ECO-04-15/16(B)

Institute of Distance and Open Learning
Gauhati University

MA/M.Sc in Economics

Fourth Semester

Optional: (B)
Econometric Methods

Contents:

Introduction:

Unit 1 : Generalised Least Squares

Unit 2 : Non-Linear Estimation

Unit 3 : Distributed Lag Models

Unit 4 : Analysis of Time Series

Unit 5 : Introduction to Simultaneous Equation Model

Contributors:

Dr. Ratul Mahanta	:	Assistant Professor Dept. of Economics Gauhati University

Editorial Team:

Prof. Kandarpa Das	:	Director IDOL, Gauhati University
Dr. Ratul Mahanta	:	Assistant Professor Dept. of Economics Gauhati University
Gitartha Goswami	:	Asst. Director IDOL, Gauhati University
Daisy Konwar	:	Accademic Consultant IDOL, Gauhati University

Cover Page Designing:

Bhaskar Jyoti Goswami	:	IDOL, Gauhati University

May, 2013

Acknowledgement

The Institute of Distance and Open Learning, Gauhati University duly acknowledges the financial assistance from the Distance Education Council, IGNOU, New Delhi, for preparation of this material.

Syllabus for MA/M.Sc Economics

Forth Semester

Optional: (B)

Econometric Methods

Unit – 1: Generalised Least Squares

Aitken's Theorem of GLS – Feasible GLS and its Properties – Heteroscedasticity: Test and Solutions – Autocorrelation: Test and Solutions.

Unit – 2: Non-Linear Estimation

Non-Linear Least Squares and Iteration process – Models with Binary Dependents Variables – Logit and Probit Models

Unit – 3: Distributed Lag Models

Concept – Koyck Model – Partial Adjustment and Adaptive Expectation Models – Estimation of Models with a Lagged Dependent Variable – Test of Autocorrelation in Auto-Regressive Models

Unit – 4: Analysis of Time Series

Components of Time Series – Fitting of Trend – Variate Difference Method – The idea of a stochastic Time Series - Stationary and Non-stationary Time Series – Autocorrelation Function and Correlelogram - the Problem of Regression Analysis with Non-stationary Time Series.

Unit – 5: Introduction to Simultaneous Equation Model

Structural and Reduced Forms – Simultaneity Bias – Informal Introduction to Identification Problem, Indirect Least Squares and Two Stage least Squares

Paper Introduction:

This paper is designed to give a basic idea of the various Econometric Methods, tools and techniques.

Unit 1 discusses the concept of generalized Least Square, the Aitken's Theorem along with the problem of Heteroscedasticity and Autocorrelation with their Test and solutions.

Unit 2 basically deals with Non-Linear Estimation giving emphasis on Logit and Probit models.

Unit 3 gives an idea about the Distributed Lag Models along with test of Autocorrelation in Auto-Regressive Models.

Unit 4 discusses about Time series Econometrics giving a basic idea of both stationary and Non-Stationary Time Series.

Unit 5 gives an introduction to simultaneous Equation Model where concepts of structural and reduced from coefficients, simultaneity Bias, Identification problems, Indirect Least Square etc are discussed.

The paper has the following five (5) units:—

Unit 1 : Generalised Least Squares

Unit 2 : Non-Linear Estimation

Unit 3 : Distributed Lag Models

Unit 4 : Analysis of Time Series

Unit 5 : Introduction to Simultaneous Equation Model

Unit -1
GENERALISED LEAST SQUARES

Contents:

1.0 Introduction
1.1 Objectives
1.2 Meaning and structure of GLS
1.3 Aitken's Theorm of GLS
1.4 Heteroscedasticity
 1.4.1 Nature of Heteroscedasticity
 1.4.2 Detection of Heteroscedasticity
1.5 Autocorrelation
 1.5.1 Nature of the problem
 1.5.2 Causes
 1.5.3 OLS Estimators in Presence of Autocorrelation
1.6 Summing Up
1.7 Self Assessment Questions
1.8 References/Suggested Readings

1.0 Introduction

In case the assumptions of the classsical linear Regression model are fulfilled, we can use OLS method for estimation of the model co-efficients. But, if it doesnot, then we have to use various method for estimation and solving the problem to find out the estimators (co-efficients) of the model.

1.1 Objectives

This unit aims to illustrate the concepts related to the various measures when the CLRM assumptions are not fulfilled.

- alternative method to OLS when the CLRM assumptions are not fulfiled i.e. GLS.
- finding the problem using OLS in presence of heteroscedasticity and selecting method for solution.
- finding the problem with regards to Autcorrelation and its various solutions.

1.2 GLS (Generalized least Squares)

1.1. One of the important assumptions of 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 to σ^2(sigma square). This is the assumption of homoscedasticity, or equal (homo) spread (scedasticity), that is equal variance., symbolically ,

$$E(u_i^2) = \sigma^2 \qquad i=1,2,....n$$

Then we can use, OLS (Ordinary least square) method to find out estimators that are BLUE. But if the variance of each distrubance term, u_i, conditional on the chosen values of explanatory variables, is not equal, i.e. symbolically,

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

Notice the subscript of σ^2, which reminds us that the conditional variance of u_i (=Conditional variances of Y_i) are no longer constant.

In such a situation if we use ordinary least square (OLS) as the estimators, what we will find that are not BLUE (Best, Linear, Unbiased).

Here we use General Least square (GLS) method. In short, GLS is OLS on the transformed variable's that satisfy the standard least squares assumptions.

For the GLS estimation, Let us consider two-variable model–

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad ....1.1$$

We can write it as for case of adgebric) mani pulation–

$$Y_i = \beta_1 X_{oi} + \beta_2 X_i + u_i \qquad ....1.2$$

Where $x_{oi}=1$ for each i. These two formulations are identical–

Now assume that the heteroscedastic variances σ_i^2 are known. Divide (1.2) through by σ_i to obtain-

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{oi}}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \qquad ....1.3$$

Where the transformed variables, are the original variables divided by (the known) σ_i.

i.e. $$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \qquad ....1.4$$

We use the notaion β_1^* and β_2^*, the parameter of transformed model, to distinguish them from the original usual parameters β_1 and β_2

The purpose of the transformation of the original model is to obtain following features of transformed error term u_i^*

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

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

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

$$=1$$

Which is a constart. That is, variance of the transformed distrubance term u_i^* is now homoscedastic. Since we are still retaining the other assumptions of the CLRM, now we can apply OLS in the transformed model which will give us the BLUE estimators. In short, the estimated β_1^* and β_2^* are now BLUE and not the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

The actual mechanism of estimating β_1^* and β_2^* are as follows- First we write down the sample regression function of 1.4

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

Now to obtain GLS estimation , we minimize -

$$\sum \hat{u}_i^{*2} = \sum \left(Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*\right)^2 \text{ ....1.5}$$

i.e where $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are the weighted least square estimators and where the weights w_i such that

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

that is, the weight is inversely poportional to the variance of u_i or Y_i conditional upon the given X_i. It is understood that ... (u_i/X_i) =var $(Y_i/X_i) = \sigma_i^2$ we can rewrite 1.5 as $\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2$ 1.6

Differentiating (1.6) with respect to $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$

$$\frac{\partial \Sigma w_i \hat{u}_i^2}{\partial \hat{\beta}_1^*} = 2\sum w_i \left(Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i\right)(-1) \text{ .... 1.7}$$

$$\frac{\partial \Sigma w_i \hat{u}_i^2}{\partial \hat{\beta}_2^*} = 2\sum w_i \left(Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i\right)(-X_i) \text{ ....1.8}$$

Setting the preceding expressions equal to Zero, we obtain the following two normal equations-

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \text{ ....1.9}$$

$$\sum w_i x_i Y_i = \hat{\beta}_1^* \sum w_i x_i + \hat{\beta}_2^* \sum w_i x_i^2 \text{ ....1.10}$$

solving simulteneously, we obtain-

$$\hat{\beta}_1^* = \overline{Y}^* - \hat{\beta}_2^* \overline{x}^* \text{ .... 1.11}$$

and $\hat{\beta}_2^* = \frac{(\Sigma w_i)(\Sigma w_i X_i Y_i) - (\Sigma w_i X_i)(\Sigma w_i Y_i)}{(\Sigma w_i)(\Sigma w_i x_i^2) - (\Sigma w_i x_i^2)}$ 1.12

its variance is given by

$$\text{Var}\left(\hat{\beta}_2^*\right) = \frac{\Sigma w_i}{\Sigma (w_i)(\Sigma w_i x_i^2) - (\Sigma w_i x_i)^2}$$

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

N. B. : 1. Weighted least square (WLS), is just a special case of more general estimating technique, GLS. In the context of hetrocedasticity, one can treat two terms WLS and GLS interchangeably.

2. $\overline{Y}* = \frac{\Sigma w_i Y_i}{\Sigma w_i}$ & $\overline{X}* = \frac{\Sigma w_i X_i}{\Sigma w_i}$

1.3 Aitkens Theorem of GLS

The alternative procedure to estimate the parameters of K variable LRM in presence of Heteroscedasticity and Auto correlation is known GLS. In the Aitkens Theorem let us ... that Ω is a panitive asymmetric definite matrix. Hence it is possible to find out another non-singular matrix (positive Definite) matrix P such that $PP' = \Omega$ ....(1)

Now multiplying equation (1) by p^{-1} and post multiplying by p'^{-1} we obtain

$$\therefore p^{-1}pp'p'^{-1} = p^{-1}\Omega p'^{-1}$$

$$\therefore I = p^{-1}\Omega p'^{-1}$$

$$\therefore p^{-1}\Omega = p'^{-1} = I \qquad ......(2)$$

Again from $PP' = \Omega$

$$\Rightarrow (pp')^{-1} = \Omega^{-1} \Rightarrow p'^{-1}p^{-1} = \Omega^{-1} \Rightarrow p^{-1}p'^{-1} = \Omega^{-1} \qquad .....(3)$$

Our regression model– $y = x\beta + u$ .... (4)

Premultiply both side by– p^{-1}

$$\therefore p^{-1}y = p^{-1}x\beta + p^{-1}u \Rightarrow y^* = x^*\beta + u^* \qquad ..... (5)$$

Then equation (5) is the modified or transformed model of the origonal model (4), Hence.

$$\therefore p^{-1}y = y^*, p^{-1}x = x^*, p^{-1}u = u^*$$

Now $E(u^*) = E(p^{-1}u) = p^{-1}E(u) = 0 \quad [\therefore G(u) = 0]$

Again $E(u^* u^{*'}) = E[(p^{-1}u)(p^{-1}u)']$

$$= E(p^{-1}uu'p - 1'$$

$$= p^{-1}E(uu')p^{-1'}$$

$$= p^{-1}\delta_u^2\Omega p^{-1} = \delta_u^2 p^{-1}\Omega p'^{-1}$$

$$= \delta_u^2 I \qquad \text{using .... (2)}$$

Thus (6) is the var– cov matrix of random distrubance term in the modified model. that is why we can use OLS method to the modified medel–

Here our $\hat{\beta}$ OLS $= (x^{*'}x^*)^{-1}x^{*'}y$

$$= \{(p^{-1}x)'(p^{-1}x)\}^{-1}(p^{-1}x)'p^{-1}y$$

$$= (x'p^{-1'}p^{-1}x)^{-1}(x'p^{-1}p^{-1}y)$$

$$= (x'\Omega^{-1}p^{-1}x)^{-1}(x'\Omega^{-1}y) \text{ using-3}$$

Which the required GLS estimator of $\hat{\beta}$

Now =

$$\text{var - cov }(\hat{\beta}GLS) = \delta_u^2(x'x)^{-1}$$

$$= \delta_u^2\{(p^{-1}x)^1 p^{-1}x\}^{-1}$$

$$= \delta_u^2(x'p^{-1'}p^{-1*}x)^{-1}$$

$$= \delta_u^2(x'\Omega^{-1}x)^{-1} \qquad .....(7)$$

Which known as the aitkens .... of the OLS estimators :-

Here

$$\hat{\beta}GLS = (x^1\Omega^{-1}x)^{-1}x'\Omega^{-1}y$$

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

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

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

Now $E(\beta GLS) = \beta + (x'\Omega^{-1}x)^{-1}x'\Omega^{-1})E(u)$

$$= \beta \therefore [E(u) = 0]$$

In the same way we can frof that the GLS estimators are efficient with the use of var-cov matrix of $\hat{\beta}$ GLS i,e

$$\text{var}-\text{cov}(\hat{\beta}GLS) = \delta_u^2 (x'\Omega^{-1}x)^{-1}$$

$$\Omega^{-1} = \begin{bmatrix} 1 & -p & 0 & 0....0 & 0 \\ -p & (1+p) & 0 & 0....0 & 0 \\ 0 & -p & (1+p) & 0....0 & 0 \\ \vdots & & & & \\ 0... & 0 & 0. & 0...-p & 1 \end{bmatrix}$$

1.4 Heterosecdasticity : Test and Solutions

1.4.1 The Nature of Heteroscedasting

Heteroscedasticity is opposite to Homosecdasticity. As one of the important assumptions of classical Linear .Regression model is that the variance of each distrubance term u_i, Conditional on the chosen values of the explanatory variables, is some constant number equal to σ^2. This is the assumption of homoncedasticity , or equal (homo) spread (Secdasticity), that is, equal variance, Symbolically,

$$E(u_i^2) = \sigma^2 \qquad i = 1,2,... 4$$

Diagrammatically-

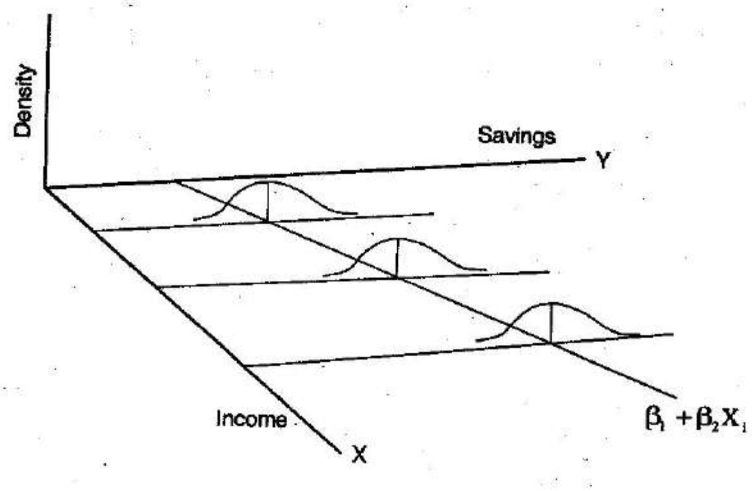
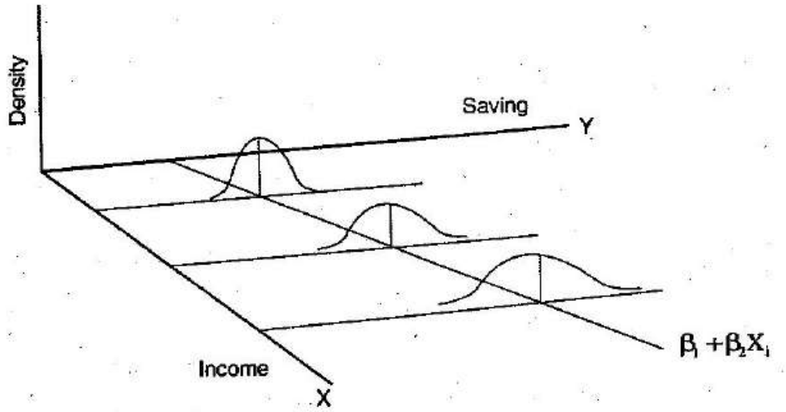


Fig 1.1 The figure shows, the conditional variance of Y_i (which is equal to that of u_i), conditional upon the given X_i, remains the same regard less of the values taken by the variable X. Again,



Above Fig 1.2 which shows that the conditional variance of Y_i, increases as X increases. Here, the variancess of Y_i are not the sam. Hence is heterosecdasticity. Symbolically,

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

Notice the subscripts of σ^2, which remind us that the conditional variances of u_i(= conditional variances of Y_i) are no longer constant.

Two make the difference between homoscedasticity and heterosecdasticity clear, assume that in the two-variable model $Y_i = \beta_1 + \beta_2 x_i + u_i$ Y represents. Savings and X represents income. Fig. 1.1 & 1.2 shows that as income increases savings also increases (i,e, on the average of income). But Fig 1.1 the variance of savings remains the same at all levels of income whereas in Fig 1.2 it increases with income. It seems that the higher income families on the average save more than the lower-income families, but there is also more variability of savings.

There are several reasons why the variances of u_i, may be variable, some of which are as follows.

1. Following the error learning models, as people learn, their errors of behavior become smaller over time.
2. As income grow, people have more discretionary income, and hance more scope for choice about the disposition of their income.
3. As data collecting techniques improve, σ_i^2 is likely to decrease.
4. Heteroscedasticity can also arise as a result of the presence of outliers.
5. Another source of heterocedasticity arises from violating Assumptions of CLRM, namely, that the regression model is correctly specified.
6. Another source of heterosecdasticity is skewness in the distribution of one or more regresors included in the model.
7. Other sources of hetrosecdasticity : As David Headry notes, (1) Incorrect data transformation (e.g. ratio or first difference

transformation) (2) Incorrect functional form (e.g. linear versus log-models)

1.4.2 Detection of HET Eroscedasticity

The important practical question is : How does one know that hetrosecdasticity is present in a specific situation? There are no hard-and-fast rules for detecting hetrosecdasticity, only a few rules of thumb.

Informal methods :

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

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 $\hat{u}_i^2$ are not the same thing as $\hat{u}_i^2$, 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.

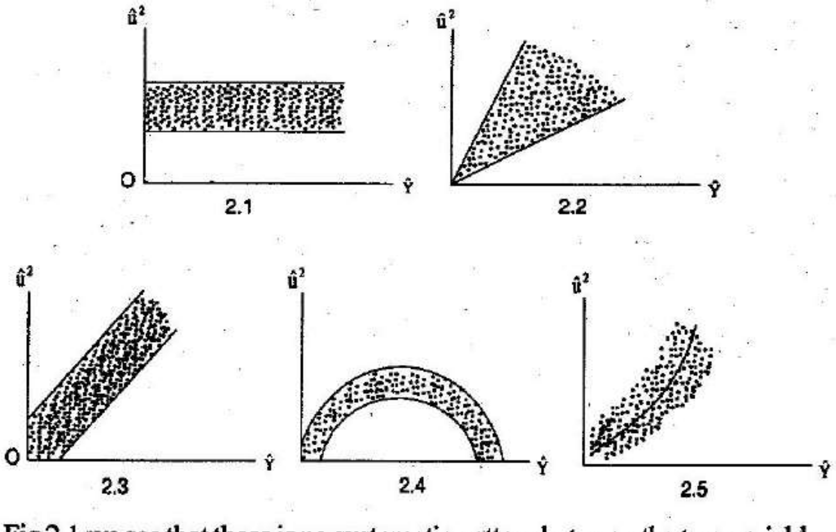
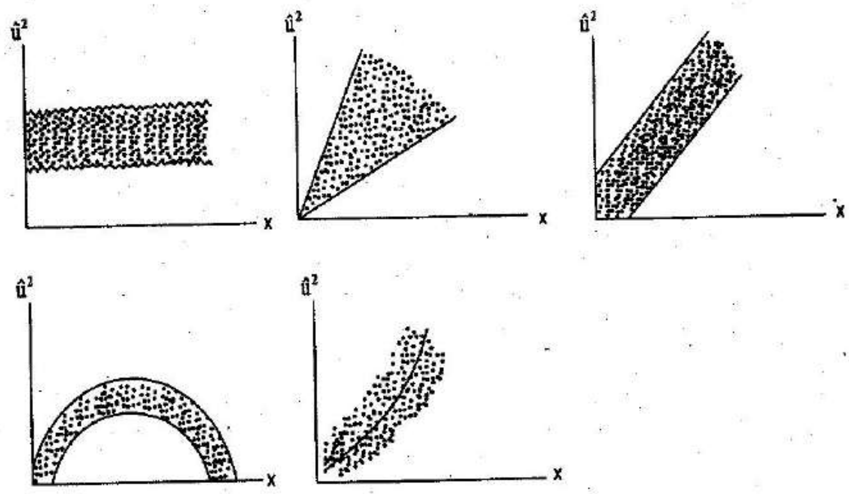


Fig 2.1 we see that these is no systematic pattern between the two variables, suggesting that perhaps no hetrosecdasticity is present in data. But Fig. 2.2 to 2.5 exhibit definite relationship (pattern) Fig 2.3 suggest a linear relationship. 2.4 & 2.5 indicate quadratic relationship between $\hat{u}_i^2$ & $\hat{Y}_i$.

Instead of plotting $\hat{u}_i^2$, against $\hat{Y}_i$, one may plot them against one of the explonatory variables, especially if plotting $\hat{u}_i^2$ against $\hat{Y}_i$ result in the pattern shown 2.1. But the relationship become same as follows.



Formal Test :

Park Test : Park formalizes the graphical methods by suggesting that σ_i^2 is some function of the explanatory variable X_i. The functional form suggested was-

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

$$\ln \sigma_i^2 = \ln\sigma^2 + \beta \ln X_i + v_i \qquad \text{....... (A)}$$

Where v_i - storhastic distrubance term.

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

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

$$= \alpha + \beta \ln X_i + v_i$$

If β turn out to be statistically significant, it would suggest that heterosecdasticity is present in the data. If it turns out to be insignificant, we may accept the assemption of homosecdasticity.

Two stage :

(i) We run OLS regression disregarding the heterosecdasticity question. We obtain $\hat{u}_i$ from this regression, and then in the second stage we run the regression.

ii) We run the regression β :

Although empirically appealing, Park test has some problems. Goldfed and quandt have argued that the error term v_i entering into (β) may not satisfy OLS asumptions and may itself be heteroscedastic.

2) Glesjer Test :

The Glesjer test is similar in spirit to the Park test. After obtaining the residuals $\hat{u}_i$ from the OLS regression, Glesjer suggests regressing the absolute values of $\hat{u}_i$, on the X variable that is thought to be closely associated with σ_i^2. In his experiments, Glesjer used the following functonal forms :

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

Where v_i is the error term.

Again as an empirical or practical matter, onemay use the Glesjar approach. But Goldfed and Quandt point out that the error term v_i has some problems, in that its expected value is nonezero, it is serially correlated and ironically it is heteroscedastic. An additional difficulty with the Glejser method is that models such as -

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

are nonlinear in the parameters and therefore cannot be estimated with the usual OLS procedure.

1st four models (preceding) give generally satisfactory results in detecting heterosecdasticity.

Spearman's Rank Correlation Test :

Spcarman's rank correlation coefficient is-

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)}\right] \text{ where } d_i\text{=difference in ranks}$$

assigned to the different characteristics of the ith individual or phenomenon and n=numbers of individuals or phenomena ranked. The preceding rank correlation coefficient can be used to detect heterosecdasticity as follows. Assume-

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Step 1. Fit the regression to the data on Y and X and obtain the residuall $\hat{u}_i$

Setp 2. Ignoring the sign of $\hat{u}_i$, that is taking their absolute value $|\hat{u}_i|$, rank both $|\hat{u}_i|$ and X_i (or $|\hat{Y}_i|$) according to an ascending or descending order and compute the Spearman's rank correlation coefficient given previously.

Step 3. Assuming that the population rank correlation coefficient p_s is Zero and n>8, the significance of the sample r_s can be tested by the t test as follows-

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \qquad \text{.... (A)}$$

With df=n-2

If the computed 't' value exceeds the critical 't' value, We may accept the hypothesis of heteronscedasticity : Otherwise we may reject it. If the regression model involves more than one X variable, r_s can be computed between $|\hat{u}_i|$ and each of the X variables separately and can be tested for statistical significance by the 't' test by equation (A).

Goldfed-Quandt. Test :

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

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \text{........... (1)}$$

Suppose σ_i^2i is positively related to Xi as

$$\sigma_i^2 = \sigma^2 X_i^2 \qquad \text{............ (2)}$$

Where σ^2 constant

Assumption (2) pastulates that σ^2 is poportional to the square of the X variable. Such an assumption has been found quite useful by Prais and Houthakker in their study of family budgets.

If (2) is appropriate, it would mean σ_i^2 would be larger, the larger the value of X_i, if that turns out to be the case, heteroscedasticity is most likely to be present in the model. To test the explicity, Goldfeld and Quandt. suggested the following steps-

Step 1 : Order or rank the observations according to the values of X_i, begining with the lowest X valve.

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

Step 3 : Fit separate OLS regression 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 small variance group) and RSS_2 that from the larger X_i values (the large variance group). These RSS each have

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

Where k is the number of parameters to be estimated, including the intercept (why?), For the two variable case k is of course 2.

Step 4

Compute the ratio

$$\lambda = \frac{RSS_2/df}{KSS_1/df} \quad ....(3)$$

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 λ of (2) follows F distribution with numerator and denominator df each of (n-c-2k)/2.

If in an application the computed λ (=F) is greater than the critical 'F' at the chosen level of signifiance, we can reject the hypothesis of homoscedasticity, that is, we can say that heteroscedasticity is very likely. Before illustrating the test, a word about omitting the c central observation is in order. For two variable model the Monte carlo experiments done by Goldfeld and Quandt suggest that c is about 8 if the sample size is about 60 and it is about 16 if the sample size is about 60. But Judge et. a note that c=4 if n=30 and c=10 if n is about 60 have been found satisfactory in practice.

But when there more than one X variable in the model, the ranking of observations, the first step in the test can be done according to any one of them. Thus in the model . $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$, we can rank order the data according to any one of these X's. If a priori we are not sure which X variable is oppropriate, we can conduct the test on each of the X variables, or via a Park test, in turn, on each X.

Breusch-Pagan-Godfrey Test :

The success of the Goldfeld- Quandt test depend not only on the value of c (the number of central observations to be omitted) but also on identifying the correct X variable with which to order the observations. This limitation of this test can be avoided if we consider the Breusch- Pagan-Godfrey (BPG) test.

To illustrate this test, consider K-variable linear regression model-

$$Y_i = \beta_1^* + \beta^2 X_{2i} + ... + \beta_k X_{Ki} + u_i \quad .... \text{(a i)}$$

Assume that the error variance σ^2 is described as

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi}) \quad .... \text{(a ii)}$$

that is σ_i^2 is a linear function of the Z's. If $\alpha_2 = \alpha_3 = ... = \alpha_m = 0$, $\sigma_i^2 = \alpha_1$ which is a constant. Therefor to test whether σ^2 is homoscedastic, one can test the hypothesis that $\alpha_2 = \alpha_3 = ... = \alpha_m = 0$ This is the basic idea behind the Breusch-Pagan-Test. The actual test procedure is as follows.

1step- Estimate (ai) by OLS and obtain the residuals $\hat{u}_1, \hat{u}_2, ... \hat{u}_n$

2step- Obtain $\tilde{\sigma}^2 = \sum \hat{u}_i^2 / n$ (this is the maximum (ML) estimator of σ^2.)

Step 3 : Construct variable p_i, devined as $p_i = \hat{u}_i^2 / \tilde{\sigma}^2$

which is simply each residual disided by $\tilde{\sigma}^2$

Step 4 : Regress p_i thus constructed on the Z's as

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi} + v_i \text{ ...... (a iv)}$$

Where v_i is the residual term of this regresion

Step 5 : Obtain the ESS (explained sum of squares) from (a iv) and define

$$\Theta = \frac{1}{2} \text{ (Ess)}$$

Assuming u_i are normally distributed, one can show that if there is homoscedasticity and if the sample size n increases indefinitely then

$$\Theta \tilde{asy} \chi^2_{m-1}$$

that is, Θ follows the chi-square distribution with (m-1) degress of freedom. (Note: asy means as asymptotically).

Therefore, if in an application the computed $\Theta = (\chi^2)$ exceeds the critical χ^2 value at the choosen level of significance, one can reject the hypothesis of homoscedasticity, otherwise one does not reject it.

White's General Heteroscedasticity Test :

Unlike Goldfeld Quandt test, which requires reordering the observations with respect to the X variable that supposedly caused heteroscedasticity or the BPG test, which is sensitive to the normality assumption, the general test of heteroscedasticity proposed by White does not rely on the normality assumption and is easy to implemeat. As an illustration the basic idea, consider the following three-variable regression model (the generalization of the k-variable model is straight forward).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_2 X_{3i} + u_i \text{ .... (1)}$$

The white test proceeds as follows-

Step 1: Given the data, we estimate (1) and obtain the residuals $\hat{u}_i$

Step 2 : We then run the following (auxillary) regression :

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \text{ .... (2)}$$

That is, the squared residuals from the original regresion are regressed on the original X variables or reggressors, their squared values, and the cross product (s) of regressors. Higher power of regressors can also be introduced. Note that there is a constant term in this equation even though the original regression may or may not contain it. Obtain the R^2 from this (auxillary) regression.

Step 3 : Order the null hypothisis that these is no heteroscedasticity, it can be shown that sample size (n) times the R^2 obtained from the auxilliary regression asymptotically follows the chi-square distribution with df equal to the number of regressors (including the constant term) in the auxilliary regression.

That is - $n.R^2_{asy2}\chi^2 df$ .... (3)

In our example, there are 5 df since there are 5 regressors in the auxilliary regression.

Step 4 : If the chi-square value obtained in (3) exceeds critical Chi-square value at chosen 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 auxilliary regression,

$\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$

A comment is in order regarding the White test. If a model has several regressors, then introducing all the regressor, their squarded, and higher powered and also cross product, quickly consume degrees of freedom. Therefore one must exercise caution using the test.

Due to above cause White test can be a test of (pure) heteroscedasticity or specification error or both. If no cross product term are present in the White test then pure heteroscedasticity or vice-versa.

Remedial Measures :

As we have seen, heteroscedasticity does not destroy the unbiasedness consistancy properties of the OLS estimators, but they are no longer efficient, not even asymptotically (i.e. large sample size).

Two approaches :

When σ_i^2 is known : the most straitforward method of correcting heteroscedasticity is by means of weighted least squares, for the estimators thus obtained are BLUE.

When σ_i^2 is not known :

White's Heteroscedasticity - consistant variances and standard Errors-- White has shown that this estimate can be performed so that asymptotically valid (large-sample) statistical infersenses can be made about the true parameter values. Incidentally white's heteroscedasticity-correeted standard errors are also known as robest standard errors.

Example :- We quote the following result from Greene.

$$\hat{Y}_i = 832.91 - 1834.2\,(\text{Income}) + 1587.04\,(\text{Income})^2$$

OLS se	= (327.3)	(829.0)	(519.1)
t	= (2.54)	(2.21)	(3.06)
White se	= (460.9)	(1243.0)	(830.0)
t	= (1.81)	(-1.48)	(1.90)

Where Y=per capita expenditure .... public schools by state in 1979 and Income = per capita income by state in 1979. The sample consisted of 50 states plus washington D.C.

As the preceding result show, (White's) heteroscedasticity-corrected standard errors are considerably larger than the OLS standard errors and therefore the estimated 't' values are much smaller than those obtained by OLS. On the basis of OLS both of the regressor are statistically significant at 5 percent level of significant. But on the basis of white test they are not. However it should be pointed out that White heteroscedasticity -correlated standard error can be larger or smaller than the uncorrelated standard errors.

Plausible Assumptions about Heteroscedasticity Pattern:

Apart from a being a large sample procedure, one drawback of the white procedure is that the estimators thus obtained may not be so efficient as those obtained by methods that transform data to reflect specific types of heteroscedasticity . To illustrate this, let us request to the two variable regression model :

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

We now consider saveral assumption about the pattern of heteroscedasticity.

Assumption 1.

The error variance is poportional to X_i^2 i,e

$$E(u_i^2) = \sigma^2 X_i^2 \quad \text{............} \quad (1)$$

It is believed that the variance of u_i is poportional to the square of the explanatory variable X. One may transform the original model as follows. Divided the original model through by X_i.

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} \quad \text{..........} \quad (1.a)$$

$$= \beta_i + \frac{1}{X_i} + \beta_2 + v_i$$ where v_i is the transformed distrubance

terms, equal to u_i/X_i . Now it is easy to verify that

$$E(v_i^2) = E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2}E(u_i^2)$$

$$\sigma^2 \text{ (using ....1)}$$

Hence the variance of v_i is now homosecdastic, and now we may proceed to apply OLS to the transformed equation 1(a), regressing Y_i/X_i on $1/X_i$.

Notice that in the transformed regression the intercept term β_2 is the slope coefficient in the original equation and the slope coefficient β_1 is the interuption the original model. To get back the original model we shall have to multiply the estimated 1(a) by X_i

Assumption 2 :

The error variance is poportional to X_i. (The square root transformation)

$$E(u_i^2) = \sigma^2 X_i \qquad \text{........(2)}$$

If it is believed that the variance of u_i, instead of being proportional to the squared X_i, is proportional to X_i itself,

Then the original model can be transformed as follows

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{x_i}} \qquad \text{........ 2 (a)}$$

$$= \beta_1\frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + v_i$$

Where $v_i = u_i/\sqrt{X_i}$ and where $X_i > 0$

Given assumption 2, one can readily verify that $E(v_i^2) = \sigma^2$, a homoscedastic situation. Therefore, one may proceed to apply OLS to 2(a) regerssing $Y_i/\sqrt{X_i}$ on $1/\sqrt{X_i}$ and $\sqrt{X_i}$.

Note an important feature of the transformed model : It has no intercept term. Therefore, one will have to use the regression through the origine model to estimate β_1 and β_2. Having run.

2 (a), he can get back to the original model simply by multiplying by 2(a) by $\sqrt{X_i}$.

Assumption 3 :

The error variance is poportional to the square of the mean value of Y.

$$E(u_i^2) = \sigma^2[E(Y_i)]^2 \qquad \text{....... (3)}$$

Equation (3) postulates that the variance of u_i is poportional to the square of the expected value of Y. Now,

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

Therefore, if we transform the original equation as follows :

$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} \qquad \text{.....} 3(a)$$

$$= \beta_1\left(\frac{1}{E(Y_i)}\right) + \beta_2 \frac{X_i}{E(Y_i)} + v_i$$

Where , $v_i = u_i / E(Y_i)$ it can be seen that $E(v_i^2) = \sigma^2$ that is, the distribances v_i are homoscedastic. Hence it is regrassion 3 (a) that will satisfy the homoscedasticity assumption of the classical linear regression model.

The transformation 3(a) is, however, inoperational because $E(Y_i)$, depend upon β_1 and β_2, which are unknown. Of course, we know $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ which is an estimator of $E(Y_i)$. Therefore we may proceeds in two steps. First, we run the usual OLS regression, disregarding the heteroscedasticity problem and obtain $\hat{Y}_i$.

Then, using the estimated $\hat{Y}_i$, we transform our model as follows :

$$\frac{Y_i}{\hat{Y}_i} = \beta_1 \frac{1}{\hat{Y}_i} + \beta_2\left(\frac{X_i}{Y_i}\right) + v_i \qquad \text{.......}3\text{ (b)}$$

Where $v_i = (u_i / \hat{Y}_i)$

Step 2:

We run regression 3(b) Although $\hat{Y}_i$ are not exactly $E(Y_i)$, they are consistant estimators, that is as the sample size increases indefinitely, they converge to true $E(Y_i)$.

Assumption 4 : A log transformation such as

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i \qquad \text{........(4)}$$

very often reduces heteroscedasticity when compared with the regression

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

This result arises because log transformation compresses the scales in which the variables are measured, thereby reducing a tenfold difference between two value to a twofold difference. Thus, the member 80 is 10 times the number 8, but ln 80=(4.3280) is about twice as large as ln 8=(2.0794).

An additional advantage of the log transformation is that the slope coefficient β_2 measures the elasticity of Y with respect to X, that is, the percentage change in Y for a percentage change in X.

Example : If Y is consumption and X is income, β_2 in (4) equation will measure income elasticity, whereas in the original model β_2 measure only the rate of change of mean consumption for a unit change in income. It is the reason why the log model are quite popular in empirical econometrics.

1.5 Autocorrelations : Test & Solutions

There are generally three types of data that are available for imperical analysis (i) Cross sectional (ii) time series and (iii) combination of cross section and time series which is also known as pooled data.

Cross sectional data are often plagued by the problem of heteroscedasticity. But the situation however likely to be very different if we are dealing with time series data, for the observations in such data follow as natural odering over time so that successive observations are likely to exhibit intercorrelations, especially if the time interval between successive observations is short, such as a day, a week, or a month rather than a year. Obviously in situation like this, the assumptions of no auto, or serial correlation in the error terms that underlies the satisfaction of CLRM will be violated. Under both heteroscedasticity and autocorrelation the usual OLS estimators, although linear, unbiased, and asymptotically (v, e, in large samples) normally distributed, are no longer of minimum variance among all linear unbiased estimators. In short, they are not efficient relative to other linear and unbiased estimators. Put it differently, they may not by BLUE. As a result, the usual t, F, and χ^2 may not valid.

1.5.1 The nature of the Problem

The term auto correlation may be defined as "Correlation between member of series of observations ordered in time [as in time series data] or space [as in cross-sectional data]." In the regression context, the classical linear regression model assumes that such auto correlation does not exist in the disturbance u_i. Symbolically-

$$E(u_i u_j) = 0 \qquad i \neq j$$

Classical linear regression model (CLRM) assumes that the distrubance term relating to any observation is not influenced by the distrubance term relating to other observations.

Let us visualize some of the plausible patterns of auto-and non auto correlations, which are givem in following figure- Fig 1 shows a pattern. Fig 2 and Fig 3 suggest upward and downward pattern and Fig 4 indicate that both linear and quadratic trend terms are present in the distrubances. Only Fig 5, indicate no systematic pattern, supporting the non auto correlation assumption of the CLRM.

Fig-1

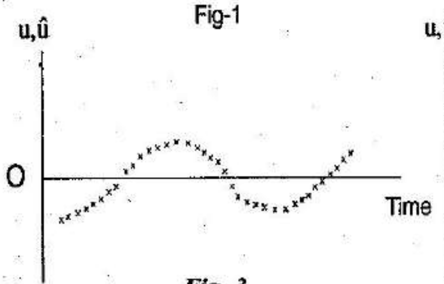


Fig-2

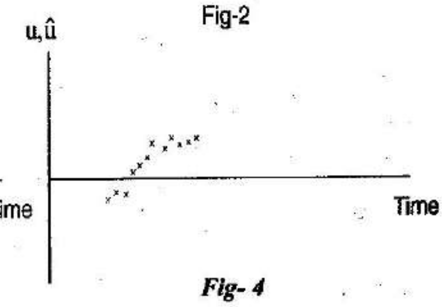
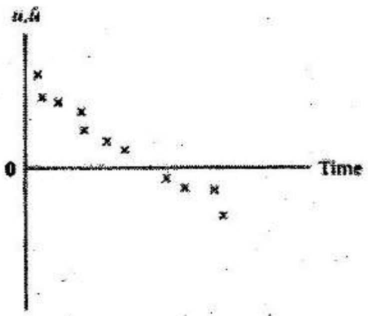


Fig- 3

Fig- 4



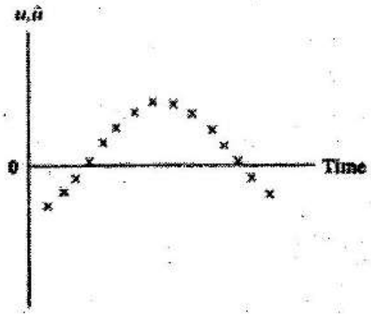
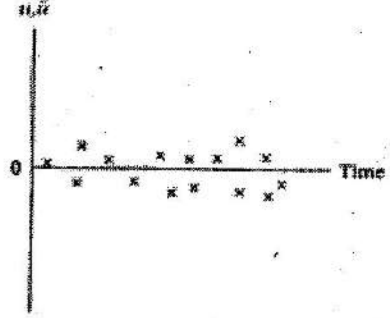


Fig- 5



However, if there is such a relationship (dependance), we have autocorrelation. Symbolically- $E(u_i u_j) \neq 0 \quad i \neq j$

In this situation, distrubance term relating to any observations is influenced by the distrubance term relating to any other observations.

The term "autocorrelation" and "serial correlation" are treated synomymously, but some authors prefer to distinguish the two terms. For example i, Tintner defines auto correlation as "lag correlation of a given series with itself, lagged by a number of time units," whereas he reserves the term serial autocorrelation to "lag correlation between two different series."

For eg.- autocorrelation between two time series such as $u_1 u_2$ .... u_{10} and u_2, u_3.......u_{11}. where the former is the latter series lagged by one time period.

Correlation between two time series such as $u_1\ u_2 .... u_{10}$ and $v_2\ v_3 ... v_{11}$, where 'U' and 'V' two different time series, is called serial autocorrelation.

1.5.2 Causes

There are the following causes why autocorrelation occurs as follows–

Inertia : Inertia or sluggishness is the salient feature of most economic problem. For instance, Time series like GNP, price indexes, production, employment, and unemployment exhibit (business) cycle. Therefore, in regressions involving time serios data, successive observations are likely to be interdependent.

Specification Bias : Excluded Variable Case :

In empirical analysis the researcher often strats with a plausible model that may not be the most 'perfect' one. After the regression analysis the researcher does the postmortem to find out whether the result accord with a priori expectation. If not surgery is begun.

An example :- Suppose we have following demand model :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \quad .... (1)$$

Where Y= quantity of beef demanded, X_2=price of beef X_3=consumer income, X_4=price of Pork and t=time, for some reason we run following regression -

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \quad .... (2)$$

If model (1) is correct than in the model (2) we are letting $v_t = \beta_4 X_{4t} + u_t$. And to the extent the price of pork affect the consumption of beef, the error or distrubance term v will reflect systematic pattern, thus creating (false) autocorrelation.

Specification Bias : Incorrect Functional Form :

Suppose the 'true' model in a cost output study is as follows -

Marginal $\text{cost}_i = \beta_1 + \beta_2\, \text{output}_i + \beta_3\, \text{output}_i^2 + u_i$ ......(1) but we fit the model as

Marginal $\text{cost}_i = \alpha_1 + \alpha_2$ output + v_i ....(2)

Model (1) shows us quadratic relationship but model (2) shows us linear relationship due to which disturbance term v_i which is in fact equal to $\text{output}_i^2 + u_i$ will reflect autocorrelation.

Cobweb Phenomenon :

This situation arise in cash of many agricultural commodities, where supply reacts to price with a lag of one time period because supply decisions take time to implement. Thus, at the begining of this year's planting of crops, farmers are influenced by the price prevaling last year, so that supply function is–

$$\text{Supply}_t,\ \beta_1 + \beta_2 P_{t-1} + u_t \ .... (1)$$

Suppose at the end of the period t, price P_t, turns out to be lower than P_{t-1}, therefore, in period t+1 farmers may very well decide to proceduce less than they did in period t. Obviously, in this case U_ts are not expected to be random and thus leading to autocorrelation.

Lags :

In regression of consumption expenditure on income of time series data, we all generally found that the consumption expenditure in the current period depend upon consumption expenditure of the previous period also.

$\text{Consumption}_t = \beta_1 + \beta_2\ \text{Income}_t + \beta_3\ \text{consumption}_{t-1} + u_t$ ....(a)

Such regression like (a) is known as autoregression because one of the explanatory variable is the lagged value of the dependent variable and if we neglect the lagged term the resulting error term will reflect a systematic pattern.

'Manipulation' of Data :

In empirical analysis, the raw data are often "manipulated". For example, quarterly data are obtained from monthy data by adding three monthy observations and dividing it by 3. It leads to introducing autocorrelation. Another source of manipulation is interpolation and extrapolation of data.

Data Transformation :

Let us consider the following model -

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \text{....... (1)}$$

Where Y= consumption expenditure and x=income since (1) hold true at every time period, it holds true also in the previous time period, (t-1) so we can write-(1) as

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad \text{.......... (1.a)}$$

$Y_{t-1}, X_{t-1}, u_{t-1}$ are known as lagged value of Y, X and u respectively, here lagged by one period.

Now if wesubtract1(a) from 1. we get

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \qquad \text{........ 1(b)}$$

Where Δ is known as first differenc operator. Let us take successive differences of the variables in question. Thus

$$\Delta Y_t = (Y_t - Y_{t-1}), \Delta X_t = (X_t - X_{t-1}) \text{ and } \Delta u_t = (u_t - u_{t-1})$$

We can rewrite 1(b) as-

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \quad \text{........ 1(c) where } v_t = \Delta u_t = (u_t - u_{t-1})$$

equation 1(c) is known as the level of form of 1(b) and as the (first) difference form. It is also known as dynamic regression models, that is, models involving lagged regressands.

Nonstatisnarity :

When the time series are non stationary there also arise autocorrelation.

1.5.3 OLS Estimator in Presence of Autocorrelation

We used a two variable linear regression model to explain the basic idea involved, namely, $Y_t = \beta_1 + \beta_2 X_t + U_t$. CLRM assumption about u_t namely, for $E(u_t - u_{t+s}) \neq 0 (s \neq 0)$, is too general to be used practically.

Generally when autocorrelation present we assume simple first order autocorrelation in linear form -

$$u_t = pu_{t-1} + \varepsilon_t \quad -1 < p < 1 \qquad \text{........ (1)}$$

Where ρ is known as the coefficient of autocovariance and where ε_t is the stochastic distrubance term such that it satisfies OLS assumption-

$$E(\varepsilon_t) = 0$$

$$Var\ (\varepsilon_t) = \sigma^2$$

$$Cov\ (\varepsilon_t, \varepsilon_{t+s}) = 0 \quad s \neq 0$$

In the engineering literature, an error term with the preceding properties is often called a white noise error term.

Equation (1) is also known as Markov-first order antoregressive scheme. Generally the ρ, the coeffiecient of autocovariance can also be interpreted as the coefficient of autocorelation at lag1.

$$\rho_s = \frac{E\{[u_t - E(u_t)][u_{t-1} - F(u_{t-1})]\}}{\sqrt{Var\ (u_t)}\sqrt{Var\ (u_{t-1})}}$$

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

Since $E(u_t)=0$ for each 't' and var (u_t)= var (u_{t-1}) because we are retaining these assumption of homoscedasticity.

Given AR (1) scheme, it can be shown that

$$Var(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

$$Cov\ (u_t, u_{t+s}) = E(u_t u_{t-s}) = p^2 \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

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

Where Cov (u_t, u_{t+s}) means covariance between error terms periods apart and where cor (u_t, u_{t+s}) means correlation between error terms periods apart. Note that because of property of covariances and corelations, Cov (u_t, u_{t+s}) =cor (u_t, u_{t+s}) and vice-versa.

Since ρ is a constant between -1 and +1 it show, under the scheme AR(1), the variance of u_t is still homoscedastic but u_t is correlated not only with its immediate past value but many several period in the past.

When $|\rho| < 1$, we say the that AR(1) process given is stationary.

Now the two variable model $Y_t = \beta_1 + \beta_2 x_t + u_t$. We know that the OLS estimator of the slope coefficient is

$$\hat{\beta}_2 = \frac{\Sigma x_t y_t}{\Sigma x_t^2} \quad \text{............... (1)}$$

and its variance is given by-

$$Var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\Sigma x_i^2} \quad \text{......... (2)}$$

Where the small latter as usual denote deviation from the mean values. Now under the AR (1) scheme the variance of this estimators can be shown as-

$$Var\left(\hat{\beta}_2\right)AR_1 = \frac{\sigma^2}{\Sigma x_t^2}\left[1 + 2\rho\frac{\Sigma x_t x_{t-1}}{\Sigma x_t^2} + 2\rho^2\frac{\Sigma x_t x_{t-2}}{\Sigma x_t^2} + 2\rho^{n-1}\frac{x_1 x_n}{\Sigma x_t^2}\right] \text{..... (3)}$$

Where $Var\left(\hat{\beta}_2\right)AR_i$ means the variance of $\hat{\beta}_2$ under the first order autoregressive scheme.

When we compare the OLS variance and AR(1) variance, we show that the former is equal to the latter times a term that depends on ρ as well as the sample covariance between the values taken by the regressor X at variuos lages. And in general we can't define whether is greater or lower than $AR_1\left(\hat{\beta}_2\right)$.

The reduced form of the two variances is

$$Var\left(\hat{\beta}_2\right)AR(1) = \frac{\sigma^2}{\Sigma x_t^2}\left(\frac{1+r\rho}{1-r\rho}\right) = Var\left(\hat{\beta}_2\right)OLS\left(\frac{1+r\rho}{1-r\rho}\right) \quad \text{....... (A)}$$

It, for example, $r = 0.6$ and $\rho = 0.8$ using formula (A) we can check $\left(\hat{\beta}_2\right)AR(1) = 2.8461\, Var\left(\hat{\beta}_2\right)$ OLS. To put it another way, $Var\left(\hat{\beta}_2\right)OLS = \frac{1}{2.8461} Var\left(\hat{\beta}_2\right)AR(1) = 0.3513\, Var\left(\hat{\beta}_2\right)AR(1)$. This is the usual OLS formula which will underestimate the variance of $\left(\hat{\beta}_2\right)AR(1)$ by about 65 percent, as we know that the result is spesific for the value of γ and ρ.

Consequences :

OLS formulas under the presence of autocorrwlation, to compute the variances and standard errors of the OLS estimator could give seriously misleading answer.

In others words the OLS estimators in presence of autocorrelations gives unbiased and consistent estimators but not minimum variance i,e

$$\text{Var}\left(\hat{\beta}_2\right) \angle \text{Var}\left(\hat{\beta}_2\right) \text{AR}(1)$$

Detecting Auto Correlation :

I) Graphical Method : The assumption of CLRM of non autocorrelation relates to the population distrubances u_t, which are not directly observable. As we know 'u_t' are not directly observable we use proxies of u_t.

There are various ways of examining the residuals. We can simply plot them against time, the time sequence plot. Alternatively we can plot the standarlized residuals against time.

To see it differently, we can plot $\hat{u}_t$ against $\hat{u}_{t-1}$, that is, plot the residual at time 't' against their value at time (t-1), a kind of empirical test of the AR(1) scheme.

The graphical method, although powerful and suggestive, is subjective or qualitative in nature.

II) The Runs Test :

When there are several residuals that are negative, then there are several residuals, which are positive, again there are several residual that are negative. If this residuals were purely random, could we observe such a pattern (like inverted U)? It its unlikely. This situation can be checked by the so-called run test, sometimes also known as the Geary test, a nonparametric test. Let us note down the sighns such as (+ or -). sign

(- - - - -) (+ + + + + + + + + + +) (- - - - - -)

of the residuals in the regression.

N.B. 1 See Damodor N Gujrati & Sangeetha for example.

Now Let

N= total number of observations = N_1+N_2

N_1= number of + symbols (i, e + residuals)

N_2= number of - symbols (i, e - residuals)

R = number of Runs.

Under the Null hypothesis - Successive outcome (residual) are independent, and assuming that $N_1>10$ and $N_2>10$. The number of runs is (asymptotically) normally distributed with-

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

$$\text{Variance } \sigma_R^2 = \frac{2N_1N_2\left(2N_1N_2 - N\right)}{(N)^2(N-1)}$$

Note $N = N_1 + N_2$

If the Null hypothesis of randomness is sustainable, following the properties of the normal distribution, we should expect that

$$\text{Prob}\left[E(R) - 1.96\,\sigma_R < R \leq E(R) + 1.96\sigma_R\right] = 0.95$$

In 95% cases, the preceeding interval will include R.

In general, If there is positive autocorrelation, the number of runs will be few and if there is negative autocorrelation the number of run will be many.

III) Durbin-Watson d Test : The most popular test for detecting autocorelation is Durbin-Watson d test. Also known Durbin-Watson d statistic.

$$d = \frac{\sum_{t=2}^{t=n}\left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=2}^{t=n}\hat{u}_t^2} \quad \text{......... (1)}$$

i,e the ratio of sum of squared differences in successive residual to the RSS. In the numerator of 'd' statistic, number of observation is (n–1) because one lost in taking successive differences.

Assumptions :

1. The regression model includes the intercept terms. If it is not present, as in case of regression through the origine, it is essential to rerum the regression with intercept to obtain the RSS.
2. The explanatory variables the X's are nonstochastic, or fixed in repeated sampling.
3. The distrubance term follow first order autoregressive scheme $u_t = \rho u_{t-1} + \varepsilon_t$.
4. The error term u_t is normally distributed.
5. The regression model does not include lagged variables of explained and explanatory variables.
6. There are no missing observations in the data.

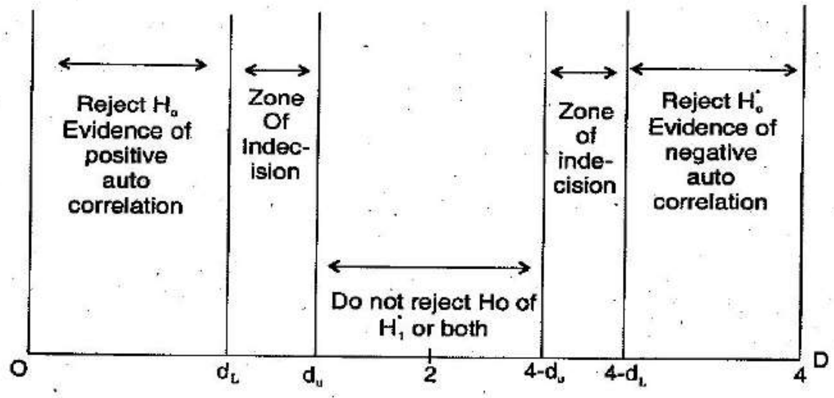
Now Expanding (1) we get-

$$d = \frac{\Sigma\hat{u}_t^2 - \Sigma\hat{u}_{t-1}^2 - 2\Sigma\hat{u}_t\hat{u}_{t-1}}{\Sigma\hat{u}_t^2} \quad \text{........ (2)}$$

Since $\Sigma\hat{u}_t^2$ and $\Sigma\hat{u}_{t-1}^2$ differ in only one observations. They are approximathy equal.

Therefore $\Sigma\hat{u}_{t-1}^2 \approx \Sigma\hat{u}_t^2$

$$d \approx 2\left(1 - \frac{\Sigma \hat{u}_t \hat{u}_{t-1}}{\Sigma \hat{u}_t^2}\right) \quad \text{......... (3)}$$



H_o = No positive autocorrelation
H_0^* = No negative autocorrelation

Now, le us take $\hat{\rho} = \frac{\Sigma \hat{u}_t \hat{u}_{t-1}}{\Sigma \hat{u}_t^2}$ .... (4)

as the sample first-order coefficient of autocorrelation, an estimator of ρ.

$$d \approx 2\left(1 - \hat{\rho}\right) \quad \text{....... (5)}$$

But since $-1 \le \rho \le 1$ implies that

$$O \le d \le 4 \quad \text{....... (6)}$$

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

When, $\hat{\rho} = O$ $d = 2$, No serial correlaton

$\hat{\rho} = +1$ $d = O$ Perfect positive correlation

$\hat{\rho} = -1$ $d = 4$ Perfect negative correlation.

Mechanism of Durbin-Watson Test :

1) Run the OLS regression and obtain the residuals.
2) Compute d from1.
3) For the given sample size and given number of explanatory variables, find out the critical d_l and d_u values.
4) Now follow the decision role given in the table.

Table 1 :

Null hypothesis	Decision	If
No positive autocorrelation	Reject	$0 \angle d \angle d_L$
No positive autocorrelation	No decision	$d_l \angle d \leq d_u$
No negative autocorrelation	Reject	$4-d_L \angle d \angle 4$
No negative correlation	No decision	$4-d_u \leq d \leq 4-d_2$
No autocorrelation, Positive/negative	Do not reject	$d_u \angle d \angle 4-d_u$

The drawback of the Durbin-Watson d test is that, if it fall in the indecisive zone, one cannot difine whether 1st order autocorrelation exist or not. So many author modified 'd' test and given the following decision.

1. $H_0: \rho=0$ versus $H_1: \rho>0$ reject H_0 at α level if $d \angle d_u$. That is there is statistically significant positive autocorrelation.
2. $H_0: \rho=0$ versus $H_1: \rho<0$, reject H_0 at α level if the estimated $(4-d) \angle d_u$. That is there is statistically significant evidence of negative autocorrelation.
3. $H_0: \rho=0$ versus $H_1: \rho \neq 0$ reject H_0 at 2α level if $d \angle du$, or $(4-d)$ $\angle du$ that is, there is statistically significant evidence of autocorrelation, positive or negative.

Durbin-Watson developed so-called h test to test serial correlation in such model where lagged values are available.

A General Test for Autocorrelation :

The Breush-Godfrey (BG) Test : Breush-Godfrey have developed a test of autocorrelation that is general in the sense that it allows for (i) non stochastic regressors, such as the lagged values of the regressand (ii) higher order autocorregressive schemes, such as AR(1), (iii) and AR(3) simple or higher order moving average of white noise error terms.

Two variable regression model to illustrate the test–

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \text{........ (1)}$$

Assume that the error term U_t follows the ρ^{th} order autoregressive, AR(ρ), scheme as follows-

$$U_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + \varepsilon_t \qquad \text{........ (2)}$$

Where ε_t – white noise error term.

The null hypothesis H_0 to be tested is that

$$H_0 : \rho_1 = \rho_2 = ...\rho_p = 0 \qquad \text{......... (3)}$$

That is, there is no serial correlations of any order. The BG test involves the following steps :-

1) Estimate (1) by OLS and obtain the residuals $\hat{u}_i$.

2) Regress $\hat{u}_t$ on original X, and $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots\ldots \hat{u}_{t-p}$. Where the latter are lagged values of the estimated residuals in step1.

$$\hat{U}_t = \alpha_1 + \alpha_2 x_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 u_{t-2} + \ldots . \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_x \quad \ldots\ldots\ldots (4)$$

and obtain R^2 from this (auxillary) regression.

3) If the sample size is large (technically, infinite), Breusch and Godfrey have shown that

$$(n-p)R^2 \approx \chi^2_\rho \quad \ldots\ldots\ldots (5)$$

That is, asymptotically, n-p times the R^2 value obtained from the auxilliary regression (4) follow chi-square distribution with ρ df. If (n-p). R^2 exceeds Chi-square value at the choosen level of significance, we reject the null hypothesis, in which case, at least one rho (2) is statistintacally significantly different from zero.

Remedial Measures :

Model Mis-Specification versus pure Autocorrelation :

Sometimes autocorrelation arises due to mis specification of the model. This mostly happens in case of time series that trend of the series were ommited.

Let the example,

$$\hat{Y}_t = 29.5192 + 0.7136\, X_t \quad \ldots\ldots. (2)$$

$$Se = (1.9423)(0.0241)$$

$$t = (15.1977)\ (29.6066)$$

$$r^2 = 0.9584\ \ d = 0.1229\ \hat{\sigma} = 2.6755$$

Now if we include the trend variable then we get-

$$\hat{Y}_t = 1.4752 + 1.3057\, X_t - 0.9032\, Y$$

$$Sc = (13.18)\ \ (0.2765)\ \ \ (0.4203) \quad \ldots\ldots (2a)$$

$$t = (0.119)\ \ (4.7230)\ \ \ (-2.1490)$$

$$R^2 = 0.9631 \quad d = 0.2046$$

The intrepretation is strait forward over time. The index of real wage has been decreasing by about 0.90 units per year. The interesting point with allowing for trend variable, the d value is still very low, suggesting that (2a) suffer from pure autoccorrelation not necessarily from specification error. To test the correct specification we regress Y on X and X^2 to test the posibility that the real wage index may be nonlinearly related to the productivity index.

Correction for Pure Autocorrelation The Method of Generalised Least Square (GLS):

As the OLS estimators are inefficient in presence of autocorrelation, we may need to solve the problem. It depends upon nature of interdependence among the distrubances.

Let us consider two variable regression model :

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \text{....... (3.1)}$$

Assuming that the error term follow the AR(1) scheme, namely

$$U_t = \rho u_{t-1} + \varepsilon_t \quad -1 \angle \rho \angle 1 \qquad \text{........ (3.2)}$$

Now we consider two cases, (1) ρ is known (2) ρ is not known but has to be estimated.

When p is known :-

If co-efficient of the first-order autocorrelation is known, the problem of autocorrelation can easily be solved. If 3.1 hold true at time 't' and also at (t-1).

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad \text{.......... (3.3)}$$

Multiplying ρ both sides of (3.3) by ρ, we obtain,

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \qquad \text{......... (3.4)}$$

Now substracting 3.4 from 3.1 gives

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \qquad \text{....... (3.5)}$$

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

Now we can rewrite 3.5 as-

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \qquad \text{........... (3.6)}$$

Where $\beta_1^* = \beta_1(1-\rho)$; $Y_t^* = (Y_t - \rho Y_{t-1})$; $X_t^* = (X_t - \rho X_{t-1})$, $\beta_2^* = \beta_2$

Since error of 3.6 satisfies the usual OLS assumptions, we can apply OLS to the transformed variable Y* and X* and obtain estimators will all properities, namely BLUE. Now GLS is nothing but OLS applied to the transformed model that satisfies the classical assumptions.

Regression 3.5 known as generalised, quasi, difference equation. In the difference process we lose observations. To avoid this lose of observations, the first observations on Y and X is transformed as follows.

$Y_1\sqrt{1-\rho^2}$ and $X_1\sqrt{1-\rho^2}$ -This transformation known as the Prais Winstern transformation.

When R is not known : The First Difference Method :

Since ρ lies between O and ± 1. One can start from the extreme position. When $\rho = 0$ no serial autocorrelation, when $\rho = 1$ perfect positive and $\rho = -1$ perfect negative autocorrelation. At first we run a regression assuming no autocorrelation, then let the Durbin-Watson or other test to justify the assumption. When $\rho \pm 1$, the generalised difference equation.

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

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \qquad \text{........ 3.7}$$

Now error term ε_t – free from serial autocorrelation.

But it is interesting that in the first difference model, there is no intercept term. We have regression through the origine. But if we forget to drop the intercept term, then the model is–

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t \qquad \text{...... 3.8}$$

The original model must have a trend in it and β_t represent the coefficient of the trend variable. The accidental 'benifit' to introduce the intercept term is to detect for the trend variable in the original model.

Another important aspect with the transforming to first difference method is that the error term series became stationary i,e

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

$(u_t - u_{t-1}) = \Delta u_t = \varepsilon_t$, the point is that the original time series was non stationary but first difference became stationary.

The first difference method may be appropriate if " ρ " is high and "d" is low. It is valid only when $\rho = 1$. To test it there are β - Webb test, to test the hypothesis that $\rho = 1$. The test statistic they used called g statistic which is define as follows-

$$g = \frac{\Sigma_2^n \hat{e}_t^2}{\Sigma_1^n \hat{u}_t^2}$$

$\hat{u}$ - OLS residual from the original regression.

e_t - residual from 1st difference regression (keep in mind there are no intercept in the 1st difference model.)

To test significance of a statistic, assuming that the level form regression contains the intercept terms, we can use Durbin-Watson table. Except that now null hypothesis is that $\rho = 1$ rather than Durbin-Watson hypothesis that $\rho = o$.

1.6 Summing Up

In this unit, we have learned about Generalized Least Square estimation methods. This method is mainly used when the assumptions of the Classical Linear Regression model are not fulfilled. After applying GLS the Regression model fulfils all the classical assumptions and then the estimation is done.

Again we have also learned the problems of Heteroscedasticity and Autocorrelation. Here, we have discussed about their causes, various detection processes of the two along with graphical methods.

1.7 Self-Assessment Questions

1. GLS is OLS on the transformed variables that satisfy the standard least squares assumptions. Discuss.

2. What are the causes of Heteroscedasticity. Discuss any of the tests of detecting Heteroscedasticity.

3. Write the main causes of Autocorrection. What happens when we apply OLS in the presence of Autocorrelation?

1.8 References/Suggested Readings

1. Johnston, J., "Econometric Methods", McGraw Hill.
2. Gujarathi. D., "Basic Econometrics", McGraw Hill.
3. Pindyck and Rubinfeld, "Econometric Models and Econometric Forecasts", McGraw Hill.
4. Greene, William, "Econometric Analysis", Macmillan.
5. Johnston and Dinardo, "Econometric Methods", McGraw Hill.

* * *

Unit-2
NON-LINEAR ESTIMATION

Contents:

2.0 Introduction
2.1 Objectives
2.2 Meaning and Structure of Non-Linear Regression Model
2.3 Estimation of the Non-Linear Regression Model
  2.3.1 Iterative Linearization Model
  2.3.2 Models with Binary Choice Dependent Variables
  2.3.3 Logit Model
  2.3.4 Probit Model
  2.3.5 Logit and Probit Model : A Comparision
2.4 Summing Up
2.5 Self Assessment Questions
2.6 References/Suggested Readings

2.0 Introduction

The major emphasis of this book is on linear regression, that is model that are linear in the parameter/ model that can be transformed so that they are linear in parameters. On occasion, however, for theoretical or empirical basis (reason) we have to consider models that are non-linear in parameter.

2.1 Objectives

This unit mainly aims to illustrate the concept of non-linear regression model and its application-

- Estimation of non-linear regression model;
- Using binary choice dependent variable model in econometrics; and
- Concepts about Logit and probit model and its usefullness.

2.2 Meaning and Structure of Non-Linear Regression Model

In this book we will basically discuss about the models that are linear in parameters, but they may or may not be linear in variables. Some model look like non linear in the variables but are inherently or intrinsically linear because with suitable transformation, they can be made linear in the parameter regression model.

Example : $Y = a + b \ln x$ ....(i)\

Another example : $Y = \beta_1 + \beta_2 x_1 + \beta_3 x_2^2$,

but we can linearise it as-

$Y = \beta_1 + \beta_2 x_1 + \beta_3 z$ Where $z = X_2^2$

But if such model cannot be linearized in the parameters they are called instrinsically nonlinear regression models. We are talking about inherenthy or instrinsically non linear model (LNRM).

Intrinscically linear Model : (C-D) production function-

$$Q = AL^{\alpha}K^{\beta}$$

$$\ln Q = \ln A + \alpha \ln L + \beta \ln K$$

$$Y = \nu + \alpha X_2 + \beta X_3$$

Q = Output, L = labour input

K = Capital input A= Constant

Intrinsically non linear function is the constant elasticity of substitution (CES) production function -

$$Y_i = A[\delta k_i^{-\beta} + (1-\delta)L_i^{-\beta}]^{-1/\beta}$$

Y= output, K=Capital input, L= labour input A= scale parameter, δ=distribution parameter ($0< \delta <1$), and β=substitution parameter ($\beta \geq -1$).

Estimation of NLRM :

To undertand the difference of estimation of NLRM, consider the following two models -

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad ....4.1$$

$$Y_i = \beta_1 e^{\beta_2 X_i} + u_i \qquad ....4.2$$

We all know 4.1 is a linear regression model and 4.2 is a nonlinear regression model. Model 4.2 known as exponential regression model and often used to measure growth of a variable, such as population, GDP or money supply.

Now suppose we consider to estimate the parameters of the two models by OLS. And we try to minimise the RSS of 4.2 . The normal equation we get are as follow : (By OLS estimation method)

$$\Sigma Y_i e^{\beta_2 X_i} = \beta_1 e^{2\hat{\beta}_2 X_i} \qquad .... 4.3$$

$$\Sigma Y_i x_i e^{\beta_2 x_i} = \hat{\beta}_1 \Sigma x_i e^{2\hat{\beta}_2 x_i} \qquad .... 4.4$$

The normal equation model of nonlinear regression have the unknowns (the $\hat{\beta}s$) both on left hand side of the equation as well as on the right hand side. As a consequence, we cannot obtain explicit solution of the unknowns in terms of known quantities. Incidently, OLS applied to a nonlinear regression model is called nonlinear least squares (NLLS).

2.3 Estimation of the Non-Linear Regression Model

2.3.1 Iterative Linearzation Model

In this model we linearize a nonlinear equation around some initial values of parameters. The linealized equation is then estimated by OLS and the initially chosen values are adjusted. These adjusted values are used to relinearize the model, and again is estimated it by OLS and readjust the estimated values. The process is continued untill there is no substantial change in the estimated values from the last couple of interations. This process is known as "Taylor series expansion" from calculas.

Let Y=f(X)+u, this can be approximated around x=a, and using tylor's is espansions–

$$Y = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{L^2} f''(a) + \frac{(x-a)^3}{L^3} f'''(a) + \ldots\ldots\ldots + u$$

now ignoring the term involving 2nd and higher order differentiation, The taylor's expansions becomes–

$$Y \approx f(a) + (x-a)f'(a) + u$$

$$Y = \underbrace{\{f(a) - af'(a)\}}_{} + \underbrace{\{xf'(ca)\}}_{} + u \quad \Big| \approx = \text{approximaly}$$

$$Y = \quad C \quad\quad mx$$

$\because$ Y= mx+c+u or c+mx+u, is a linear regression model.

Now we can expand it for 'K' variable regression model.

$$\text{Let } Y = f(X_1, X_2, \ldots\ldots X_n), X = a$$

$$\therefore Y = f(a_1, a_2 \ldots\ldots a_n))$$

Now

$$Y = f(a_1, a_2, \ldots\ldots a_n) + (X_1 - a_1)\frac{\partial f}{\partial X_1}\{a_1 a_2 \ldots\ldots a_n) +$$

$$(x_2 - a_2)\frac{\partial}{\partial X_2} f(a_1, a_2 \ldots a_n) + (X_3 - a_3)\frac{\partial f}{\partial X_3}(a_1, a_2 \ldots a_n) +$$

term involving second and higher differtiation

Now ignoring 2nd & higher order we get-

$$Y = f(a_1 a_2 \ldots a_n) + \Sigma(x_i - a_i)\frac{\partial}{\partial x_i} f(a_1 a_2 \ldots\ldots a_n)$$

Application of this Function :

The Function is non Linear inherently. For $\beta_{1\ldots\ldots}\beta_k$ using Taylor's expansion and ignoring term including 2nd and higher ordre we obtain-

$$Y = F\left(\bar{\beta}_1, \bar{\beta}_2 \ldots \bar{\beta}_k\right) + \sum_{i=i}^{k}\left(\beta_i - \bar{\beta}_i\right)\frac{\partial f}{\partial \beta_i}(\bar{\beta} \ldots\ldots \bar{\beta}_k) + U$$

Where $\bar{\beta}_1, \bar{\beta}_2 \ldots \bar{\beta}_k$ are some initial value of $\overset{*}{\beta}_1, \overset{*}{\beta}_2 \ldots \overset{*}{\beta}_k$ ..

Now

$$\underbrace{Y - \{F\left(\bar{\beta}_1, \bar{\beta}_2 \ldots \bar{\beta}_k\right) + \sum_{i=1}^{k} \bar{\beta}_i \frac{\partial}{\partial \bar{\beta}_i} f\left(\bar{\beta}_i \ldots \bar{\beta}_k\right)}_{Z_t} = \sum_{i=1}^{k} \beta_i \underbrace{\frac{\partial}{\partial \bar{\beta}_i} f\left(\bar{\beta}_i \ldots \bar{\beta}_k\right)}_{w_t}$$

$Z_t = \Sigma \beta_i w_{it}$

or $Z_t = Z_t = \beta_1 w_{1t} + \beta_2 w_{2t} + \ldots\ldots + \beta_k w_{kt} + u$ – (A)

Now, Model A is a linear model.

The advantage of this method is that value of the final estimation can be used for testing significane of the regression coefficient seperately by using 't' values. We can also test the significance of the coefficient jointly by using 'F' value.

R^2 value will not be a measure of goodness of fit and R^2 will measure what percentage of 'Z' is explained by the model.

But we are interested about Y_t of the original model. The R^2 known as -

$$R^2 = 1 - \frac{\Sigma\left(Y_t - \hat{Y}_t\right)^2}{\Sigma\left(Y_t - \bar{Y}_t\right)^2} = 1 - \frac{RSS}{TSS}$$

2.3.2 Models with Binary Choice Dependent Variables

This model is used to explaine and predict the choice of an individual variable with two alternative decission or choices. We have to used a dummy variable–

Y=1 for the particular choice

Y=0 for the other.

The particular choice influenced by a number of factors may be captured in the form of variable $X_1 \ldots X_n$.

$Y_t = f(X_1 X_2 \ldots X_n)$

Suppose,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k x_{kt} + u_t$$

$$= x_t \beta + u_t$$

Let Y can take '1' and '0' with probability P and (1-P)

$$E(Y_t) = 1xP + 0x(1-P) = P$$

$$\because 0 \le P \le 1, \therefore 0 \le E(Y_t) \le 1$$

Incase of Linear function, these is no guarantee that $E(Y_t)$ will lie within 0 and 1. So, we transform the function as 'F' such that $F(X_t \beta)$. i,e

$Y_t = F(X_t \beta) + u_t$, now $F(X_t \beta)$ can lie betwee 0 to 1 as $x + \beta$ goes from $-\alpha$ to $+\alpha$. Here F is the cumulative distribution function.

Let X is a continuous random variable then PDF of X is denoted by f(X). Now if X is a discrete random variable then we get specific value of probability mass function for a given value of Xi, e P (X=a) =f (a)

But incase of continuous random variable we does not obtain specific value for it, i,e P (x=a) is not relevant because x lies between certain limit say a and b, that is $P(a \angle x \angle b) = \int_a^b f(x)dx$

Now f(x) is called a distribution function and it shows the cumulative frequency upto X=a, then

$$P(x \leq a)$$

Frog Fig

$$f(a) = \sum_{x=-\alpha}^{a} f(x)$$

$$= \int_{-\alpha}^{a} f(x)dx$$

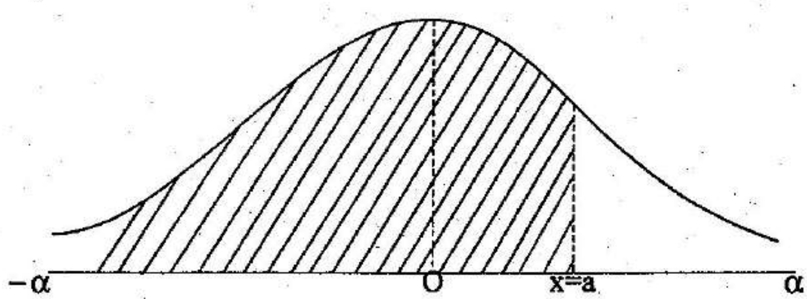
Here $f(-\alpha) = 0$

$$f(\alpha) = 1 \quad \text{i,e} \quad f(x) = \int_{-\alpha}^{\alpha} f(x)dx = 1$$

If we increase more and more of x then the value of distribution f(x) will be increased, so it is a non decreasing function :

Fig for function

$$f(a) = \sum_{x=-\alpha}^{a} f(x)$$



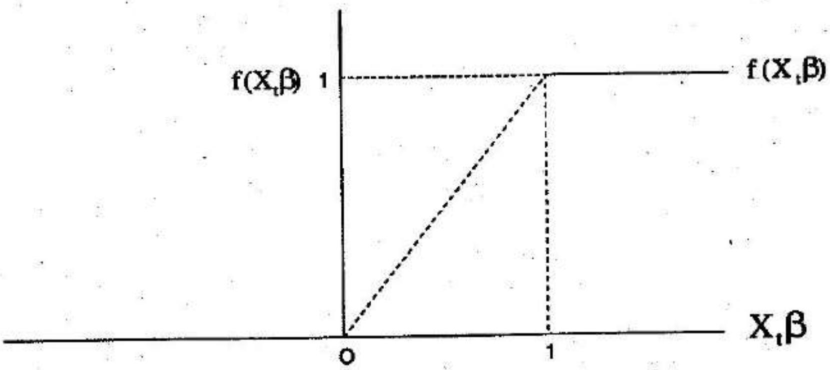
Now under the linear probability model.

$$y_t = f(X_t\beta) + u_t, \text{ but } f(X_t\beta) = 0, \text{ if } X_t\beta < 0$$

$$= X_t, \text{if} \quad 0 \leq X_t\beta \leq 1$$

$$= 1, \text{i} \quad \int X_t\beta > 1$$

The above distribution of $f(X_t\beta)$ can be diagramratically represent as follows–



The LPM is used in Linear regression model because if we use OLS or GLS in non-linear regression model then it will be very complex. In LPM the problem of heteroscedasticity exists as all the points are concentrated usually on Y=1 or Y=0.

2.3.3 Logit Model

We take the example of home ownership in relation to income, the LMP model-

$$P_i = E(Y=1|X_i) = \beta_1 + \beta_2 X_i \qquad ...... (6.1)$$

X = income, Y= 1 means family own house. Now following representation of home ownership-

$$P_i = E(Y=1|X_i) = \frac{1}{1+e^{-\beta_1+\beta_2 X_i}} \qquad ...... (6.2)$$

We can reasite it as -

$$P_i = \frac{1}{1+e^{-Z_i}} = \frac{e^{Z_i}}{1+e^{Z_i}} \qquad ...... (6.2)$$

Where $Z_i = \beta_1 + \beta_2 X_i$ ...... (6.3)

Equation 6.3 represent what is known as the (cumulative), logistic distribution function. Here z_i range from $-\alpha$ to α, P_i is non linearly related to Z_i. But we cannot use OLS procedure as it is a non linear the model in terms of X and also β.

If P_i, is the probability of owning house, then (1-Pi) is the probability of not owning the house–

$$1 - P_i = \frac{1}{1+e^{Z_i}} \qquad .... (6.4)$$

Therefore we can write

$$\frac{P}{1-P_i} = \frac{1+e^{zi}}{1+e^{-zi}} = e^{zi} \qquad .... (6.5)$$

Now, the $P_i/(1=P_i)$ is simply the odds ratio in favour of owning a house- the ratio of the probability that a family will own a house to the probability that it will not own a house.

Now, if we take natural log of (6.5) we obtain a very interesting result namely-

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i$$

$$=\beta_1+\beta_2 X_i \qquad \text{......... (6.6)}$$

that L, the log of the odds ratio, is not only linear in X, but also linear in parameters. 'L' is called the logit and hence name logit model for model like 6.6.

2.3.4 Probit Model

In some applications, the normal CDF has been found useful. The estimating model that emerge from normal CDF such as

$$F(x) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

is popularly known as the probit model.

F is the standard normal CDF, which is written explicity in the present context with an example- To motivate the probit model, assume that in our home ownership example decision of the i^{th} family to own a house or not depend on an unobservable utility index I_i, that is determined by one or more explanatory variables, say income X, in such a way that the larger the value of the Index Ii, the greater the probability of the family owing a house.

We express the Index I_i as-

$$I_i = \beta_1+\beta_2 X_i \qquad \text{....... (7.1)}$$

Where X_i is the income of the ith family. Given the assumption of normality, the probability that I_i, is less than or equal to I_i can be computed from the standardized normal CDF as

$$P_i = P(Y=1\,|\,x) = P(I_i^* \le I_i) = P(Z_i < \beta_1+\beta_2 x_i) = F(\beta_1+\beta_2 x_i) \quad \text{....(7.2)}$$

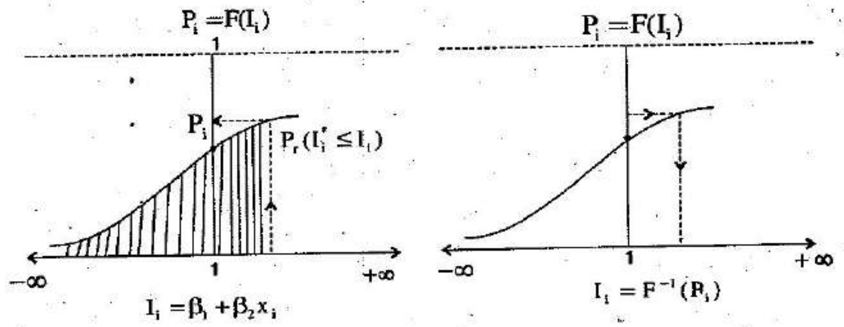
Where P(Y=1/x) means the probability that an event occur given the value's of the x, or explanatory variables and where Zi is the standard normal variable i,e

$$Z \sim N(0, \sigma^k)$$

Now, $$F(I_i = \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{Ii} e^{-z^2}/zdz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{\beta_1+\beta_2 x_i} e^{-z^2/zdz}$$

Since P represemts the probility that an event will occur, here the probability of owing a house, it is measured by the area of the standard normal curve from $-\alpha$ to Ii as shown in Fig bellow-



(a) (b)

a) Given I_i, read F_i from the ordinate

b) Given F_i, I_i read from the abscissa

2.3.5 Logit and Probit Model : A Comparision

Even the logit and probit model gives us similar qualitative answer. Our question is between Logit and Probit which model is preferable (?) In most application the model are quite similar the main difference is that thelogistic distribution has slightly fattertails, as shown in figure bellow. The conditional probability Pi, approaches Zero at a slower rate in logit than in probit. Therefore, ther is no compelling reason to choose one over the other. In practice many researcher choose the logit model because of its comparative mathematical simplicity.

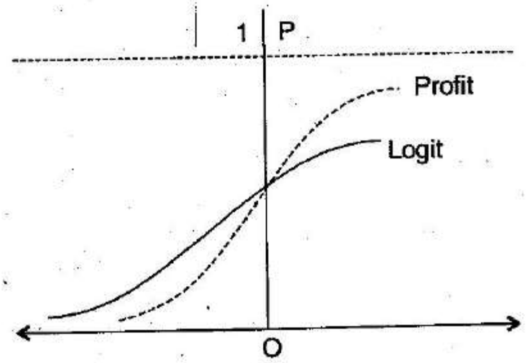


Fig - Logit and profit cumulative distribution

2.4 Summing Up

This unit was about Non-Linear Regression models where we have basically learned about the meaning of non-linearity in Econometric models. Estimation of NLRM through various models viz. Iterative Linearization Model, Binary choice Model, Logit and Probit model. Among the four, Logit and Probit models gives us almost similar qualitative answers but due to its mathematical simplicity Logit has become more popular among the Economists.

2.5 Self Assessment Questions

1. Show with the help of example why models which are not intrinsically linear are not a big problem in Econometrics.
2. Discuss the Logit model showing how the non-linear model is estimated with the help of Logit.
3. What is the main difference between Logit and Probit Modes?

2.6 References/Suggested Readings

1. Johnston, J., "Econometric Methods", McGraw Hill.
2. Gujarathi. D., "Basic Econometrics", McGraw Hill.
3. Pindyck and Rubinfeld, "Econometric Models and Econometric Forecasts", McGraw Hill.
4. Greene, William, "Econometric Analysis", Macmillan.
5. Johnston and Dinardo, "Econometric Methods", McGraw Hill.

* * *

Unit- 3
DISTRIBUTED LAG MODELS

Contents:

3.0 Introduction
3.1 Objectives
3.2 Role of Time or Lag in Econometrics
3.3 Reasons for Lags
3.4 Estimation of the Distributed Lags Model
  3.4.1 Ad Hoc Estimation of Distributed Lag Model
  3.4.2 Koyck Model
  3.4.3 Partial Adjustment Model
3.5 Estimation of Autoregressive Model
3.6 The Model of Instrumental Variable
3.7 Detecting Autocorrelation in Auto-Regressive Model
3.8 Summing Up
3.9 Self Assessment Questions
3.10 References/Suggested Readings

3.0 Introduction

Time lag play on important part of the economic applications. In formulating an economic model, we have to consider both the current as well as lag of time (past) value as incase of explanatory variable. In this unit we deal with that types of problem.

3.1 Objectives

Autoregressive and distributed lag models, are used extensively in econometric analysis, and in this chapter we take a look at such model with a view of finding out-

- What is the role of lags in economics
- What are the reason for the lags
- Is there are any theoritical justification for the commonly used lagged models in emperical econometrics
- What is the relationship, if any, between autoforegressive and distributed-lage models? Can one derive one from other.

In Regression analysis involving time series data, If the regression model includes not only current but also the lagged (Past) values of explanatory variables (X), then it is called a distributed lag model.

$$Ex - Y_t = \alpha + \beta_o X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$$

Again if the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called an autoregressive model.

$$Ex - Y_t = \alpha + \beta x_t + \gamma Y_{t-1} + u_t$$

also known as dynamic model.

3.2 Role of Time or Lag in Econometrics

In Economics there is the dependence of a variable Y, (dependent variable) on another variable (s) X explanatory variable (s) very often, Y responds to X with a lapse of time. Such a lapse of time is called a lag.

We may write K distributed lag model-

$$Y_t = \alpha + \beta_o x_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_k X_{t-k} + u_t \qquad ...... (8.1)$$

Here- β_o - is know as the short run multiplier or impact multiplier, because it gives the change in the mean value of Y following a unit change in x in the same time period. ($\beta_o + \beta_1$) for the next period and so on. The partial sums are called interim or intermediate multipliers. Finally, after 'k' period we obtain -

$$\sum_{i=0}^{k} \beta_i = \beta_0 + \beta_1 + \beta_2 + ... + \beta_k = \beta \qquad ...... (8.2)$$

It is known as the long run or total distributed lag multiplier, provided the sum β exists-

If we define - $\beta_i^* = \frac{\beta_i}{\Sigma \beta_i} = \frac{\beta_i}{\beta}$

It is the "standardized" β_i. Partial sums of the standarized β_i, then gives the poportion of the long run, or total impact felt by a certain period.

Example of Distributive Lag Model :

The Consumption Function :

Suppose a person received a permanent salary increase of $2000 in annual pay. Now what will be effect of this increase income of the person's consumptions?

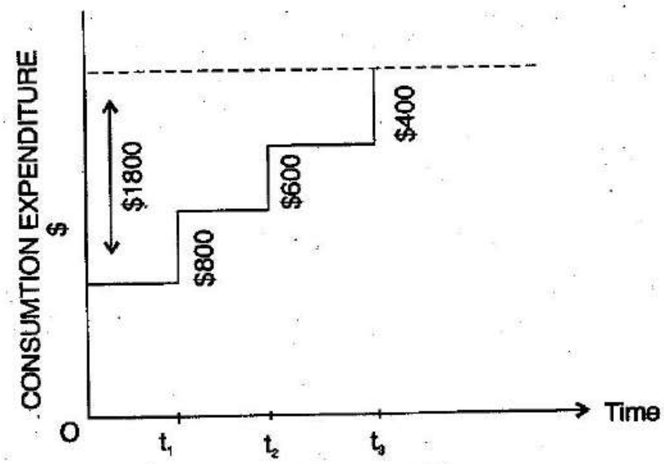
The person does not opend all increase immediately. Thus our recipients may decide to increase consumption expenditure by $ 800 in the 1st year following the salary increase, by another $ 600 in the next year, by another $ 400 in the following year, saving the remainder. By the end of the third year, the person's annual consumption expenditure will be increased by $ 1800. We can thus write the consumption function as-

$$Y_t = \text{Constant} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + u_t \qquad ....(1)$$

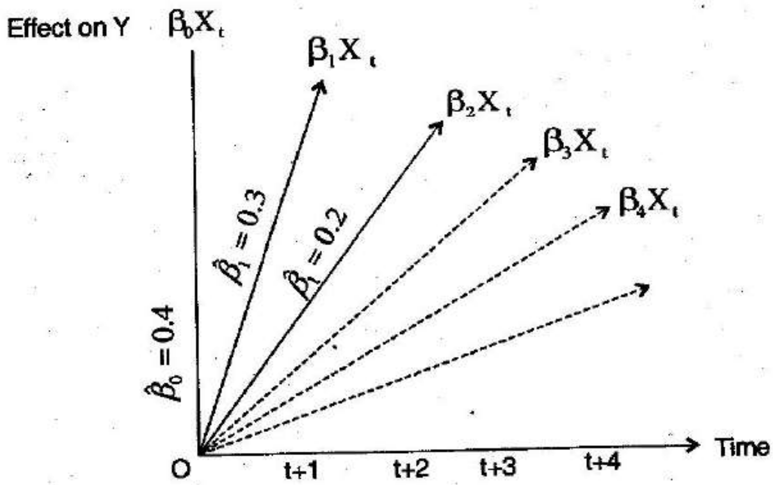
Y= consumption expenditure and X is income.

Now, equation (1) shows that the effect of an increase in income of $ 2000 is spread, or distributed over a period of 3 years. Such model is called

distributed lag models because the effect of a given cause (income) is spread over a time periods.
Using the definition of short run multiplier i,e short-run marginal propersity is 0.4 and long-run multiplier or long-run marginal propersity to consume is 0.4 and long-run multiplier or long-run marginal propersity to consume 0.4+0.3+0.2=0.9 i,e As income increase by $1 the consumer spend 40 percent in the year of increase, 30 percent in the next year and by 20 percent in the following year. diagramatic Representative–



Example of distributed-lag



The effect of a unit change in X on Y at time t and on subsequant time periods.

3.3 Reasons for Lags

There are several reasons but the three main reasons are :-

1) **Psychological Reason :** As a result of the force of habit, people do not change their consumption habits immediatly following a price decrease or perhaps income increase because the change may have some immediate disutility.
2) **Technological Reasons :** Suppose the price of capital relative to labour declines, making substitution of capital for labor economically feasible. Of course, the addition of capital takes times (gestation period). Moreover if the drop in price is expected to be temporary, they will not rush to substitute capital for labour.
3) **Institutional Reasons :** These reasons also contribute to lags. For ex:- employer often gives their employees a choice among several health insurance plans, but once a choice is made an employee may not switch to another plan for atleast one year.

3.4 Estimation of the Distributed Lags Model

Suppose we have following distributed lag model in one explanatory variable–

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-i} + \beta_2 X_{t-2} + ...+ u_t \qquad ...... (9.1)$$

We have not defined the length of the lag, that is how far back into the past we want to go. Such model is called on infinite (lag) model.

Whereas the model where the length of the lag is defined this is called finite distributed lag model. ex-k-specified model.

3.4.1 Ad Hoc Estimation of Distributed Lag Model

As we assumed X_t is nonstochastic and atleast uncorrelated with distrubance term, the X_{t-1}, X_{t-2} and so on, are non-stochastic too. Now we can apply ordinary least square method in the equation (9.1). This approach was proposed by Tin berger and Alt. They suggested that to estimate on sequential basis i,e first we regress Y_t on X_t, then on X_{t-1} and X_{t-2} and so on. The sequentional procedure stop when the regression co-efficient of the lagged variables. Start becoming statistically insignificant or atleast coefficient of one variable changes its signs as positive to negative or vice-versa.

For Example :-

Alt regressed fuel oil consumption Y on new orders X. Based on the quarterly data for the period 1930-1939, the result were following -

$$\hat{Y}_t = 8.37 + 0.171X_t$$

$$\hat{Y}_t = 8.27 + 0.111X_t + 0.064X_{t-1}$$

$$\hat{Y}_t = 8.27 + 0.109X_t + 0.071X_{t-1} - 0.055X_{t-2}$$

$$\hat{Y}_t = 8.32 + 0.108X_t + 0.063X_{t-1} + 0.022X_{t-2} - 0.020X_{t-3}$$

As we have shown in the above model that second regression as the 'best' one because last two equation the sign of X_{t-2} are changable and in the last one, X_{t-3} was negative, which may be difficult to interpret economically. Although the procedure is straight forward, it has the following drawbacks-

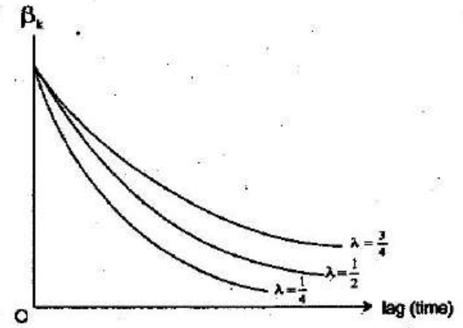
1) There are no a priori guide as to what is the maximum length of the lag.
2) Due to successive lags, there are fewer degrees of freedom left. It makes statistical inference somewhat shaky. Because the economist are not lucky to have long series data so that they can go on estimating numerous lags.
3) More importantly, in economic time series data, successive value (lags) tend to be highly correlated i,e there are presence of multicollinearity.
   It leads the standard error tend to be large in relation to the estimated coefficients. As result the 't' value decreases and lagged coefficient becomes statintically insignificant.
4) The sequential search for the lag length opens the researcher to the charge of data-mining. Here both nominal and true level of significance are used to test statistical hypothesis and it becomes an important issue of sequential research.

3.4.2 Koyck Model

Koyck has proposed an ingenious method of estimating distributed -lag models. Suppose ws start the infinite lage model 9.1. Assuming that the β's are all of the same sign. Koyck assumes that they declines geometrically as follows-

$$\beta_k = \beta_0 \lambda^k \quad k = 0,1 \quad .... (9.2)$$

Where λ, such that $0<\lambda<1$ is known as the rate of decline or decay of the distributed lag model and where $1-\lambda$ known as the speed of adjustment. What 9.2 postulates is that each successive β coefficient is numerically less than each proceeding β (this statement since $\lambda<1$), implying that as one go back into the distant past, the effect of the lag on Y_t^E, become progressively smaller, a quite plausible assumption. Koyek scheme is depticted in Fig–



Koyck Scheme (declining geometric distribution)

Features of Koyck scheme : (1) by assuming nonnegative values for λ Koyek rules out the β's from changing sign (2) by assuming $\lambda < 1$, he gives lesser weight to the distant β's than the current ones and (3) he ensure that the sum of the β's, which gives the long run multiplier, is finit namely,

$$\sum_{k=0}^{\alpha} \beta_k = \beta_0 \left(\frac{1}{1-\lambda} \right) \qquad \text{....... (9.3)}$$

As a result of (9.2), the infinite lag model (9.1) may be written as -

$$Y_t = \alpha + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + ... + u_t \quad \text{....... (9.4)}$$

But now Koyek suggests an ingeneous way out. He lags (9.4) by one period-

$$Y_{t-1} = \alpha + \beta_0 x_{t-1} + \beta_0 \lambda x_{t-2} + \beta_0 \lambda^2 x_{t-3} + ... + u_{t-1} \quad \text{....... (9.5)}$$

Then he multiplies (9.5) by λ to obtain -

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \beta_0 \lambda^3 x_{t-3} + ... + \lambda u_{t-1} \qquad \text{...... (9.6)}$$

Substracting 9.6 from 9.4 we gets–

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 x_t + (u_t - \lambda u_{t-1}) \qquad \text{...... (9.7)}$$

or rearranging–

$$Y_t = \alpha(1-\lambda) + \beta_0 x_t + \lambda Y_{t-1} + v_t \qquad \text{...... (9.8)}$$

where $v_t = (u_t - \lambda u_{t-1})$, a moving average of u_t and u_{t-1}

This procedure is known as Koyek transformation. Before that we had to estimate α and infinite number of β's. But now we have to estimate only three unknowns, α, β_0 and λ .

Features :

1) We started with distributed-lag model but ended up with an autoregressive model because Y_{t-1} appear as one of the explanatory variables.
2) The presence of Y_{t-1} is likely to create some statistical problems Y_{t-1} and Y_t, is stochastic, which means we have a stochastic explanatory variable in the model.
3) In the original model we have disturbance term u_t but in the transformation model, the disturbance term is $v_t = (u_t - \lambda u_{t-1})$. The latter shows there are serial correlation in addition with stochastic explanatory variable Y_{t-1}.
4) The presence of lagged Y violates one of the assumption underlying Durbin-Watson d test. Therefore, we will have to develop an alternative to test for serial correlation in the presence of lagged Y. It is Durbin-h test.

The Median Lag :

The median lag is the time required for the first half, or 50 percent, of the total change in Y following a unit sustained change in X.

Koyck Model : Median lag $= -\frac{\log 2}{\log \lambda}$

Thus if $\lambda = 0.2$, the median lag is 0.4306

The Mean Lag :

Provided all β_k are positive, the mean, or average, lag is defined as–

$$\text{Mean lag} = \frac{\Sigma_0^\infty k\beta_k}{\Sigma_0^\alpha \beta_k}$$

Which is simply the weighted average of all the lags invloved, with the respective β coefficients serving as weights.

Koyck Model : Mean lag $= \frac{\lambda}{1-\lambda}$

Thus if $\lambda = \frac{1}{2}$ the mean lag is 1.

3.4.3 Partial Adjustment Model

The adaptive expectation model is one way of rationalizing the Koyek Model. More Nerlove presided Stock adjustment or Partial adjustment model (PAM). To illustrate the model; consider the flexible accelator model of economic theory. It assumes that there is an equalibrium, optional, desired, or long run amount of capital stock needed to produce a gievn output under the given state of technology and rate of interest etc. For simplicity assume that this desired state of capital Y_t^* is a linear function of output X as follows–

$$Y_t^* = \beta_0 + \beta_1 x_t + u_t \qquad \text{.......... (10.1)}$$

Since disered stock of capital is not directly observable, Nerlove postulate the following hypothesis known as partial adjustment, or stock adjustment hypothesis

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1}) \qquad \text{...... 10.2}$$

Where δ such that $0 \angle \delta \angle 1$ known as the coefficient of adjustment and where $Y_t - Y_{t-1}$ = actual change and $(Y_t^* - Y_{t-1})$- desired change.

Since, $Y_t - Y_{t-1}$, the change in capital stock between two periods, is nothing but investment, 10.2 can be writen as–

$I_t = \delta (Y_t^* - Y_{t-1})$ ....(10.3) when It= Investment in time period t

Equation 10.2 state that the actual change of capital stock in any given time period 't' is some fraction δ of the desired change for the period.

If $\delta = 1$ meanse that actual stock of capital is equal to the desired stock.

If $\delta = 0$ it means that nothing changes since actual stock at time 't' is the same as that observed in the previous time period.

The formula - (10.2) can be written as –

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

Now substitution (10.1) into (10.4) we dstain

$$\left.\begin{aligned} Y &= \delta(\beta_0 + \beta_1 X_t + u_t) + (1-\delta)Y_{t-1} \\ &= \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta u_t \end{aligned}\right\} \text{...........10.5}$$

The model is called the parfial adjustment model (PAM). Since (10.1) represent long run, or equlibrium, demand for capital stock, 10.5 can be called short run capital stock since in the short run the existing capital stock may not necessarity be equal to its long run level.

Combination of Adaptive Expectations and Partial Adjustment Models :

Consider the following model :-

$$Y_t^* = \beta_0 + \beta_1 X_t^* + u_t$$

Y_t^* = desired capital stock and X_t^* = expected level of output.

Since both Y_t^* and X_t are not directly observable, . One could use the partial adjusment mechanism for Y_t^* and the adaptive expectation model for X_t^* to arrive at the following estimating equation.

$$Y_t = \beta_0\delta\gamma + \beta_1\delta\gamma X_t + [(1-\gamma) + (1-\delta)]Y_{t-1}(1-\delta)(1-\gamma)y_{t-2} + [\delta u_t - \delta(1-\gamma)u_{t-1}]$$

$$= \alpha_0 + \alpha_1 X_t + \alpha_2 X_{t-1} + \alpha_3 Y_{t-2} + v_t$$

Where $v_t = \delta[u_t - (1-\gamma)u_{t-1}]$ This model too is autoregressive, the only difference from the purely adaptive expectation model being that Y_{t-2} appears alongwith Y_{t-1} as an explanatory variable. Like Koyck and the AE models, the errors term follows a moving average process. Another feature of the model is that although the model is linear in the α's, it is nonlinear in original parameter.

3.5 Estimation of Autoregressive Model

From our discussion this far we have the following three models -

Koyck-

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (u_t - \lambda u_{t-1}) \quad \text{.......(11.1)}$$

Adaptive Expectation–

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + [u_t - (1-\gamma)u_{t-1}] \quad \text{.......(11.2)}$$

Partial adjustment–

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta u_t \quad \text{.......(11.3)}$$

All these models have following common form–

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + v_t \quad \text{.......(11.4)}$$

that is they are all autoregressive in nature.

Here we can not use least square methods of estimation, the reasons are two fold– The presence of stochastic explanatory variables and the possibility of serial correlation. Even if stochastic, for the application of classical leat square theory, the stochastic explanatory variable Y_{t-1} must be distributed independently of the distrubance term v_t. To determined it, we have to know the properities of v_t. Suppose the original distrubance term u_t satisfies assumptions of homoscedastcicity, no autocorrelation, unbiasedness etc. But v_t may not be so. Koyck model's error term, we can show it is serially correlated-

$$E(v_t v_{t-1}) = -\lambda\sigma^2 \qquad \text{......(11.5)}$$

Which is nonzero (unless λ happens to be zero). And since Y_{t-1} appears in the Koyck model as an explantory variable, it is bound to be correlated with v_t. As a matter of fact, it can ve shown,

$$\text{cov}[y_{t-1}, (u_t - \lambda u_{t-1})] = -\lambda\delta^2 \qquad \text{......(11.6)}$$

In Koyck model as well ab in the adaptative expectations model the stochastic explanatory variable Y_{t-1} is correlated with the error term v_t (?) As noted previously, If an explanatory variable in a regression model is correlated with the stochastic distrubance term, the OLS estimators are not only biased but also not even consistent, that is even if the sample size is increased indefinitely, the estimators do not approximate their true population values. Therefore, estimation of Koyck and adaptive expectation models by the usual OLS procedure may yeild seriously misleading results.

3.6 The Model of Instrumental Variable (IV)

We cannot apply OLS to obtain consistent estimator when Y_{t-1} is correlated with disturbance term v_t. But if this correlation is removed OLS can be applied. To accomplish this, Liviatan has proposed following solutions Let us suppose that we find a proxy for Y_{t-1} that is highly correlated with Y_{t-1} but is uncorrelated with v_t, where v_t is the error term appearing in the Koyek or adaptive expectation model. Such a proxy is called an instrumental variable (IV), Liviatan suggest X_{t-1} as the instrumental variable for Y_{t-1} and further suggests that the parameters of the regression (11.4) can be obtained

$$\left.\begin{aligned} \Sigma Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1 \Sigma X_t + \hat{\alpha}_2 \Sigma Y_{t-1} \\ \Sigma Y_t X_t &= \hat{\alpha}_0 \Sigma X_t + \hat{\alpha}_1 \Sigma X_t^2 + \hat{\alpha}_2 \Sigma Y_{t-1} X_t \\ \Sigma Y_t X_{t-1} &= \hat{\alpha}_0 \Sigma X_{t-1} + \hat{\alpha}_1 \Sigma X_t X_{t-1} + \hat{\alpha}_2 \Sigma Y_{t-1} X_{t-1} \end{aligned}\right\} - (A)$$

Notice if we were to apply OLS directly to (11.4), the usual OLS normal equation would be-

$$\left.\begin{aligned}\Sigma Y_t &= n\hat{\alpha}_o + \hat{\alpha}_1 \Sigma X_t + \hat{\alpha}_2 \Sigma Y_{t-1} \\ \Sigma Y_t x_t &= \hat{\alpha}_o \Sigma X_t + \hat{\alpha}_1 \Sigma X_t^2 + \hat{\alpha}_2 \Sigma Y_{t-1} X_t \\ \Sigma Y_t Y_{t-1} &= \hat{\alpha}_o \Sigma Y_{t-1} + \hat{\alpha} \Sigma x_t Y_{t-1} + \hat{\alpha}_2 \Sigma Y_{t-1}^2\end{aligned}\right\} - (B)$$

The difference between the two sets of normal equations should reading be apparent. Liviatan has shown that the $\hat{\alpha}$ estimated from (A) are consistent whereas those estimated from (B) may not be consistent because Y_{t-1} and $v_t[= u_t - \lambda u_{t-1}$ or $u_t - (1-\gamma)u_{t-1}]$ may be correlated whereas X_t and X_{t-1} are correlated with v_t although easy to apply in practice once a suitable proxy is found, Liviatan approach is likely to suffer from the multicolinerity problem because X_t and X_{t-1}, which enter the normal equation in (A) are likely to be highly correlated. The implication then is that although the Liviatan procedure yields consistent estimates, the estimators are likely to be inefficient.

As the finding of a good proxy always is not an easy task, so one may have to resort to maximum likelihood techniques, which are beyond the scope of the book.

3.7 Detecting Autocorrelation in Auto-Regressive Model

The serial correlation in error term v_t make estimation problem more complex. In the stock adjustment model the error term v_t did not (first order) have serial correlation if the errom term u_t in the original model was serially uncorrelated, whereas in the Koyck and adaptive expectation model v_t was serially correlated even if "u_t" was serially independent.

So the main question is how does one know if there is serial correlation in the error term appearing in the autoregressive model?

Durbin himself has proposed as large sample test of first order serial correlation in auto regressive models. This test is called the 'h statistic'.

<u>h statistic</u>-

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\text{var}(\hat{\alpha}_2)]}} \quad \text{.......... (13)}$$

Where n is the sampel size, Var $(\hat{\alpha}_2)$ is the variance of the coeffecient of the lagged $Y_t = (Y_{t-1})$ and $\hat{\rho}$ is an estimate of the first order serial correlation ρ.

For large sample, Durbin has shown that, if $\rho = 0$, the h statistic of (13) follows the standard normal distribution. That is

$$h a \tilde{s} y N(0,1) \quad \text{.......... (13.1)}$$

asy–means asymptotically

In practice one can estimate ρ as

$$\hat{\rho} \approx 1 - \frac{d}{2} \quad \text{.......... (13.2)}$$

It is interesting to observe that although we cannot use the 'Durbin d' to test for autocorrelation in autoregressive models, we can use it as an input in compuling the h statistic.

Important Features of 'h' Statistic :

1. It does not matter how many X variable or how many lagged values of Y are included in the regression model. To compute h, we need to consider only the variance of the coefficient of lagged Y_{t-1}.
2. The test is not applicable if $[n\text{Var}\hat{\alpha}_2]$ exceeds 1. In practice it usually dest not happen.
3. Since the test is a large sample test, its application in small samples is not strictly justified, as shown by Inder and Kiviet. It has been suggested that the Breusch-Godtrey (BG) test, also known as the lagrange multiplier test, is statistically more powerfull not only in large sample but also in finite or small, samples and is therefore preferable to the h test.

Let us illustrate the use of the h statintic's with our example where n=30, $\hat{\rho} \approx (1-\frac{d}{2}) = 0.4972$ and var $(\hat{\alpha}_2) = 0.0239$. Putting these values in equation 13, we get–

$$h = 0.4972\sqrt{\frac{30}{1-30(0.0239)}} = 5.1191$$

Since the 'h' value has the standard normal distribution under the null hypothesis, the probability of obtaining such a high value is very small. Recal that probability that a standard normal variable exceeds the value of ± 3 is extremly small. In the present context there is (positive) autocorrelation. Of course bear in the mind that h follows standard normals distribution asymptotically.

3.8 Summing Up

In this unit, we have discussed about distributive by models, which include Reasons for Logs, Estimations of Distributive Lag Model etc. The unit gives a precise analysis Koyek approach to Distributive Lag model where we start with a distributive-lag model but eventually end with an autoregressive model. Another model discussed in this unit is Partial Adjustment Model.

After that we have discussed about Liviatan's method of Instrumental Variables and Durbin's "h" test of Detecting Autocorrelation in Autoregressive model.

3.9 Self Assessment Questions

1. Discuss the concept of Distributive Lag Model with the help of a suitable example.
2. Discuss the Koyek approach to Distributive Lag Model.
3. What are the important features of Durbin's "h" statistic?

3.10 References/Suggested Readings

1. Johnston, J., "Econometric Methods", McGraw Hill.
2. Gujarathi. D., "Basic Econometrics", McGraw Hill.
3. Pindyck and Rubinfeld, "Econometric Models and Econometric Forecasts", McGraw Hill.
4. Greene, William, "Econometric Analysis", Macmillan.
5. Johnston and Dinardo, "Econometric Methods", McGraw Hill.

* * *

Unti-4
ANALYSIS OF TIME SERIES

Contents:

4.0 Introduction
4.1 Objectives
4.2 Statistic Process
    4.2.1 Stationary Stochastic Process
4.3 Non Stationary Time Series
    4.3.1 Non-Stationary Stochastic Process
    4.3.2 Unit Root Stochastic Process
4.4 Trend Stationary and Difference Stationary Stochastic Process
    4.4.1 Pure Random Walk
    4.4.2 Random Walk with Drift
    4.4.3 Deterministic Trend
    4.4.4 Random Walk with Drift and Deterministic Trend
    4.4.5 Deterministic Trend with Stationary AR(1) Components
4.5 Test of Stationarity
    4.5.1 ACF and Correlogram
    4.5.2 The Augmented Dickey Fuller Test
4.6 Summing Up
4.7 Self Assessment Questions
4.8 References/Suggested Readings

4.0 Introduction

We have that there are three types of data in Economics, viz. Time Series, Cross Sectional and Pooled. One of the important kind is Time Series data. In time series data we study the behaviour of one or two variables in different time periods.

4.1 Objectives

- introducing the concept of stationary stochastic process and non-stationary stochastic process differentiating between trend stationary and difference stationary stochastic process; and
- finding the various ways to test stationarity of a given time series.

4.2 Stochastic Processes

A random or stochastic process is a collection of random variables orderd in time. If we assume Y denote a random variable, and if it is continuos, we denote it as Y(t) but if it is discrete, we denote it as Y_t.

4.2.1 Stationary Stochastic Processes

A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two periods depend only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. In the time series, such a stochastic process is called weakly stationary, or covariance stationary, or second order stationary, or in wide sense, stochastic process.

To explain weak stationary, Let Y_t be a stochastic time series with these properities-

Mean	$E(Y_t)=\mu$	.....23.1.1
Variance	$Var(Y_t)=E(Y_t-\mu)^2=\sigma^2$	.....23.1.2
Co-variance	$\gamma_k=E[(Y_t-\mu)(Y_{t+k}-\mu)]$	.....23.1.3

Where γ_k, the covariance at lag k, is the co-variance between the values of Y_t and Y_{t+k}, that is two Y values K periods apart. If k=o, we obtain γ_o, which is simply the variance of Y=(σ^2), if $k=1$, γ_1 is the covariance between two adjacent values of Y.

In short, if a time series is stationary, if its mean, variance and autocovariances (at various lag) remain the same no matter at what point we measure them. That is they are time invariant. Such time rises will tend to return to mean.

4.3 Non Stationary Time Series

If a time series is not stationary in the sense just defined, it is called a nonstationary time series. In other words non-stationary time series will have a time varying mean or a time varying variance or both.

Stationary time series are so important, because if a time series is nonstationary, we can study its behaviour only for the time period under consideration. Each set of time series data will therefore be for a particular episode.

Before we move on, we mention a special type of stochastic process (time series) namely a purely random or white noise, process, as it has zero mean, constant variance σ^2, and is serially uncorrelated.

4.3.1 Non-stationary Stochastic Process

In case of non-stationary time series, one often encounters the classic example being the random walk model (RWM). Examples are, asset prices, such as stock exchange rates. We distinguish two types of random walks (1) random walk without drift (no constant or interupt term) (2) random walk with drift (i,e a constant term is present).

(1) Random Walk Without Drift :

Suppose u_t is a white noise error term with mean 0 and variance σ^2. Then the series Y_t is said to be a random walk model if.

$$Y_t = Y_{t-1} + u_t \qquad ....23.1.4$$

Random walk model (as 23.1.4) shows, the value of Y at time 't' is equal to its value at time (t−1) plus a random shock, thus it is an AR model.

Now from 23.1.4 we can write

$$Y_1 = Y_0 + u_1$$
$$Y_2 = Y_1 + u_2 = Y_0 + u_1 + u_2$$
$$Y_3 = Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3$$

In general, if the process started at some time 0 with a value of Y_0 we have.

$$Y_t = Y_0 + \Sigma u_t \qquad .....23.1.5$$

Therefore- $E(Y_t) = E(Y_0 + \Sigma U_t) = Y_0$ 23.1.6

$$\therefore Var(Y_t) = t\sigma^2 \qquad .....23.1.7$$

Here mean of Y is equal to the initial value which is constant. But as 't' increases its variance increases indefinitely, thus violating a condition of stationarity.

(2) Random Walk With Drift :

Let modify (23.1.4) as follows-

$$Y_t = \delta + Y_{t-1} + u_t \qquad .....23.1.8$$

Where is δ known as drift parameter. Then-

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \qquad .....23.1.9$$

it show that Y_t, drift upward or downward, depending on δ being positive or negative.

It can be shown-

$$E(Y_t) = Y_0 + t.\delta \qquad .....23.1.10$$

$$Var(Y_t) = t\sigma^2 \qquad .....23.1.11$$

It shows that random walk with drift model are non stationary.

4.3.2 Unit Root Stochastic Process

Let us write random walk model-23.1.4 as

$$Y_t = \rho Y_{t-1} + u_t$$

This model is known as the Markov first order autoregressive model. If $\rho = 1$, we face what is know as unit root problem. Thus, the term nonstationary, random walk, and unit root can be treated as synonymous. If however, $|0| \le 1$, that is if the absolute value of ρ is less than one, then it can be shown that the time series, Y_t is stationary.

4.4 Trend Statinary (TS) and difference Stationary Stochastic Process

Broaldy speaking, if the trend in a time series is completely predictable and not variable, we call it a deterministic trend, whereas if it is not predictable, we call it a stochastic trend. To make this defination more normal let us consider following series -

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \qquad ....23.1.12$$

Where u_t is a white hoise error term and where t (time) is measured choronolagically. Now we have the following possibilities.

4.4.1 Pure Random Walk

If in 21.2.12 $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 1$, we get

$$Y_t = Y_{t-1} + u_t \qquad .... 23.1.13$$

If is RWM without drift and it is therefore nonstationary. But if we write-

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \qquad .....23.1.14$$

became stationary, which we can call difference stationary because ΔY_t is the first difference of Y_t as noted before.

4.4.2 Random Walk With Drift

If in 23.1.12 $\beta_1 \neq 0, \beta_2 = 0 \& \beta_3 = 1$ we get

$$Y_t = \beta_1 + Y_{t-1} + u_t \qquad .....23.1.15$$

Which is a random walk drift and therefore nonstationary. If we write it as-

$$(Y_t - Y_{t-1}) = \Delta Y_t = \beta_1 + u_t \qquad .....24.1$$

This means Yt will exhibit a positive ($\beta_1 > 0$) or negative ($\beta_1 < 0$) trend. Such a trend is called a stochastic trend. Equation 23.1.14 is DSP process because the nonstationarity in Y_t can be eliminated by taking first difference of the time series.

Fig for Random walk model without drift:

As this world variance depend on time t. Thus, where $\alpha = 0$ and β_1 i,e AR proses= $Y_t = \alpha + \beta Y_{t-1} + u_t$. The time series will sealter more and more as it increases around the same mean (Y_0). Shown in the diagram below daigram–

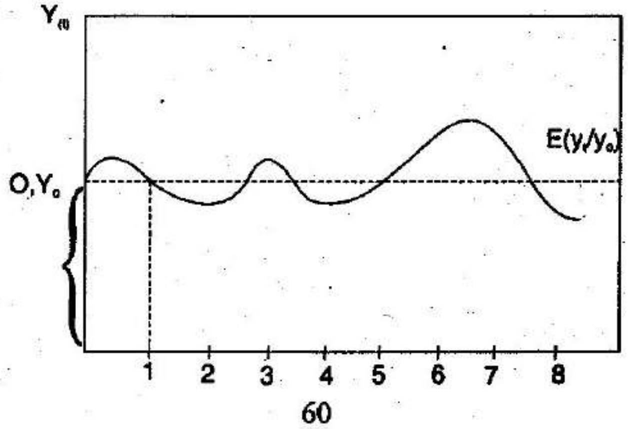
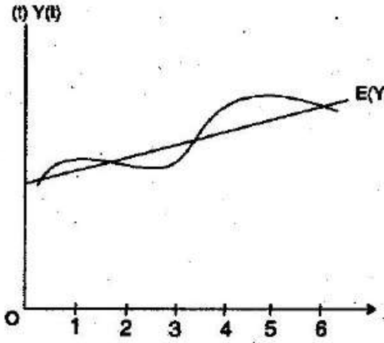


Figure for Random walk model with drift :

Fig-1 : $\alpha < 0, \beta = 1$ Fig–2 : $\alpha > 0, \beta = 1$



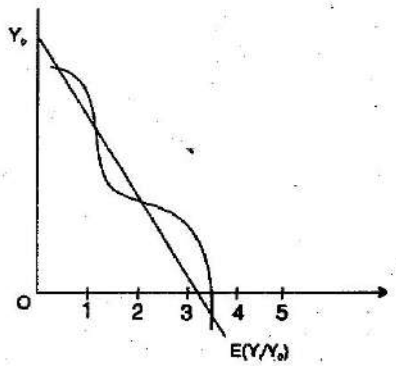
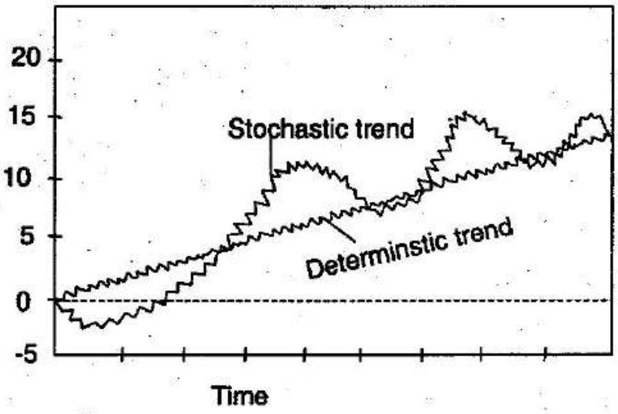


Figure For Determinstic versus stochastic trend.



4.4.3 Determinstic Trend

If in (23.1.12), $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 = 0$ then we, get-

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad ......24.2$$

which is called a trend stationary process (TSP). Although the mean of Y_t is $\beta_1 + \beta_2^t$, which is not constant, if varience = (σ^2) is constant. Once the value of β_1 and β_2 are known, the mean can be foreast perfectly. Therefore if we subtract the mean of Y_t form Y_t, the result will be stationary. Hence, the name trend stationary. The procedure of removing the (determinstic) trend called detrending.

4.4.4 Random Walk With Drift and Determinstic Trend

If in C 23.1.12, $_1\beta \neq 0, \beta_2 \neq 0, \beta_3 = 1$. we obtain

$$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + u_t \qquad .... 24.3$$

We have obtained a random walk with drift and a deterministic trend, which can be seen if we write this equations as–

$$\Delta Y_t = \beta_1 + \beta_2 t + u_t \qquad ....24.4$$

which means that Y_t is nonstationary.

4.4.5 Determinstic Trend With Stationary AR (1) Components

If on (23.1.12), $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0$ then we, get-

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \qquad ....24.5$$

Which is stationary around the determinstic trend.

4.5 Test of Stationary

With the above elaboration probably the reader has got a good idea about the nature of stationary stochastic process and their importance. In practice we face two important problems.

1) How do we find out that a given series is stationary?
2) If we find out that a given time series is non-stationary, is there a way that it can be made stationary?

The second questions answer given by following the various method latter. But the first questions answer follows the following test-

4.5.1 Autocorrelation Function (ACF) and Correlogram

One simple test of stationarity is based on the so-called autocorrelation function (ACF). The ACF at lag k, denoted by ρ_k, is defined as

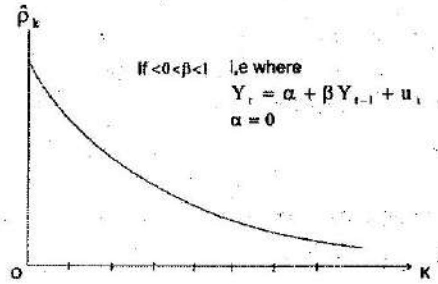
$$\rho_k = \frac{\vartheta_k}{\vartheta_0} = \frac{\text{covariance of lag k}}{\text{variance}} \qquad ..... 25.1$$

Since, covariance at lag k and variance are as defined before.

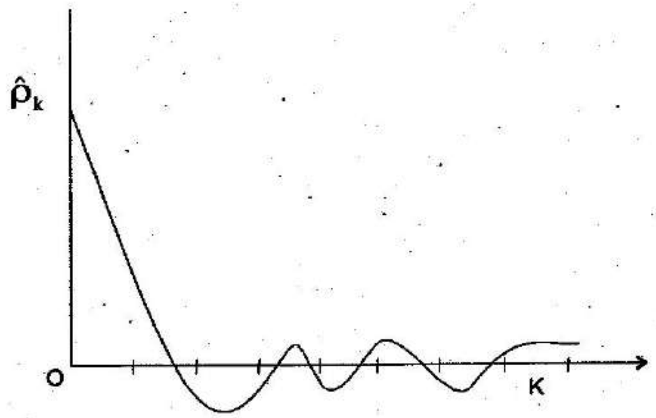
As both variance and co-variance are measured in terms of same units of measurement, ρ_k is a unitless, or pure, number. It lies between -1 and +1, as any correlations coefficients does. If we plot ρ_k against, k, the graph we obtain is know population correlogram.

Fig of Corrolegram :-

i,e we plot $\hat{\rho}_k$ against k.



If $-1<\beta<0$ then,



Since in practice we only have a sample of a stochastic process, we can only compute the sample autocorrelation function (SAFC) $\hat{\rho}_k$. To compute this we must first compute the sample covariance at lag k, $\hat{\gamma}_k$ and the sample variance $\hat{\gamma}_0$ which is defined as-

$$\hat{\gamma}_k = \frac{\Sigma\left(Y_t - \bar{Y}\right)\left(Y_{t+k} - \bar{Y}\right)}{n} \qquad ......25.2$$

$$\hat{\gamma}_0 = \frac{\Sigma\left(Y_t - \bar{Y}\right)^2}{n} \qquad ......25.3$$

Where 'n' sample size and $\bar{Y}$ is the sample mean. Therefore, the autocorrelation function at lag k is

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \qquad ......25.4$$

Which is simply the ratio of sample covariance, (at lag k) to sample variance. A plot of $\hat{\rho}_k$ again k is known as the sample correlogram.

Statistical significance of Autocorrelation coefficients :

The statistical significance of any $\hat{\rho}_k$ can be judged by its standard error. Bartlett has shown that if a time series is purely random, i,e it exhibits white noise, the sample autocorrelation coefficient $\hat{\rho}_k$ are approximathy-

$$\hat{P}_k \approx N(0, 1/n) \qquad \text{...... 26.1}$$

That is, in large sample the sample autocorrelation coefficients are normally distributed with zero mean and variances equal to one over the sample size.

Now let us turn to the estimation. This is simple enough, All we have to do is to take the first differences of Y_t and regress them one Y_{t-1} and see if the estimated slope coefficient in this regression (δ) is zero or not. If it is zero, we conclude that Y_t is nonstationary. But if it is negative, we conclude that Y_t is stationary. Under the null hypothesis that $\delta = 0$ (i,e $\rho = 1$), the 't' value of the estimated coefficient of Y_{t-1} does not follow the 't' distribution even in large samples, that is it does not haveasymptotic normal distribution.

Alternatively Dickey and Fuller have shown that under the null hypothesis that $\delta = 0$, the estimated 't' value of the coefficient of Y_{t-1} follows the (τ)(tau) statistics. These authors have computed the critcal values of the tau statistic on the basis of Monte-Carlo simulations. In literature the tau statistic known as the Dickey-Fuller (DF) test, in honor of its discoveries. Interesting, if the hypothesis that $\delta = 0$ is rejected (i,e the time series is stationary), we can use the usual (student) t test.

To allow for various possibilities, the DF test is estimated in three different forms, i,e under the three null hypothesis-

Y_t is a random walk $\Delta Y_t = \delta Y_{t-1} + u_t$ 26.2

Y_t is a random walk with drift $\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$ 26.3

Y_t is a random walk with $\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$ 26.4
drift around a stochastic trend :

Where t is the time or trend variable. In each case, the null hypothesis is that $\delta = 0$, i,e there is a unit root- the time series is stationary. If the null hypothesis is rejected, it means Y_t is a stationary time arises with zero mean in case of (26.2), that Y_t is stationary with a non zero mean $[\beta_1/1 - P]$ in case of 26.3 and Yt is statinary around a determinstic trend in 26.4.

It is extremly important to note that the critical values of the tau test to test the hypothesis that $\delta = 0$, are different for each of the preceding three specifications of the DF test.

4.5.2 The Augmented Dickey-Fuller Test

In the model (26.2) (26.3) and (26.4) it was assumed that the error term u_t was uncorrelated. But in case the u_t are correlated, Dickey-Fuller have developed a test, known as augmented Dickey Fuller (ADF) test. The test is conducted by "augmenting" the preceeding three equations by adding the lagged values of dependent variable ΔY_t

The ADF test here consists of estimating the following regression :-

$$\Delta Y_t = \beta_1 + \beta_{2t} + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \qquad ....26.5$$

Where ε_t is a pure white noise error term and where

$\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2}), \Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$ etc.

The number of lagged difference terms to include is often determined empirically. The idea is to include enough terms so that the error term in 26.5 is serially uncorrelated. In ADF we still test whether $\delta = 0$ and the ADF test follows the same asymptotic distribution as the DF statistic, so the critical values can be used.

Testing the significane of more than one coefficient : The F Test : Suppose we estimate a model and test the hypothesis $\beta_1 = \beta_2 = 0$, i,e the model is RWM without drift and trend. To test this join hypothesis we used F test as discussed earlier.

4.6 Summing Up

In this unit we got a basic idea of Time Series Econometrics. Concepts of Stationary Stochastic Process and Non-Stationary Stochastic process are discussed. In case of stationary process we will have time invariant mean, variance and covariance but in case of non-stationary series we will have a time varying mean or a time varying variance or both. After that we have discussed different types of non-stationary series, Autocorrelation function and a brief concept of Augmented Dickey-Fuller Test.

4.7 Self Assessment Questions

1. What is the difference between Stationary Time Series and Non-Stationary Time Series?
2. Discuss the two types of classic Random Walk Model of non-stationary time series.
3. How can we find out whether a given time series is stationary or not?

4.8 References/Suggested Readings

1. Johnston, J., "Econometric Methods", McGraw Hill.
2. Gujarathi. D., "Basic Econometrics", McGraw Hill.
3. Pindyck and Rubinfeld, "Econometric Models and Econometric Forecasts", McGraw Hill.
4. Greene, William, "Econometric Analysis", Macmillan.
5. Johnston and Dinardo, "Econometric Methods", McGraw Hill.

* * *

Unti-5
INTRODUCTION TO SIMULTANEOUS EQUATION MODEL

Contents:

5.0 Introduction
5.1 Objectives
5.2 Meaning and Structure of Simultaneous Equation Models
5.3 Simultaneity Bias
5.4 Identification Problem: Introduction
  5.4.1 Under identifiaction
  5.4.2 Just or Exact Identification
  5.4.3 Over identification
5.5 Rules of Identification
  5.5.1 Order Condition of Identifiability
  5.5.2 Rank Condition of Identifiability
5.6 Estimation of Just Identified Equation
5.7 Estimation of an Over Identified Equation
5.8 Summing Up
5.9 Self Assessment Questions
5.10 References/Suggested Readings

5.0 Introduction

There are situations where there is a two-wag-flow of influence among economic variable's i,e one economic variable (s), affects another economin variable (s) and in turn, is affected by it (them). And this leads us to consider- simulteneous -equation models, model in which there is more than one regression equation, one for each independent variable.

5.1 Objectives

This units mainly aims to illustrate the concep of if simultaneous- equation model and its economic applications. More specifically it seeks to deal with-

- Various concepts related to simultaneous equation model;
- Finding the estimatators of the simultaneous equation model; and
- Problem related to identification of the simultaneous equation model.

5.2 Meaning and Structure of Simultaneous Equation Models

In many situations, the cause and effect relationship are 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 are two way or on simultaneous relationship between Y's and X's, which makes the distinction between dependent on independent (explaratory) variable.

Ex – The demand and supply function.

or $$Y_{1i} = \beta_{10} + \beta_{12} Y_{2i} + \gamma_{11} x_{1i} + u_{1i} \qquad .... (14.1)$$

$$Y_{2i} = \beta_{20} + \beta_{21} Y_{1i} + \gamma_{21} x_{1i} + u_{2i} \qquad .... (15.2)$$

Here Y_1 and Y_2 are mutually dependent or endogenous variables, X_i is and exogeneous variable and u_1 & u_2 are the stochastic distrubance term.

5.3 Simultaneity Bias

As we stated previously, the method of least square may not be applied to the simultaneous equation model, if one or more explanatory variables are correlated with the disturbance term in that equation. Because the estimators thus obtained are inconsistent.

Ex :- Consider simple Keynsian model of income determination.

Consumption function : $C_t = \beta_0 + \beta_1 Y_t + u_t \quad 0 < \beta_1 < 1$ ....(14.1)

Income identity : $Y_t = C_t + I_t (= S_t)$ ....(14.2)

C=Consumption, Y=income, I=Investment, S=Savings, t=time, u=stochastic distrubance term. β_0 and β_1 are parameters.

Assuming that $E(u_t) = 0, E(u_t^2) = \sigma^2, \quad E(u_t, u_{t+i}) = 0$

$i \neq 0$ and cov $(I_t u_t)$=0 i,e. assumption of CLRM.

To prove that Y_t and u_t are correlated, we proceed as follows. substitute-(14.1) in (14.2) to obtain–

$$Y_t = \beta_0 + \beta_1 Y_t + u_t + I_t$$

Now $$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} U_t \qquad .... (15.1)$$

Now $$E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t \qquad .... 15.2$$

Where use is made of the fact that $E(u_t)$= 0 and I_t =exogeneous or predetermined, has as its expected value I_t.

Now, substracting 15.2 from 15.1 we get-

$$Y_t - E(Y_t) = \frac{u_t}{1-\beta_1} \qquad .... 15.3$$

Moreover- $u_t - E(u_t) = u_t$ 15.4

Where $\text{cov}(Y_t, u_t) = E[Y_t - E(Y_t)][u_t - E(u_t)]$

$$= \frac{E(u_t^2)}{1-\beta_1}$$

$$= \frac{\sigma^2}{1-\beta_1}$$

Since σ^2 is positive by assumption, the covariance between Y_t and u_t given in (15.5) is bound to be different from zero. As a result Y_t and u_t in 14.1 are expected to correlated which violates assumption of CLRM and where estimators are inconsistent.

To show that the OLS estimators $\hat{\beta}_1$ is an inconsistent estimator of β_1 –

$$\hat{\beta}_1 = \frac{\Sigma(Ct-\bar{c})(Y_t-\overline{Y})}{\Sigma(Y_t-\overline{Y})^2} \qquad \text{........ 15.6}$$

$$= \frac{\Sigma c_t y_t}{\Sigma y^2}$$

$$= \frac{\Sigma c_t y_t}{\Sigma y_t^2}$$

Substitation for c_t from 14.1 we obtain

$$\hat{\beta}_1 = \frac{\Sigma(\beta_0+\beta_1 Y_1+u_t)y_t}{\Sigma y_t^2} \qquad \text{...... 15.7}$$

$$= \beta_1 + \frac{\Sigma(y_t u_t)}{\Sigma y_t}$$

where $\Sigma y_t = 0$ and $(\Sigma Y_t y_t / \Sigma y_t^2) = 1)$

If we take the expectation of (15.7) of both sides, we obtain-

$$E\left(\hat{\beta}\right) = \beta_1 + E\left[\frac{\Sigma y_t u_t}{\Sigma y_t^2}\right] \qquad \text{......15.8}$$

Unfortunatly we cannot evaluate $E(\Sigma y_t u_t \mid \Sigma y_t^2)$, since the expectations operator is a linear operator.

But if the sample size increases indefinitely then we can resortto the concept of consistent estimators and find out and what happen to $\hat{\beta}_1$ as the sample size increase indefinity.

Now an estimator is said to be consistent if its probability limit or plim for short, is equal to its true value (population value).

Applying the rules of probability—

$$\text{Plim}\left(\hat{\beta}_1\right) = \text{Plim}(\beta_1) + \text{Plim}\frac{(\Sigma y_t u_t)}{(\Sigma y_t^2)}$$

$$= \text{Plim}(\beta_1) + \text{Plim}\left(\frac{\Sigma y_t u_t / n}{\Sigma y_t^2 / n}\right)$$

$$= \beta_1 + \frac{\text{Plim}(Ey_t u_t / n)}{\text{Plim}(Ey_t^2 / n)} \qquad \text{............ (15.9)}$$

In 15-9 where in the 2nd step, we have divided $\Sigma y_t u_t$ and Σy_t^2 by the total number of observations in the sample size 'n' so that the quantities in the parentheses are now the sample covariances between 'Y' and 'U' and the sample variance of Y, respectively.

In other words, (15.9) state that the probability limit of $\hat{\beta}_1$ is equal to the true β_1 plus the ratio of the plim of the sample co-variance between Y and u to the plim of the sample variance Y. If the sample size 'n' increase indefinitely, one would expect the sample covariance between Y and u to approximate the true population covariance. $E[Y_t - E(Y_t)]\,[U_t - E(U_t)]$ which is equal to $[\sigma^2/1-\beta_1)]$. Similar is the case for sample variance.

$$\text{So, } \operatorname{Plim}\left(\hat{\beta}_1\right) = \beta_1 + \frac{\sigma^2/(1-\beta_1)}{\sigma^2 Y}$$

$$= \beta_1 + \frac{1}{1-\beta_1}\left(\frac{\sigma^2}{\sigma_y^2}\right) \qquad \text{...... (15.10)}$$

Given that $0<\beta_1<1$ and σ^2 and σ_t^2 bothpositive, plim $(\hat{\beta}_1)$ always greater than β_1 i.e. $\hat{\beta}_1$ is a biased estimator.

5.4 Identification Problem: Introduction

To introduce our discussion, the following notation and definition. The general M equation model in M enogenous, on jointly dependent variable may be written as-

$$\left.\begin{aligned}
Y_{1t} &= \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \dots + \beta_{1M}Y_{Mt} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \dots + \gamma_{1k}X_{kt} + u_{1t} \\
Y_{2t} &= \beta_{21}Y_{1t} + \dots + \beta_{23}Y_{3t} + \dots + \beta_{2M}Y_{Mt} + \gamma_{21}X_{1t} + \gamma_{12}X_{2t} + \dots + \gamma_{1k}X_{kt} + u_{2t} \\
Y_{3t} &= \beta_{31}Y_{1t} + \beta_{32}Y_{2t} + \dots + \beta_{3M}Y_{Mt} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \dots + \gamma_{3k}X_{kt} + u_{3t} \\
Y_{MT} &= \beta_{M1}Y_{1t} + \beta_{M2}Y_{2t} + \dots + \beta_{M,M-1}Y_{M-1,t} + \gamma_{M1}X_{1t} + \gamma_{M2}X_{2t} + \dots + \gamma_{Mk}X_{kt} + u_{Mt}
\end{aligned}\right\} .16.11$$

where Y_1, Y_2, Y_M=M endogeneous or jointly dependent variable.
$X_1, X_2 \dots X_k$=K predetermind variable (one of these X variables may take a value of unity to allow for the intercept term in each equation).
$u_1, u_2 \dots u_M$=M stochastic distrubances.
$t=1,2 \dots T$=Total numbers of observations.

β 's= Coefficicats of the endogoneous variables

γ 's= coeffictents of the predetermined variables.

Two types of variable :-

(1) Endogeneous, that is, those (whose value are) determined within the model. They are regarded as stochastic.
(2) Predetermined, that is those (whose values are) determined outside the model. This variables are regarded as non-stochastics.

Again pre-determined variables are divided into two catagories-exogeneous, (current as well as lagged) and lagged endogeneous.

X_{1t}- current exogeneous

$X_{1(t-1)}$- lagged exogeneous variables.

$Y_{1(t-1)}$- lagged endogeneous variables.

The equation like (16.11) known as structural, or behavioural, equation because they may portray the structure (of an economic model) of an economy or the behaviour of an economic agents (consumer/producers). β's and γ's structural paramenters or coefficient.

A reduce form equation is one that expresses an endogeneous variable solely in terms of the predetermined variables and the stochastic distrubances.

Example :-

Consumption function :-

$$C_t = \beta_0 + \beta_1 Y_t + U_t \quad 0 \angle \beta_1 \angle 1 \qquad ......19.2$$

Income Identiy-

$$Y_t = C_t + I_t \qquad ....19.3$$

Now 19.2 is substitude to 19.3 and we obtain-

$$Y_t = \pi_0 + \pi_1 I_t + w_t \qquad ....19.4$$

$$\left.\begin{aligned} \text{Where } \pi_0 &= \frac{\beta_0}{1-\beta_1} \\ \pi_1 &= \frac{1}{1-\beta_1} \\ w_t &= \frac{U_t}{1-\beta_1} \end{aligned}\right\} \qquad ....19.5$$

19.4 is the reduced form equation. π_0 and π_1 reduced form coefficients.

Note the Interesting Features :

since only the pre-determined variables and stochastic distrubances appear on the right sides of these equations, and since the predetermined variables are assumed to be the uncorrelated with the distrubances terms the OLS method can be applied to estimate the coefficients of the reduced- form equations. From the estimated reduced form coefficients one may estimate the structural coefficients (the β's). This procedure is known as indirect least squares (ILS), and the estimated structural coefficients are called ILS estimates.

By identification problems we mean whether numerical estimates of the parameters of a structural equation can be obtained from the estimated reduced form coefficients. If it can be done, then we say that equation is identified. If this cannot be done, then we say that the equation under consideration is unidentified, or under identified.

5.4.1 Under identifiaction : Demand and Supply Model

Demand function $Q_t^d = \alpha_o + \alpha_1 P_t + u_{1t}$ $\alpha_1 < o$ 19.6

Supply function $Q_t^s = \beta_o + \beta_1 P_t + u_{2t}$ $\beta_1 > 0$ 19.7

Equlibrium quation : $Q_t^d = Q_t^s$

By equlibrium we obtain $\alpha_o + \alpha_1 P_t + u_{1t} = \beta_o + \beta_1 P_t + u_{2t}$ ...19.8

Showing (19.8) we obtain the equlibrium price-

$$P_t = \pi_o + v_t \quad .....19.9$$

$$\text{Where } \pi_o = \frac{\beta_o - \alpha_o}{\alpha_1 - \beta_1} \quad .....20.1$$

$$v = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \quad .....20.2$$

Now substituting Pt from (19.9) into (19.6) we obtain the following equlibrium quantity

$$Q_t = \pi_1 + w_t \quad ...... 20.3$$

$$\text{Where, } \pi_1 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} \quad ...... 20.4$$

$$W_t = \frac{\alpha_1 u_2 t - \beta_1 u_1 t}{\alpha_1 - \beta_1} \quad ...... 20.5$$

The error terms v_t and w_t are linear. Combinations of the original error terms u_1 and u_2. From the reduced for equation there are four structural coefficient $\alpha_0, \alpha_1, \beta_0$ and β_1 but there is no unique way of estimating them because there are only two reduced form equations. For exactly identified equations if there are K unknowns then we must have K (independent) equations.

5.4.2 Just or Exact Identification

Demand function : $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}$ $\alpha_1 < 0, \alpha_2 > 0$ ...20.6

Supply function : $Q_t = \beta_0 + \beta_1 P_t + u_{2t}$ $\beta_1 > 0$ 20.7

Where I= Income of the consumer, an exogeneous variable, and all other variables are as defined previously with respect to the former model. Here we include one additional variable in the demand function namely, income. Using the market clearing mechanism, quantity demanded = Quantity supplied, we have

$$\alpha_0+\alpha_1 P_t+\alpha_2 I_t+U_{1t}=\beta_0+\beta_1 P_t+U_{2t} \qquad ......20.8$$

Solving the equation 20.8 provides the following equlibrium value of P_t.

$$P_t=\pi_0+\pi_1 I_t+v_t \qquad ...... 20.9$$

Where reduced form coefficient are

$$\left.\begin{aligned} \pi_0 &= \frac{\beta_0-\alpha_0}{\alpha_1-\beta_1} \\ \pi_1 &= \frac{\alpha_2}{\alpha_1-\beta_1} \\ v_t &= \frac{u_{2t}-u_{1t}}{\alpha_1-\beta_1} \end{aligned}\right\} \qquad ....20.10$$

Now substituting the equlibrium value of P_t into the preceding demand or supply function, we obtain the following equlibrium quantity–

$$Q_t=\pi_2+\pi_3 I_t+w_t \qquad ......21.1$$

$$\left.\begin{aligned} \pi_2 &= \frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1} \\ \pi_3 &= \frac{\alpha_2\beta_1}{\alpha_1-\beta_1} \\ W_t &= \frac{\alpha_1 U_{2t}-\beta_1 U_{1t}}{\alpha_1-\beta_1} \end{aligned}\right\} \qquad ....21.2$$

Since 20.9 and 21.1 are both reduced form equations, we can use the OLS to estimate the parameters. Now in the demand and supply function there are five structural co-efficients - $\alpha_0, \alpha_1, \alpha_2, \beta_1$ and β_2. But there are four equation to estimate them, namely four reduced form co-effiecient π_0, π_1, π_2 and π_3 given in 20.10 and 21.2. Hence unique solution of all structural coeffecients is not possible. But it is possible that supply function can be estimated (identified)

$$\beta_0=\pi_2-\beta_1\pi_0$$

$$\beta_1=\frac{\pi_3}{\pi_1} \qquad .....21.3$$

But there is no unique way to identify the demand function, therefore it remains underidentified. But notice an interesting fact : It is the presence of an additional variable in the demand function that enable us to identify the and supply function.

5.4.3 Over identification

For certain goods and services income as well as wealth of the consumer's is an impotant determinant of demand.

Demand Function :

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \qquad .....21.4$$

Supply function :

$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \qquad .....21.5$$

Where in addition to the variable already defined, R represents wealth, for most goods and services, wealth, like income, is expected to have a positive effect on consumption.

Equating demand to supply, we obtain the following equlibrium price and quantity -

$$P_t = \pi_0 + \pi_1 I_t + \pi_2 R_2 + \pi_3 P_{t-1} + v_t \qquad .....21.6$$

$$Q_t = \pi_4 + \pi_5 I_t + \pi_6 R_t + \pi_7 P_{t-1} - 1 + w_t \qquad .....21.7$$

Where,

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \qquad \pi_1 = \frac{\alpha_2}{\alpha_1 - \beta_1}$$

$$\pi_2 = \frac{\alpha_3}{\alpha_1 - \beta_1} \qquad \pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_4 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} \qquad \pi_5 = \frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} \qquad ....21.8$$

$$\pi_6 = \frac{\alpha_3 \beta_1}{\alpha_1 - \beta_1} \qquad \pi_7 = \frac{\alpha_1 \beta_1}{\alpha_1 - \beta_1}$$

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \qquad v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

The preceding demand and supply model contains seven structural coefficients, but there are eight equations to estimate them- the eight reduced form coefficients given in (21.8). that is, the number of equations is greater than the number of unknowns. Therefore, unique estimation of all the parameters of our model is not possible. From the reduced for coefficients we can obtain

$$\beta_1 = \frac{\pi_6}{\pi_2}$$

or $\beta_1 = \frac{\pi_5}{\pi_1}$, that is, threre are two estimates of the price coefficient in the supply function, and there is no guarantee that these two values or solutions will be identical.

5.5 Rules of Identification

There are two condition :

i) The order condition of identifiability.

ii) The Rank conditionof indetifiability.

M= Number of endogeous variables in the model

m= Number of endogenous variables in a given equation.

K= Number of predetermined variables in the model including the intercept.

k = Number of predetermined variables in a given equation.

5.5.1 Order Condition of Identifiability

A necessary (but not sufficient) condition of identification, known as the order condition, may be stated in two different ways-

1st : In a model of "M" simulteneous equation in order for an equation to be identified, it must exeludes at least M-1 variables (endogeneous as well as predetermined) appearing in the model. If it excludes exactly (M-1) variables, the equation is just identified. If it excludes more than M-1 variables, if is overidentified.

2nd rule : In a model of M simulteneous equation, in order for an equation to be identified, the number of predetermined variables excluded from the equation must not be less than the number of endogeneous variables includes in that equation less 1, that is $K-k>m-1$, if $K-k=m-1$, the equation is just identified, but if $k-k>m-1$ it is overidentified.

5.5.2 Rank Condition of Identifiability

The order condition discussed previously is a necessary but not sufficient condition for identification, that is, even it is it is satisfied, if may happen that the equation is not identified. We need one sufficient condition for identification. This is provided by the rank condition of identification. Which may be stated as-

In a model containing M equations in M endogeneous variables, an equation is identified if and only if at least one nonzero determinant of order (M-1) (M-1) can be constructed from the coefficients of the variables (both endogeneous and predetermined) excluded from that particular equation but included in the other equations of the model.

An illustration of the rank condition of identification- where the Y variables are endogneous and the X variables are predetermined.

$$Y_{1t}-\beta_{10} \quad -\beta_{12}Y_{2t}-\beta_{13}Y_{3t}-\gamma_{11}X_{1t}=u_t \quad ....21.1$$

$$Y_{2t}-\beta_{20} \quad -\beta_{23}Y_{3t}-\gamma_{21}X_{1t}-\gamma_{22}X_{2t}=U_{2t} \quad ....21.2$$

$$Y_{3t} - \beta_{30} - \beta_{31} Y_{1t} - \gamma_{31} X_{1t} - \gamma_{32} X_{2t} = u_{3t} \quad ....21.3$$

$$Y_{4t} - \beta_{40} - \beta_{41} Y_{1t} - \beta_{42} Y_{2t} - \gamma_{43} X_{3t} = u_{4t} \quad ....21.4$$

To facilitate identification, let us write the preceding system in following Table which is self explanatory.

Table 1 Cofficient of the Variables

Equation No	1	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
20.1	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
20.2	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
20.3	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
20.4	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

Table 2

Equation No	No. of Predetermined variable excluded K-k	No. endogeneous variable included less one (m-1)	Identified
20.1	2	2	Exactly
20.2	1	1	Exactly
20.3	1	1	Exactly
20.4	2	2	Exactly

Let us first apply the order condition of identification, as shown in Table 2. By order condition all equations are identified. Now check with Rank condition. First equation, which excludes 3 variables, Y_4, X_2 and X_3. For this equation to be identified we must have atleast one nonzero determinant of order 3x3 from the coefficient of variable excluded from this but included in other equations. In present case there is only one matrix, call it A

$$\det A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix} \quad .... 21.5$$

It can be seen that the determined of this matrix is zero :

$$\det A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix} = 0 \quad .... 21.6$$

Since the determined is zero, the rank of the matrix 21.5, denoted by ρ (A) is less than 3. Therefore equation 20.1 does not satisfy the rank condition and hence is not identified.

Note :- The rank condition tells us whether the equation under considerations is identified or not, whereas the order condition tell us if it is exactly identified or overidentified.

To apply the rank condition one may proceeds as follows :

1. Write down the system in a tabular form, as show in above two table.
2. Strike out the coefficient of the row in which the equation under consideration appears.
3. Strike out the column's corresponding to those coefficients in 2 which are none zero.
4. The entires left in the table will then gives only the coeffiecients of the variables included in the system but not in equation under consideration. From these entries form all possible matrices like A,of order M-1 and obtain the corresponding determinants. If atleast one nonzero determinant can be found the equation in question is (just or over) identified.

5.6 Estimation of Just Identified Equation

The Methods of Identical least squares (ILS) :

For a just or exactly identified equation (structural), the method of obtaining the estimators of the structural coefficient form OLS estimates of the reduced form coefficients known as the method of indirect least squares (ILS). The estimation follows the following step-

Step 1 : We first obtain reduced form equations, where the endogeneous variables is a function of predetermined variables (endogenous or lagged endogeneous) and the stochastic error term(S).

Step 2 : We apply OLS to the reduced form equations individually. This is possible due to step 1.

Step 3 : We obtain estimates o fthe original structural coefficient from the estimated reduced form coefficients obtained in step 2. As we know, if an equation is exactly identified, there is a one-to-one correspondence between the structural and reduced form coefficients.

Example :

Demand function : $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_{1t}$ 20.1

Supply function : $Q_t = \beta_0 + \beta_1 P_t + u_{2t}$ 20.2

Q= Quantity, P=Price, X=Income

X= Exogeneous

Reduced from equation :-

$P_t = \pi_0 + \pi_1 X_1 + w_t$ 20.3

$Q_t = \pi_2 + \pi_3 X_t + v_t$ 20.4

Where the π's are the reduced form coefficients and are (nonlinear) combinations of the structural coefficients. w and v are linear combination of structural distrubance term u_t and u_2.

Reduced form equations contain only one endogeneous variable and which is a functions of 'X' exogeneous variables so we can use OLS

$$\hat{\pi}_1 = \frac{\Sigma P_t x_t}{\Sigma x_t^2} \quad .....20.5$$

$$\hat{\pi}_0 = \bar{P} - \hat{\pi}_1 \bar{X} \quad .....20.6$$

$$\hat{\pi}_3 = \frac{\Sigma q_t x_t}{\Sigma x_t^2} \quad .....20.7$$

$$\hat{\pi}_2 = \bar{Q} - \hat{\pi}_3 \bar{x} \quad .....20.8$$

Where $\bar{Q}$ and $\bar{P}$ are the sample mean value of Q and P. and $\hat{\pi}_i$ are the consisten estimators. Now as we determined the supply function it is exactly identified. Therefore its parameter can be estimated from the reduced form coefficients as follows-

$$\beta_0 = \pi_2 - \beta_1 \pi_o \text{ and } \beta_1 = \frac{\pi_3}{\pi_1}$$

Hence, the estimates of these parameters can be obtained from the estimates of the reduced form coefficients as -

$$\hat{\beta}_0 = \hat{\pi}_2 - \hat{\beta}_1 \hat{\pi}_0 \quad ...... 20.9$$

$$\hat{\beta}_1 = \frac{\hat{\pi}_3}{\hat{\pi}_1} \quad ......20.10$$

Which are the ILS estimators. Note that the parameters of the demand functions cannot be estimated.

5.7 Estimation of an Over Identified Equation

The Method of two stage-least Squares (2SLS) :

Income function = $Y_{1t} = \beta_{10} + ... + \beta_{11} Y_{2t} + \gamma_{11} X_{1t} + \gamma_{12} X_{2t} + u_{1t}$ ..21.2

Money supply function : $Y_{2t} = \beta_{20} + \beta_{21} Y_{1t} + ... + u_{2t}$ 21.3

The variable X_1 and X_2 are exogeneous. In this type of model if we apply OLS then we obtain inconsistent estimators, as the Y_{1t} and the distrubance term u_{2t} are correlated. In such case we used 2SLS developed by Henri Theil and Robert Basmann. The methods involve following steps:-

Step 1 :

To get rid of the likely correlation between Y_1 and u_2, regress first Y_1 on all the predetermined variables in the whole system, not just that equations. In present case, this means regressing Y_1 on X_1 and X_2 as follows:

$$Y_{1t} = \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \hat{\pi}_2 X_{2t} + \hat{u}_t \quad .....21.4$$

Where $\hat{u}_t$ are the usual OLS residuals. From equation 21.4 we obtain-

$$\hat{Y}_{1t} = \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \hat{\pi} X_{2t} \quad .....21.5$$

$\hat{Y}_{1t}$ is an estimate of the mean of Y conditional upon the fixed X's.
So we can writes 21.4 as

$$Y_{1t}=\hat{Y}_{1t}+\hat{u}_t \qquad ......21.6$$

Stage 2 : The over identified money supply equation can be written as-

$$\left.\begin{aligned} Y_{2t} &= \beta_{20}+\beta_{21}(\hat{Y}_{1t}+\hat{u}_t)+u_{2t} \\ &= \beta_{20}+\beta_{21}\hat{Y}_{1t}+(u_{2t}+\beta_{21}\hat{u}_t) \\ &= \beta_{20}+\beta_{21}\hat{Y}_{1t}+\hat{u}_t^* \end{aligned}\right\} \qquad ......21.7$$

Where $u_t^* = u_{2t} + \beta_{21}\hat{u}_t$

Comparing 21. 3 and 21.7, we see that they are very similar in appearance, the only difference being that Y_1 is in replaced by $\hat{Y}_1$. The advantage of is that $\hat{Y}_1$ it uncorrelated with U_t^* in 21.7 asymptotically, that is large sample as the sample size increase indefinitely. But in original model Y_1 and U_2 is correlated.

As this two-stage procedure indicates, the basic idea behind 2SLS is to 'purify' the stochastic explanatory variable Y_1 of the influence of the distrubance U_2. The goal is accomplished by performing the reduce form regression of Y_1 on all the predetermined variables in the system (stage1), and obtaining the estimates $\hat{Y}_{1t}$, and replacing Y_{1t} in the original equation by the estimated $\hat{Y}_{1t}$ and applying OLS to the equations thus transformed (stage 2). The estimator thus obtained are consistent, that is they coverge to their true value as the sample size increase indefinitely.

Features of 2SLS :

1. It can be applied to an individual equation on the system without directly taking into account any other equation (s) in the system.
2. Unlike ILS, which provides multiple estimates of parameters in the overidentified equations, 2SLS provides only one estimate per parameter.
3. It is easy to apply because all one needs to know is the total number of exogeneous or predetermined variables in the system without knowing any other variable in the system.
4. Although specially designed to handle over identified equations, the method can also be applied to exactly identified equations. Where ILS and 2SLS estimates will be identical.
5. If yhe R^2 values in the reduced form regeression are very high, say in excess of 0.8, The classical OLS estimates and 2SLS estimates will be very close. This result means, that the estimated values of the

endogeneous variable are very close to their actual values, and hence the latter are less likely to be correlated with the stochastic distrubances in the original structural equations. If not i,e if R^2 low then the $\hat{Y}$'s will be very poor proxies for the original Y's.

6. Notice, in reporting the ILS regression, we did not state the standard errors of the estimated coefficients. But we can do this for the 2SLS estimates because the structural coefficients are directly estimated from the second stage of (OLS) regression.

5.8 Summing Up

In this unit we have learned about simultaneous equation model and its various components like structural and reduced form coefficients, simultaneity Bias etc. It has also forwarded an informal introduction to identification, over identification or under identification etc. Estimation methods in case of just identified equation namely ILS and in case of over identified equation viz. 2SLS are also discussed here.

5.9 Self Assessment Questions

1. Discuss the concept of simultaneity Bias with the help of suitable example.
2. Discuss the order and Rank condition of identification.
3. Discuss the methods of ILS and 2SLS. Under what condition they give identical result?

5.10 References/Suggested Readings

1. Johnston, J., "Econometric Methods", McGraw Hill.
2. Gujarathi. D., "Basic Econometrics", McGraw Hill.
3. Pindyck and Rubinfeld, "Econometric Models and Econometric Forecasts", McGraw Hill.
4. Greene, William, "Econometric Analysis", Macmillan.
5. Johnston and Dinardo, "Econometric Methods", McGraw Hill.

* * *