from $X_3$ by simply adding to it the following numbers, which were taken from a table of random numbers; 2, 0, 7, 9, 2. Now there is no longer perfect collinearity between $X_2$ and $X_3^*$. However, the two variables are highly correlated because calculations will show that the coefficient of correlation between them is 0.9959.

## ** The Ballentine view of multicollinearity:

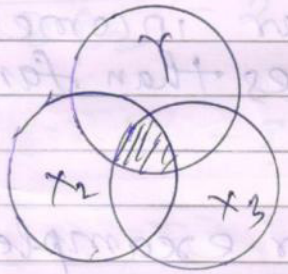



(a) No multicollinearity exists $X_2$ and $X_3$

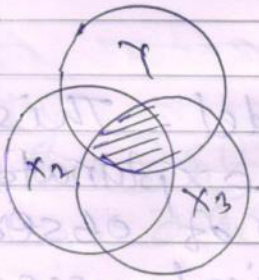
(b) Low multicollinearity exists between $X_2$ & $X_3$



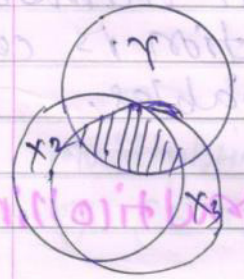

(c) Moderate multicollinearity exists between $X_2$ & $X_3$




(d) High Multicollinearity exists between $X_2$ and $X_3$




(e) Very high multicollinearity exists between $X_2$ and $X_3$




## Sources of Multicollinearity:

There are several sources of multicollinearity. As Montgomery and Peck note, multicollinearity may be due to the following factors:

i) The data collection method employed, for example, sampling over a limited range of

the values taken by the regressors in the population.

2) Constraints on the model or in the population being sampled. For example, in the regression of electricity consumption on income ($X_2$) and house size ($X_3$) there is a physical constraint in the population in that families with higher incomes generally have larger homes than families with lower incomes.

3) Model specification, for example, adding polynomial terms to a regression model, especially when the range of the X-variable small.

4) An overdetermined model. This happens when the model has more explanatory variables than the number of observations. This could happen in medical research where there may be a small number of patients about whom information is collected on a large number of variables.

* Estimation in the presence of multicollinearity

We consider three variable regression model. Using the deviation form as,

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \quad \ldots (2.44)$$

$$\hat{u}_i = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i} \quad \ldots (2.45)$$

$$\hat{u}_i^2 = (y_i - \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i})^2$$

$$\Sigma \hat{u}_i^2 = \Sigma (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})^2 \quad \cdots (2.46)$$

Differentiating eqⁿ (2.46) w.r.t. $\hat{\beta}_2$ & $\hat{\beta}_3$, we get & equating with zero, we get.

$$\frac{\partial \Sigma \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \Sigma (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})(-x_{2i}) = 0$$

$$\therefore -\Sigma y_i x_{2i} + \hat{\beta}_2 \Sigma x_{2i}^2 + \hat{\beta}_3 \Sigma x_{3i} x_{2i} = 0$$

$$\therefore \Sigma y_i x_{2i} = \hat{\beta}_2 \Sigma x_{2i}^2 + \hat{\beta}_3 \Sigma x_{3i} x_{2i}$$

$$\cdots (2.47)$$

Now $$\frac{\partial \Sigma \hat{u}_i^2}{\partial \hat{\beta}_3} = 2 \Sigma (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})(-x_{3i}) = 0$$

$$\therefore \Sigma y_i x_{3i} = \hat{\beta}_2 \Sigma x_{2i} x_{3i} + \hat{\beta}_3 \Sigma x_{3i}^2 \quad \cdots (2.48)$$

Multiplying eqⁿ (2.47) by $(\Sigma x_{3i}^2)$ & eqⁿ (2.48) by $(\Sigma x_{3i} x_{2i})$, we get.

$$(\Sigma y_i x_{2i})(\Sigma x_{3i}^2) = \hat{\beta}_2 (\Sigma x_{2i}^2)(\Sigma x_{3i}^2) + \hat{\beta}_3 (\Sigma x_{3i} x_{2i}) \cdot (\Sigma x_{3i}^2)$$

$$(\Sigma y_i x_{3i})(\Sigma x_{3i} x_{2i}) = \hat{\beta}_2 (\Sigma x_{2i} x_{3i})(\Sigma x_{3i} x_{2i}) + \hat{\beta}_3 (\Sigma x_{3i}^2) \cdot (\Sigma x_{3i} x_{2i})$$

$$(\Sigma y_i x_{2i})(\Sigma x_{3i}^2) - (\Sigma y_i x_{3i})(\Sigma x_{3i} x_{2i}) = \hat{\beta}_2 \left[ (\Sigma x_{2i}^2)(\Sigma x_{3i}^2) - (\Sigma x_{2i} x_{3i})(\Sigma x_{3i} x_{2i}) \right]$$

$$\therefore \hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad \cdots (2.49)$$

similarly,

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad \cdots (2.50)$$

Which gives the OLS estimators of the population partial regression coefficients $\beta_2$ and $\beta_3$, resp. & the variances is given by the following formulas

$$\text{Var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \sigma^2 \quad \cdots (2.51)$$

or equivalently,

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \quad \cdots (2.52)$$

since $r_{23}$ is the sample coefficient of correlation between $x_2$ & $x_3$.

$$\& \quad V(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad \cdots (2.53)$$

or equivalently

$$\text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)} \quad \ldots (2.54)$$

## Case 1: Perfect Multicollinearity

Assume that, $X_{3i} = \lambda X_{2i}$ where $\lambda$ is a nonzero constant. Substituting this into eqn (2.59), we obtain

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2) - (\lambda \sum y_i x_{2i})(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2) - \lambda^2 (\sum x_{2i}^2)^2}$$

$$= \frac{0}{0} \quad \ldots (2.55)$$

Which is an indeterminate expression. Similarly, $\hat{\beta}_3$ is also indeterminate.

& $$\text{Var}(\hat{\beta}_2) = \frac{\lambda^2 \sum x_{2i}^2 \cdot \sigma^2}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2) - (\lambda \sum x_{2i} x_{3i})^2}$$

$$= \frac{\lambda^2 \sum x_{2i}^2 \sigma^2}{0}$$

$$= \infty$$

Ily $\text{Var}(\hat{\beta}_3) = \infty$

Hence the $\text{var}(\hat{\beta}_2)$ & $\text{Var}(\hat{\beta}_3)$ are infinite.

Therefore, in the case of perfect multicolli-
-nearity $\hat{\beta}_2$ & $\hat{\beta}_3$ have indeterminate expression & $Var(\hat{\beta}_2)$ & $Var(\hat{\beta}_3)$ are infinite.

Case-2: Less than perfect multicollinearity

Consider $x_{3i} = \lambda x_{2i} + v_i$ - - (2.56)

where $\lambda \neq 0$ and $v_i$ is a stochastic error term such that $\sum x_{2i} v_i = 0$.

In this case, estimation of regression coefficients $\beta_2$ and $\beta_3$ may be possible. Substituting eq (2.56) into (2.49), we get

$$\hat{\beta}_2 = \frac{\sum (y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum y_i x_{2i} + \sum y_i v_i)(\lambda \sum x_{2i}^2)}{\sum (x_{2i}^2)(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum x_{2i}^2)^2} \quad \ldots (2.57)$$

Where use is made of $\sum x_{2i} v_i = 0$.
a similar expression can be derived for $\hat{\beta}_3$.

Note: If $v_i$ is sufficiently small, say, very close to zero eqn (2.56) will indicate almost perfect collinearity and we shall be back to the indeterminate case eqn (2.55).

## Consequences of multicollinearity:

1) In case of perfect multicollinearity the OLS estimators are indeterminate and also their variances are infinite.

2) In case of near or high multicollinearity, the following are the consequences.

(i) Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult.
(When $r_{23} = 1$ then variances are infinite)

(ii) Because of consequence (i), the confidence interval tend to be much wider, leading to the acceptance of zero null Hypothesis. (i.e. the true population coefficient is zero).

(iii) Because of consequence (i), t-ratio of one or more coefficients tends to statistically insignificant.

(iv) Although the t ratio of one or more coefficients is statistically insignificant, $R^2$, the overall measure of goodness of fit, can be very high.

(v) The OLS estimators and their standard errors can be sensitive to small change in data.

## Detection of Multicollinearity

Having studied the nature and consequences of multicollinearity, the natural question is: How does one know that collinearity is present in any given situation, especially in models involving more than two explanatory variables. Here it is useful to bear in mind Kmenta's warning:

(i) Multicollinearity is a question of degree and not a kind. The meaningful distinction is not between the presence and the absence of multicollinearity, but between the its various degrees.

(ii) Since multicollinearity refers to the condition of the explanatory variables that are assumed to be nonstochastic, it is a feature of the sample and not of the population.

Therefore, we do not "test for multicollinearity" but can, if we wish, measure its degree in any particular sample.

Since multicollinearity is essentially a sample phenomenon, arising out of the largely nonexperimental data collected in most social sciences, we do not have one unique method of detecting it or measuring its strength. What we have are some rules of thumb, some informal and some formal, but rules of thumb all the same. We now consider some of these rules.

1. High $R^2$ but few significant t ratios:

As noted, this is the classic symptom of multicollinearity. If $R^2$ is high, say, in excess of 0.8, the F test in most cases will reject the hypothesis that the partial slope coefficients are simultaneously equal to zero, but the individual t tests will show that none or very few of the partial slope coefficients are statistically different from zero.

Although this diagnostic is sensible, its disadvantage is that "it is too strong in the sense that multicollinearity is considered as harmful only when all of the influences of the explanatory variables on Y cannot be disentangled.

2. High pair-wise correlations among regressors:

If the pair-wise or zero-order correlation coefficient between two regressors is high, say, in excess of 0.8, then multicollinearity is a serious problem. The problem with this criterion is that, although high zero-order correlations may suggest collinearity, it is not necessary that they be high to have collinearity in any specific case.

To put the matter somewhat technically, high zero-order correlations are a sufficient but not a necessary condition for the existence of multicollinearity because it can exist even though the zero-order or simple correlations are comparatively low (say, less than 0.50). To see this relationship, suppose we have a four variable model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

and suppose that,

$$x_{4i} = \lambda_2 x_{2i} + \lambda_3 x_{3i},$$

where $\lambda_2$ and $\lambda_3$ are constants, not both zero obviously, $x_4$ is an exact linear combination of $x_2$ and $x_3$, giving $R^2_{4.23} = 1$, the coefficient of determination in the regression of $x_4$ on $x_2$ and $x_3$.

$$R^2_{4.23} = \frac{r^2_{42} + r^2_{43} - 2 r_{42} r_{43} r_{23}}{1 - r^2_{23}} \quad \cdots (2.58)$$

But since $R^2_{4.23} = 1$ because of perfect collinearity, we obtain

$$1 = \frac{r^2_{42} + r^2_{43} - 2 r_{42} r_{43} r_{23}}{1 - r^2_{23}} \quad \cdots (2.59)$$

It is not difficult to see that eqn (2.59) is satisfied by $r_{42} = 0.5$, $r_{43} = 0.5$ and $r_{23} = -0.5$. which are not very high values.

Therefore, in models involving more than two explanatory variables, the simple or zero order correlation will not provide an infallible guide to the presence of multicollinearity. Of course, if there are only two explanatory variables, the zero order correlation will suffice.

3. Examination of partial correlations:
Because of the problem just mentioned in relying on zero-order correlations, farrar and Glauber have suggested that one should look at the partial correlation coefficients. Thus, in the regression of Y on $X_2$, $X_3$ and $X_4$, a finding that $R^2_{1.234}$ is very high but $r^2_{12.34}$, $r^2_{13.24}$ and $r^2_{14.23}$ are comparatively low may suggest that the variables $X_2$, $X_3$, and $X_4$ are highly intercorrelated and that at least one of these variables is superfluous.

4. Auxiliary regressions:
Since multicollinearity arises because one or more of the regressors are exact or approximately linear combinations of other regressors, one way of finding out which X variable is related to other X variables is to regress each $X_i$ on the remaining X variables and compute the corresponding $R^2$, which we designate as $R^2_i$; each one of these regressions is called an auxiliary regression, auxiliary to the main regression of Y on the X's. Then the following the relationship between F and $R^2$, the variable follows the F distribution with $(k-2)$ and $(n-k+1)$ d.f.

$$F_i = \frac{R^2_{x_i.x_2x_3\cdots x_k} / (k-2)}{(1 - R^2_{x_i.x_2x_3\cdots x_k}) / (n-k+1)} \quad \cdots\cdots (2.60)$$

In eqn (2.60) n stands for the sample size, k stands for the number of explanatory variables including the intercept term and $R^2_{x_i.x_2x_3\cdots x_k}$ is the coefficient of determination in the regression of variable $x_i$ on the remaining x variables.

If the computed f exceeds the critical $F_i$ at the chosen level of significance, it is taken to mean that the particular $x_i$ is collinear with other x's; if it does not exceed the critical $F_i$, we say that ~~variable~~ it is not collinear with other x's, in which case we may retain that variable in the model. If $F_i$ is statistically significant, we will still have to decide whether the particular $x_i$ should be dropped from the model.

Note:

Istead of formally testing all auxillian $R^2$ values, one may adopt Klien's rule of thumb, which suggests that multicollinearity may be a troublesome problem only if the $R^2$ obtained from an auxiliary regression is greater than the overall $R^2$.

## 5. Eigenvalues and condition index:

We can derive eigen values of $X'X$ and the condition number k defined as,

$$k = \frac{\text{Maximum eigenvalues}}{\text{Minimum eigenvalues}} \quad \cdots (2.61)$$

and the condition index (CI) defined as

$$CI = \sqrt{\frac{\text{Maximum eigenvalue}}{\text{Minimum eigenvalue}}} = \sqrt{k} \quad \ldots (2\text{-}62)$$

Then we have this rule of thumb, If k is between 100 and 1000 there is moderate to strong multicollinearity and if it exceeds 1000 there is severe multicollinearity.

Alternatively, if CI ($=\sqrt{k}$) is between 10 and 30, there is moderate to strong multicollinearity and if it exceeds 30 there is severe multicollinearity.

### 6. Tolerance and Variance inflation factor:

The speed with which variances and covariances increase can be seen with the variance-inflating factor (VIF), which is defined as

$$VIF = \frac{1}{1 - r_{23}^2} \quad \ldots (2\text{-}63)$$

VIF shows how the variance of an estimator is inflated by the presence of multicollinearity. As $r_{23}^2$ approaches to 1, the VIF approaches infinity. That is as the extent of collinearity increases, the variance of an estimator increases, and the limit it can become infinite. As can be seen, if there is no collinearity between $X_2$ and $X_3$, VIF will be 1.

Thus, for the k-variable model, the the variance of the partial regression coefficients,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \quad \cdots (2.64)$$

Where $\hat{\beta}_j$ = partial regression coefficient of regressor $X_j$.

$R_j^2 = R^2$ in the regression of $X_j$ on the remaining $(k-2)$ regressions

[Note: There are $(k-1)$ regressors in the k-variable regression model]

$$\sum x_j^2 = \sum (X_j - \bar{X}_j)^2$$

Eqn (2.64) can be also write,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2} VIF_j \quad \cdots (2.65)$$

Tolerance:

The inverse of the VIF is called tolerance (TOL),

$$TOL_j = \frac{1}{VIF_j} = (1 - R_j^2) \quad \cdots (2.66)$$

When $R_j^2 = 1$ (i.e. Perfect collineanity), $TOL_j = 0$ and when $R_j^2 = 0$ (i.e. no collineanity), $TOL_j = 1$

As a rule of thumb, if the VIF of a variable exceeds 10, which will happen if $R_j^2$ exceeds 0.9, that variable is said

to be highly collinear.

Note: VIF (or tolerance) as a measure of collinearity is not free of criticism. As eqn (2.64) shows that var($\hat{\beta}_j$) depends on three factors: $\sigma^2$, $\sum x_j^2$ and $VIF_j$. A high $VIF_j$ can be counterbalanced by a low $\sigma^2$ or a high $\sum x_j^2$. To put it differently a high VIF is neither necessary nor sufficient to get high variances and high standard errors. Therefore, high multicollinearity, as measured by a high VIF, may not necessarily cause high standard errors. In all this discussion, the terms high and low are used in a relative sense.

Remedial measures:

What can be done if multicollinearity is serious. As in case of detection there is no clearcut method, because multicollinearity is essentialy a sample problem. One can try the following rules of thumb to address the problem of multicollinearity, the success depending on the severity of the collinearity problem.

1. A priori information:

Suppose we consider the model,

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \text{---(2.67)}$$

where Y = consumption, $X_2$ = income, and $X_3$ = wealth. As noted before income and wealth variables tend to be highly collinear.

But suppose a priori we belive that $\beta_3 = 0.10\beta_2$; that is, the rate of change of consumption with respect to wealth is one-tenth the corresponding rate with respect to income. We can then run the following regression,

$$Y_i = \beta_1 + \beta_2 X_{2i} + 0.10\beta_2 X_{3i} + u_i$$
$$= \beta_1 + \beta_2 X_i + u_i \quad \ldots (2.68)$$

Where $X_i = X_{2i} + 0.1 X_{3i}$. One we obtain $\hat{\beta}_2$, we can estimate $\hat{\beta}_3$ from the postulated relationship between $\beta_2$ and $\beta_3$.

2. Combining cross-sectional and time series data:

A variant of the extraneous or a prior information technique is the combination of cross-sectional and time-series data, known as pooling the data.

Suppose we want to study the demand for automobiles in the United States and assume we have time series data on the number of cars sold, average price of the car, and consumer income, suppose also that,

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t \quad \ldots (2.69)$$

where $Y$ = number of cars sold,
$P$ = average price,
$I$ = income
and $t$ = time.

Our objective is to estimate the price elasticity $\beta_2$ and income elasticity $\beta_3$.

In time series data the price and income variables generally tend to be highly collinear. Therefore, if we ~~have cross-sectional~~ run the preceding regression, we shall be faced with the usual multicollinearity problem.

A way out of this has been suggested by Tobin. He says that, if we have cross-sectional data, we can obtain a fairly reliable estimate of the income elasticity $\beta_3$ because in such data, which are at a point in time, the prices do not vary much. Let the cross-sectionally estimated income elasticity be $\hat{\beta}_3$. Using this estimate, we may write the preceding time series regression as,

$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t \quad \text{---(2.40)}$$

where $Y^* = \ln Y - \hat{\beta}_3 \ln I$, that is $Y^*$ represents that value of $Y$ after removing from it the effect of income. We can now obtain an estimate of the price elasticity $\beta_2$ from the preceding regression.

3. Dropping a variable(s) and specification bias:

When faced with severe multicollinearity, one of the "simplest" things to do is to drop one of the collinear variables. Thus, in ~~our~~ we take one example, consumption-income-wealth If we assume that consumption expenditure is linearly related to income and wealth. The data are given in table (2.2).

Table 2.2:
Hypothetical data on consumption expenditure Y, Income $X_2$, AND WEALTH $X_3$.

| Y ($) | $X_2$ ($) | $X_3$ ($) |
|---|---|---|
| 70 | 80 | 810 |
| 65 | 100 | 1009 |
| 90 | 120 | 1273 |
| 95 | 140 | 1425 |
| 110 | 160 | 1633 |
| 115 | 180 | 1876 |
| 120 | 200 | 2052 |
| 140 | 220 | 2201 |
| 155 | 240 | 2435 |
| 150 | 260 | 2686 |

The Regression model for above data (Table 2.2) is,

$$\hat{Y}_i = \hat{B}_1 + \hat{B}_2 X_{2i} + \hat{B}_3 X_{3i}$$

$$\hat{Y}_i = 24.7747 + 0.9415 X_{2i} - 0.0424 X_{3i}$$

$$(6.7525) \quad (0.8229) \quad (0.0807)$$

$$t = (3.6690) \quad (1.1442) \quad (-0.5261)$$

$$R^2 = 0.9635, \quad \bar{R}^2 = 0.9531, \quad df = 7. \qquad (2.71)$$

Regression (2.71) (from $R^2$), shows that income and wealth together explain about 96 percent of the variation in consumption expenditure, and yet neither of the slope coefficients is individually statistically insignifi-

- Cant but also it has the wrong sign. A priori, one would expect a positive relationship between consumption and wealth. Although $\hat{\beta}_2$ and $\hat{\beta}_3$ are individually statistically insignificant, if we test the hypothesis that $\beta_2 = \beta_3 = 0$ simultaneously, this hypothesis can be rejected on the basis of F-value calculated in eqn (2.72).

ANOVA

| Source of Variation | SS | d.f. | MSS |
|---|---|---|---|
| Due to regression | 8565.5541 | 2 | 4282.7770 |
| Due to residual | 324.4459 | 7 | 46.3494 |
| Total | 8890.0000 | 9 | |

$$\therefore F = \frac{4282.7770}{46.3494} = 92.4019 \quad \cdots (2.72)$$

This F-value is obviously highly significant.

Our example, shows dramatically what multicollinearity does. The fact that the F test is significant but the t values of $X_2$ and $X_3$ are individually insignificant means that the two variables are so highly correlated that it is impossible to isolate the individual impact of either income or wealth on consumption. As a matter of fact, if we regress $X_3$ on $X_2$, we obtain:

$$\hat{X}_{3i} = 7.5454 + 10.1909 X_{2i} \quad \cdots (2.73)$$
$$(29.4758) \quad (0.1643)$$

Which shows that there is almost perfect collinearity between $X_3$ and $X_2$.

Now let us see what happens if we regress Y on $X_2$ only.

$$\hat{Y}_i = 24.4545 + 0.5091 X_{2i}$$
$$(6.4138) \quad (0.0357)$$
$$t = (3.8128) \quad (14.2432) \quad \text{---} (2.74)$$
$$R^2 = 0.9621$$

In eqn (2.71) the income variable was statistically insignificant, whereas now it is highly significant.

If instead of regressing Y on $X_2$, we regress it on $X_3$, we obtain.

$$\hat{Y}_i = 24.411 + 0.0498 X_{3i}$$
$$(6.874) \quad (0.0037)$$
$$t = (3.551) \quad (13.29) \quad \text{---} (2.75)$$
$$R^2 = 0.9567$$

Now, we see that wealth has a significant impact on consumption expenditure, where as in (2.71) it had no effect on consumption expenditure.

In above example, If we drop wealth variable, we obtain regression eqn (2.74) which shows that, whereas in the original

model the income variable was statistically insignificant, it is now highly significant.

But in dropping a variable from the model we may be committing a specification bias or specification error. specification bias arises from incorrect specification of model used in the analysis. Thus if economic theory says that income and wealth should both be included in the model explaining the consumption expenditure, dropping the wealth variable would constitute specification Bias.

From the above discussion it is clear that dropping a variable from the model to alleviate the problem of multicollinearity may lead to the specification bias. Hence the remedy may be worse than the disease in some situations because, whereas multicollinearity may prevent precise estimation of the parameters of the model, omitting a variable may seriously mislead us as to the true values of the parameters.

4. Transformation of variables:

Suppose we have time series data on consumption expenditure, income and wealth. One reason for high multicollinearity between income and wealth in such data is that over time both the variables tend to move in the same direction. One way of minimizing this dependence is to proceed as follows.

If the relation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

holds at time t, it must also hold at time (t-1) because of the origin of time is arbitrary anyway. Therefore, we have

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \quad \cdots (2.77)$$

If we substract eqn (2.77) from (2.76), we obtain.

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2,t-1}) + \beta_3 (X_{3t} - X_{3,t-1}) + v_t \quad \cdots (2.78)$$

Where $v_t = u_t - u_{t-1}$. Eqn (2.78) is known as the first difference form because we run the regression, not on the original values but on the difference of successive values of the variables.

The first difference regression model often reduce the severity of multicollinearity, because, although the levels of $X_2$ and $X_3$ may be highly correlated, there is no a priori reason to believe that their differences will also be highly correlated.

Another commonly used transformation in practice is the ratio transformation. Consider the model,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \cdots (2.79)$$

where Y = consumption expenditure,
$X_2$ = GDP
& $X_3$ = Total population.

Since GDP and population grow over time, they are likely to be correlated. One solution to this problem is to express the model on a per capita basis, that is, by dividing eqn (2.76) by $X_{3t}$, to obtain:

$$\frac{Y_t}{X_{3t}} = \beta_1\left(\frac{1}{X_{3t}}\right) + \beta_2\left(\frac{X_{2t}}{X_{3t}}\right) + \beta_3 + \left(\frac{u_t}{X_{3t}}\right) \quad \text{----(2.80)}$$

Such a transformation may reduce collinearity in the original variables.

For instance, the error term $v_t$ in eqn (2.78) may not satisfy one of the assumptions of the CLRM, namely, that the disturbances are serially uncorrelated.

Similarly, in the ratio model eqn (2.80) the error term $(u_t/X_{3t})$ will be heteroscedastic, if the original error term $u_t$ is homoscedastic. Hence, Again the remedy may be worse than the disease of collinearity.

5. Additional or new data.

Since multicollinearity is a sample feature, it is possible that in another sample involving the same variables collinearity may not be so serious as in the first sample. Sometimes simply increasing the size of the sample (if possible) may attenuate the collinearity problem.

For example, in the three-variable model, we saw that

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

Now as the sample size increases, $\sum x_{2i}^2$ will generally increase. Therefore, for any given $r_{23}$, the variance of $\hat{\beta}_2$ will decrease, thus decreasing the standard error, which will enable us to estimate $\beta_2$ more precisely.

6. Reducing collinearity in polynomial regression:

In practice though, it has been found that, if the explanatory variable(s) are expressed in the deviation form (i.e. deviation from the mean value), multicollinearity is substantially reduced. But even then the problem may persist, in which case one may want to consider techniques such as orthogonal polynomials.

7. Other methods of remedying multicolli-
-nearity:

Multivariate statistical techniques such as factor analysis and principal components or techniques such as ridge regression are often employed to solve the problem of multicollinearity.

## * Ridge Regression:

When the method of least square is applied to non-orthogonal data, very poor estimator of the regression coeff-icient can be obtained. In this case the variance of OLS estimators of the regression coefficient may be consideriably inflated and the length of interval of OLS estimator is to long on average. This implies that absolut value of OLS estimates are to large and hence they are unstable. That is there magnitude and sign may change considerably given different sample.

The problem with the method of least square is the requirement that $\hat{\beta}$ be an unbiased estimator of $\beta$. The Gauss - Markov theorem shows us the OLS estimators has minimum variance in the class of linear unbiased estimaters. but there is no gaurantly that this variance will be small. The variance of $\hat{\beta}$ is large. implying that the confidence interval on $\beta$ would be widen and the point estimator of $\hat{\beta}$ is unstable.

One way to elibriate this problem is drop to requirement that the estimator of $\beta$ be unbiased. Suppose that, we can find a biased estimator say $\hat{\beta}^*$ that has the smaller variance than the unbiased estimator $\hat{\beta}$. The MSE($\hat{\beta}^*$) is defined as,

$$MSE(\hat{\beta}^*) = E[(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)']$$

By allowing a small amount of biase in $\hat{\beta}^*$. A number of procedures have been develope or obtaining bias estimator of regression coefficient. One of this procedure is ridge regression proposed by Hoerl and Kennard (1970).

The Ridge estimater is found by slight modification by normal equation. Specially, we define the ridge estimater soln of.

$$\therefore (X'X + \gamma I)\hat{\beta}_R = X'Y$$

$$\therefore \hat{\beta}_R = (X'X + \gamma I)^{-1} X'Y, \quad \gamma > 0$$

Where $\gamma$ is a constant selected by Analyst. It is to note that, when $\gamma = 0$, then ridge estimator is converted in to the OLS estimater.

The ridge estimator is a linear transformation of OLS estimator,

$$\hat{\beta}_R = (X'X + \gamma I)^{-1} X'Y$$

$$= (X'X + \gamma I)^{-1} X'X.\hat{\beta}$$

$$= Z\hat{\beta}$$

Therefore,

$$\therefore E(\hat{\beta}_R) = Z E(\hat{\beta})$$

$$= Z\beta, \quad (\because E(\hat{\beta}) = \beta$$

$$\neq \beta$$

Which shows that, $\hat{\beta}_R$ is a bias estimator of $\beta$. We usually refer the constant $r$ as biasing parameter,

$$\text{Var}(\hat{\beta}_R) = \text{Var}(Z\hat{\beta})$$

$$= Z\,\text{var}(\hat{\beta})\,Z'$$

$$= Z\,\sigma^2 (X'X)^{-1} Z'$$

$$= \sigma^2 (X'X + rI)^{-1} X'X (X'X)^{-1} X'X (X'X + rI)^{-1}$$

$$= \sigma^2 (X'X + rI)^{-1} X'X (X'X + rI)^{-1}$$

A bias of $\hat{\beta}_R$ is,

$$E(\hat{\beta}_R - \beta) = E[Z\hat{\beta} - \beta]$$

$$= Z\,E(\hat{\beta}) - \beta$$

$$= Z\beta - \beta$$

$$= [(X'X + rI)^{-1} X'X - I]\beta$$

$$= [(X'X + rI)^{-1} X'X - (X'X + rI)^{-1}(X'X + rI)]\beta$$

$$= [(X'X + rI)^{-1}(X'X - X'X - rI)]\beta$$

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

$$E(\hat{\beta}_R - \beta) = -r(X'X + rI)^{-1}\beta$$

The MSE of ridge estimator is,

$$MSE(\hat{\beta}_R) = Var(\hat{\beta}_R) + (bias(\hat{\beta}_R))(bias(\hat{\beta}_R))'$$

$$= Var(\hat{\beta}_R) + \gamma^2 (X'X+\gamma I)^{-1} \beta\beta' (X'X+\gamma I)^{-1}$$

$$MSE(\hat{\beta}_R) = \sigma^2 (X'X+\gamma I)^{-1} X'X (X'X+\gamma I)^{-1} + \gamma^2 (X'X+\gamma I)^{-1} \beta\beta' (X'X+\gamma I)^{-1}$$

Taking trace on both side, we get.

$$tr(MSE(\hat{\beta}_R)) = tr\{ \sigma^2 (X'X+\gamma I)^{-1} X'X (X'X+\gamma I)^{-1} + \gamma^2 (X'X+\gamma I)^{-1} \beta\beta' (X'X+\gamma I)^{-1} \}$$

$$= \sum_{j=1}^{k} \left[ \frac{\sigma^2 \lambda_j}{(\lambda_j+\gamma)^2} + \frac{\gamma^2 \beta_j^2}{(\lambda_j+\gamma)^2} \right]$$

$$= \sum_{j=1}^{k} \left[ \frac{\sigma^2 \lambda_j + \gamma^2 \beta_j^2}{(\lambda_j+\gamma)^2} \right]$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are be characterstic root of $X'X$. If $\gamma > 0$ note that bias in $\hat{\beta}_R$ increases with $\gamma$. However, the variance decreases as $\gamma$ increases. In using ridge regression we would like to choose the value of $\gamma$. Such that the reduction in variance term is greater than in increasing the squared bias. If this can be done the MSE of Ridge estimator $\beta_R$ will less than the variance of OLS estimator.

$\gamma \propto$ bias

$\gamma \propto \frac{1}{\text{variance}}$

# Simultaneous-Equation Models

We have seen the concept of single equation models. i.e. models in which there was a single dependent variable Y and one or more explanatory variables, the X's. In such models the emphasis was on estimating and/or predicting the average value of Y conditional upon the fixed values of the X variables. The cause-and-effect relationship, if any, in such models therefore ran from the X's to the Y.

But in many situations, such a one-way or unidirectional cause-and-effect relationship is not meaningful. This occurs if Y is determined by the X's, and some of the X's are, in turn, determined by Y. In short, there is a two way or simultanes relationship between Y and (some of) the X's, which makes the distinction between dependent and explanatory variables of dubious value. It is better to lump together a set of variables that can be determined simultaneously by the remaining set of variables - precisely what is done in simultaneous equation models. In such a models there is more than one equation - one for each of the mutually or jointly, dependent or endogenous variables. And unlike the single equation models, in the simultaneous equation models one may not estimate the parameters of a single equation without taking into account information provided by other equations in the system.

What happens if the parameters of each equation are estimated by applying, say, the method of OLS, disregarding other equation in the system &

We know that, one of the crucial assumptions of the method of OLS is that the explanatory X variables are either non-stochastic or, if stochastic are distributed independently of the stochastic disturbance term. If neither of these conditions is met, then the least square estimators are not only biased but also inconsistant; i.e. as the sample size increases indefinitely, the estimators do not converge to their true (population values. Thus, the following hypothetical system of equations.

$$Y_{1i} = \beta_{10} + \beta_{12} Y_{2i} + \gamma_{11} X_{1i} + u_{1i} \quad \text{---(3.1)}$$

$$Y_{2i} = \beta_{20} + \beta_{21} Y_{1i} + \gamma_{21} X_{1i} + u_{2i} \quad \text{---(3.2)}$$

Where $Y_1$ and $Y_2$ are mutually dependent or endogeneous variables and $X_1$ is an exogenous variable and where $u_1$ and $u_2$ are the stochastic disturbance terms, the variables $Y_1$ and $Y_2$ are both stochastic. Therefore, unless it can be shown that the stochastic explanatory variable $Y_2$ in eqn (3.1) is distributed independently of $u_1$ and the stochastic explanatory variable $Y_1$ in eqn (3.2) is distributed independently of $u_2$.

Application of classical OLS to these equations will lead to inconsistent estimates.

## Examples of Simultaneous Equation Models:

### 1) Demand-And-Supply Model:

As is well know, the price P of a commodity and the quantity Q sold are determined by the intersection of the demand-and-supply curves for that commodity. Thus, assuming for simplicity that the demand-and-supply curves are linear and adding the stochastic disturbance terms $u_1$ and $u_2$, we may write the empirical demand-and-supply functions as;

Demand function $Q_t^d = \alpha_0 + \alpha_1 P_1 + u_{1t}$

$\alpha_1 < 0$

- - - (3.3)

Supply function $Q_t^s = \beta_0 + \beta_1 P_t + u_{2t}$

$\beta_1 > 0$

- - - (3.4)

Equilibrium condition, $Q_t^d = Q_t^s$

Where $Q^d$ = quantity demanded

$Q^s$ = quantity supplied.

$t$ = time.

and the $\alpha$'s and $\beta$'s are the parameters. A prior, $\alpha_1$ is expected to be negative (downward sloping demand curve) and $\beta_1$ is

expected to be positive (upward-sloping supply curve).

Now it is not too difficult to see that P and Q are jointly dependent variables. If, for example, $u_{1t}$ in eqn (3.3) changes because of change in other variables affecting $Q_t^d$ (such as income, wealth, and tastes), the demand curve will shift upward if $u_{1t}$ is positive and downward if $u_{1t}$ is negative. These shifts are shown in fig. 3.1.

As the figure shows, a shift in the demand curve changes both P and Q. Similarly, a change in $u_{2t}$ (because of strikes, weather, import or export restrictions, etc) will shift the supply curve, again affecting both P & Q. Because of this simultaneous dependance between Q and P, $u_t$ and $P_t$ in eqn (3.3) and $u_{2t}$ and $P_t$ in eqn (3.4) can not be indep. Therefore, a regression of Q on P as in eqn (3.4) would violate an important assumption of the classical linear regression model, namely, the assumption of no correlation between the explanatory variables and the disturbance term.

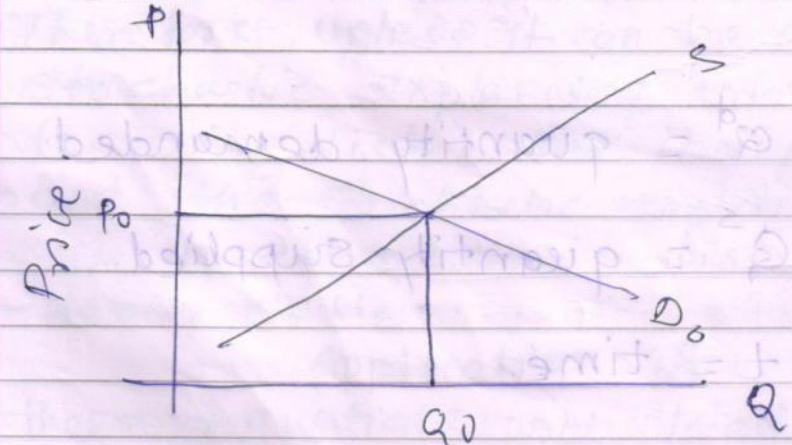



fig. 3.1 Quantity

## 2) Keynesian Model of Income Determination:

Consider the simple keynesian model of income determination:

Consumption function: $C_t = \beta_0 + \beta_1 Y_t + u_t$

$0 < \beta_1 < 1$ ...(3.5)

Income identity: $Y_t = C_t + I_t \ (= S_t)$ ...(3.6)

Where, C = consumption expenditure
Y = income
I = investment
S = savings
t = time
u = stochastic disturbance term

$\beta_0$ and $\beta_1$ = Parameters.

The parameter $\beta_1$ is known as the marginal propensity to consume (MPC). From economic theory, $\beta_1$ is expected to lie between 0 and 1. Eqⁿ (3.5) is the (stochastic) consumption function and eqⁿ eqⁿ (3.6) is the national income identity.

## * Inconsistency of OLS Estimators:

To show that OLS are bias and inconsistant, we will show that one or more of the explanatory variables are correlated with disturbance term. ~~In that equation~~ Let us consider the simple keynesian model of income determination. Suppose that we want to estimate

the parameters of consumption function (3.5). Assuming that,

$$E(u_t) = 0, \quad E(u_t^2) = \sigma^2,$$

$E(u_t u_{t+j}) = 0$ (for $j \neq 0$), which are the assumption of classical linear regression model, we first show that $Y_t$ and $u_t$ in eqn (3.5) are correlated and then prove that $\hat{\beta}_1$ is inconsistent estimator of $\beta_1$.

To prove that $Y_t$ and $u_t$ are correlated, we proceed as follows. Substitute eqn (3.5) into (3.6) to obtain,

$$Y_t = \beta_0 + \beta_1 Y_t + u_t + I_t \quad \text{---(3.7)}$$

$$Y_t - \beta_1 Y_t = \beta_0 + u_t + I_t$$

$$Y_t(1-\beta_1) = \beta_0 + u_t + I_t$$

$$Y_t = \frac{\beta_0}{(1-\beta_1)} + \frac{1}{(1-\beta_1)} I_t + \frac{1}{(1-\beta_1)} u_t$$

$$\text{---(3-8)}$$

Now

$$E(Y_t) = \frac{\beta_0}{(1-\beta_1)} + \frac{1}{(1-\beta_1)} I_t \quad \text{--- (3-9)}$$

Where use is made of the fact that $E(u_t) = 0$ and that $I_t$ being exogenous, or predetermined.

Therefore, subtracting (3.9) from (3.8), we get

$$Y_t - E(Y_t) = \frac{U_t}{1-\beta_1} \quad \ldots (3.10)$$

Moreover $U_t - E(U_t) = U_t \quad \ldots (3.11)$

$$\text{cov}(Y_t, U_t) = E[Y_t - E(Y_t)][U_t - E(U_t)]$$

$$= \frac{E(U_t^2)}{1-\beta_1} \quad (\text{from } (3.11) \text{ \& } (3.10))$$

$$= \frac{\sigma^2}{1-\beta_1} \quad \ldots (3.12)$$

Since $\sigma^2$ is positive by assumption, the covariance between Y and U given in eqn (3.12) is bound to be different from zero. As a result $Y_t$ and $U_t$ in (3.5) are expected to be correlated, which violates the assumption of the CLRM that the disturbances are independent or at least uncorrelated with the explanatory variables. Hence OLS estimators in this situation are inconsistent.

To show that OLS estimator $\hat{\beta}_1$ is an inconsistent estimator of $\beta_1$ because of correlation between $Y_t$ and $U_t$,

$$\hat{\beta}_1 = \frac{\sum (C_t - \bar{C})(Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

$$= \frac{\sum C_t y_t}{\sum y_t^2}$$

$$\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum y_t^2}$$

from eqn (3.5)

$$\hat{\beta}_1 = \frac{\sum (\beta_0 + \beta_1 x_t + u_t) y_t}{\sum y_t^2}$$

$$= \frac{\sum \beta_0 y_t + \sum \beta_1 x_t y_t + \sum u_t y_t}{\sum y_t^2}$$

$$= \frac{\beta_0 \sum y_t}{\sum y_t^2} + \beta_1 \frac{\sum x_t y_t}{\sum y_t^2} + \frac{\sum u_t y_t}{\sum y_t^2}$$

$$= \frac{\beta_0 (0)}{\sum y_t^2} + \beta_1 (1) + \frac{\sum y_t u_t}{\sum y_t^2}$$

$$= \beta_1 + \frac{\sum y_t u_t}{\sum y_t^2} \quad \cdots (3.13)$$

since $\sum y_t = 0$ & $\frac{\sum x_t y_t}{\sum y_t^2} = 1$

If we take expectation of eqn (3.13) on both sides, we get.

~~$E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\sum x_t u_t}{\sum y_t^2}\right)$~~

$$E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\sum y_t u_t}{\sum y_t^2}\right) \quad \cdots (3.14)$$

Now an estimator is said to be consistent if its probability limit or plim is equal to its true (population) value.

Applying the rules of probability limit to eqn (3.13), we obtain.

$$\text{plim}(\hat{\beta}_1) = \text{plim}(\beta_1) + \text{plim}\left(\frac{\sum y_t u_t}{\sum y_t^2}\right)$$

$$= \text{plim}(\beta_1) + \text{plim}\left(\frac{\sum y_t u_t / n}{\sum y_t^2 / n}\right)$$

$$= \beta_1 + \frac{\text{plim}(\sum y_t u_t / n)}{\text{plim}(\sum y_t^2 / n)} \quad \text{---(3.15)}$$

Hence we can say that from eqn (3.15), $\text{plim}(\hat{\beta}_1)$ is not equal to the true (population) value. Hence $\hat{\beta}_1$ is inconsistent.