BLOCK 2

ADVANCED TOPIC IN REGRESSION ANALYSIS

INTRODUCTION TO BLOCK 2

Block 2 is on Advanced Topics in Regression Analysis. It has four units (units 4 to 7). **Unit 4** is on Auto-Regressive and Distributed Lag Models. Two basic features of lags [viz. 'factors contributing to lags' and 'market equilibrium and lags'] are explained first. Methods of obtaining solutions to such distributed lag models are discussed next. Under this, three methods viz. (i) the OLS method, (ii) adhoc estimation method and (iii) Koyck's approach are explained. The unit also covers some 'alternative approaches to solving distributed lag models'. Under this, four approaches are discussed. These are: (i) partial adjustment model, (ii) adaptive expectations model, (iii) blended approach and (iv) Almon's approach.

Unit 5 is on Binary Dependent Variable Models. This unit covers three models viz.: (i) the Linear Probability Model (LPM), (ii) the Logit Model and (iii) the Probit Model. Under LPM, the unit explains issues like (i) non-fulfillment of probability limits, (ii) R^2 as measure of goodness of fit, (iii) limitations of LPM and (iv) corrective measures. The Logit and Probit models cover issues of estimation, interpretation and application.

Unit 6 is on 'Simultaneous Equation Models I'. The unit explains (i) consequences of simultaneity with reference to simple regression models and the Keynesian Model of Income Determination, (ii) the identification problem in situations of under identification and over identification and (iii) the conditions which help us to test for the identification status.

Unit 7 continues the discussion on Simultaneous Equation Models with an exposure to three methods of estimation viz. ILS, IV and 2 SLS. Under each, aspects like characteristics, assumptions, properties and estimation procedure are explained.

UNIT 4 AUTO-REGRESSIVE AND DISTRIBUTED LAG MODELS*

Structure

4.0 OBJECTIVES

After reading this unit, you will be able to:

- define the term 'lags' with examples;
- describe the significance of 'lags' in economics with illustrations;
- outline the factors that contribute to the 'lag effect';
- discuss the effect of lags on 'market equilibrium' with suitable examples;
- state the expressions for (i) distributed lag models, (ii) auto-regressive model and (iii) auto-regressive distributed lag models;
- explain the approach to applying the OLS procedure for the case of distributed lag models;
- demonstrate the 'adhoc approach' of estimating the distributed lag models stating its limitations;

* Prof. Sushil Haldar, Jadavpur University.

- show that the Koyck's approach to estimating the distributed lag models helps in overcoming an 'infinite series situation';
- bring out the limitations of Koyck's approach;
- compare the approach in the Nerlove's 'Partial Adjustment Model' with that of Koyck's transformation specifying the common features between the two;
- highlight the features of the 'adaptive expectations model';
- derive the expression under the 'blended approach' for estimating the parameters of a distributed lag model'; and
- write a note on 'Almon's approach' to estimating a distributed lag model.

4.1 INTRODUCTION

There is always some time lag between a cause and its effect. This is irrespective of the discipline of events i.e. the events may be from social sciences, physical sciences, medical sciences, etc. When the time lag between a cause and its effect is short, we say that the two events are contiguous. In general, in social sciences, non-contiguous causes sets in motion other processes which, in turn, results in time lag effect. For instance, farmers are interested in producing more output in the current year based on the good price received in the last year. Similarly, current wage earning depends on past period's human capital accumulation. Likewise, greater parental education leads to a variety of lifestyle benefits (like greater income, social capital, and advanced reading and verbal skills). These, in turn, influence their children's educational attainment. Note that a causal claim between parental education levels and children's education levels is not warranted unless there is a theory to suggest otherwise. Clearly, more educated parents are more likely to encourage their children to education than the uneducated parents. Higher education by those children helps them succeed better in life with a lag effect.

4.2 SIGNIFICANCE OF LAGS IN ECONOMICS

Lags are nothing but time related delays between cause and effect. To understand lags, it is essential to grasp the dynamics in the economy. It is difficult to know which actions cause what consequences. Due to this, a lag between decision and outcome is on account of unpredictability. For instance, the 'relative income hypothesis' of consumption function states that the present consumption is not only influenced by present levels of income, but also by levels of consumption in a previous period. It is difficult for a family to reduce a level of consumption once attained. Therefore, the aggregate ratio of consumption to income is assumed to depend on the relative level of present income to past income. Likewise, the 'accelerator theory of investment' uses time lag in consumption to suggest that investment intentions depend on the pace of growth in economic activity. In other words, the acceleration principle draws a connection between changing consumption

patterns and capital investment. More specifically, the acceleration principle states that if the appetite for consumer goods increase, the demand for equipment and other investments necessary to make these goods will also increase. This means, if a population's income increases, and as a result, its residents consume more, there will be a corresponding increase in investment. We can therefore write this as:

$$I_t = \beta(C_t - C_{t-1}) \qquad (4.1)$$

where I_t stands for investment at time t, $\beta > 0$, C_t and C_{t-1} are the consumption at time t and (t-1) respectively. One can simply formulate an 'optimal control capital accumulation model' with an exogenous time lag between investment and the accumulation of the capital stock. Such models help in analysing the influence of time lags in dynamic systems. Theoretical developments have shown that optimal investment paths for a finite time lag are cyclic (as opposed to monotonic paths) for capital accumulation. At the firm level, investment projects depend on past outputs, on expectations about future profits and on capital stock and interest rates. Therefore, we can write this as:

$$I_t = f(Y_t, Y_{t-1}.., \pi_{t,} K_{t-1}, r_{t\backslash}, r_{t-1}) \qquad (4.2)$$

where, Y = level of output, π = profit, K = capital stock and r = rate of interest. The decision to invest in research and development (R & D) expenditure and its payoff in terms of increased productivity involve considerable time lag. Likewise, the time lag between the invention of an idea and its development to a commercially applicable stage is large. With change in technology and its adoption, it takes time before all the old machines are replaced by the better new ones. The lag introduced in the process of such diffusion bear important economic implications making the studying of 'distributed lag models' important.

Certain economic decisions are explicitly driven by a history of related activities. For instance, energy demand by individuals is a function not only of current prices and income, but also of accumulated stocks of energy. Huge capital would have been invested in generating such stocks. The demand for durable goods also depend on both present and past levels of income. Together, they determine the amount saved for the acquisition of the durables, stocks of durables, current prices, etc. We can write this as:

$$Q_D^D = D(Y_t, Y_{t-1}, ..., S_{D,(t-1)}, P_t) \qquad (4.3)$$

where Q_D^D stands for quantity demanded for durable goods, Y is the level of income, S_D is the stock or acquisition of durables and P is the price.

4.2.1 Factors Contributing to Lags

There are several factors contributing to making the consideration of time lags important in economics. These are as follows:

i) Time lags play an important role in the 'effectiveness of an economic policy'. For instance, interest rate cuts can takes several months to show their full effect. In case of fiscal policy, if the government plans to increase spending, it takes several months for the impact of fiscal stimulus to take effect.

ii) Contractual obligations prevent firms from switching from one source of labour or raw material to another. The same logic is applied in case of fixed deposit schemes. Once a particular scheme is opted by a depositor, it cannot be redeemed because of change in interest rate. Likewise, employers may give their employees a choice among several health insurance plans. Once a choice is made, the employee cannot switch to another plan. Such cases come under 'contractual constraints'.

iii) Lag in economic policies are also a result of 'technology and imperfect information'. A firm employing more labour relative to capital cannot substitute capital for labour even if the price of capital drops. Similarly, when a new machine is introduced due to technological development in the market, its demand is initially low. This is due to uncertainty on the efficiency of the new machinery. Once the uncertainties settle down, over a time lag, the demand for new machinery picks up.

iv) Due to old habits or inertia, people do not quickly react to purchasing new goods, take new economic decisions, change business behaviour, etc. A time lag operates due to 'psychological reasons'.

v) Lags are of paramount importance in governmental 'decision making'. It is crucial for the government to know how fast, or after what time period, the economic units react to changes in policy variables. For instance, how will consumers react to the imposition of a sales tax or a credit squeeze? How will firms react to tax concessions and other incentives for investment and innovation? What will be the effect of devaluation in currency? How fast will investors react to changes in interest rates? These are some examples of 'decision delay' or the time lag effect on decisions.

vi) In modelling the response of economic variables to policy stimuli, it is expected that there will be long lags between policy changes and their impacts. The length of lag between changes in monetary policy and its impact on important economic variables such as output, employment and investment has been a subject of analysis for decades.

We can, therefore, conclude that in a dynamic world of continuous adjustment, any adjustment process takes effect only after due length of time period depending on the nature of particular phenomenon.

4.2.2 Market Equilibrium and Lags

Suppose that in the market for a single good, the supply and demand equations for period *t* are given by:

$$S_t = \alpha_1 + \beta_1 P_{t-1} \quad (4.4)$$

$$D_t = \alpha_2 + \beta_2 P_t \quad (4.5)$$

where S_t = supply at time *t*, D_t = demand at time *t*, P_{t-1} = price at $(t-1)$, P_t = price at *t*, $\alpha_1 < \alpha_2$, $\beta_1 > 0$ and $\beta_2 < 0$. It is assumed that the price is set in each period to clear the market. Note that this is a model with lagged supply in which the supply depends on the previous period's price. Further, since production or supply takes time, the adjustments on the supply side will not be instantaneous. In other words, the effect can be perceived in the market only after a time lag. Now, if the slope of the supply function (β_1) is positive, but the slope of the demand function (β_2) is negative, we need to study the behaviour of price. Such a model of market instability is known as the 'cobweb model'. This is called Cobweb because the path taken by the observed price and quantity resembles a cobweb. Since the equilibrium condition is, $D_t = S_t$, we can write:

$$\alpha_1 + \beta_1 P_{t-1} = \alpha_2 + \beta_2 P_t \quad (4.6)$$

$$\Rightarrow \beta_2 P_t - \beta_1 P_{t-1} = \alpha_1 - \alpha_2 \quad (4.7)$$

Equation (4.7) is basically a first order non-homogenous difference equation. Hence, the solution of this equation consists of two parts (a) a particular integral and (b) a complementary function. The particular integral is obtained by assuming $P_t = P_{t-1} = ... = \overline{P}$. Thus, from equation (4.6), we get:

$$\beta_2 \overline{P} - \beta_1 \overline{P} = \alpha_1 - \alpha_2$$

$$\Rightarrow \overline{P}(\beta_2 - \beta_1) = \alpha_1 - \alpha_2 \Rightarrow \overline{P} = \frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1} \quad (4.8).$$

Equation (4.7) is invariant over time. It is the inter-temporal equilibrium value of price at which demand equals supply. This is positive since both the numerator and denominator are negative (since $\alpha_1 < \alpha_2$, $\beta_2 < 0$ and $\beta_1 > 0$). Now, the complementary function can be obtained from a trial solution. We use the trial solution $P_t = Ax^t$ in the Equation (4.7) i.e. its left hand side to get:

$$\beta_2 Ax^t - \beta_1 Ax^{t-1} = 0$$
$$Ax^{t-1}(\beta_2 x - \beta_1) = 0$$
$$(\beta_2 x - \beta_1) = 0 \because Ax^{t-1} \neq 0$$

Therefore, $x = \frac{\beta_1}{\beta_2}$. Thus, since the general solution is the sum of particular integral and complementary function, this means:

$$P_t = \overline{P} + Ax^t = \overline{P} + A\left(\frac{\beta_1}{\beta_2}\right)^t \qquad (4.9_a)$$

In Equation (4.9_a), the expression A is an arbitrary constant which can be estimated from the initial conditions. In order to find the value of A, we consider the case of $t = 0$. This implies:

$$P_0 = \overline{P} + A\left(\frac{\beta_1}{\beta_2}\right)^0 \Rightarrow P_0 = \overline{P} + A \Rightarrow A = (P_0 - \overline{P}) \Rightarrow A = \left(P_0 - \frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1}\right).$$

The complete solution of the first order difference equation therefore is:

$$P_t = \overline{P} + (P_0 - \overline{P})\left(\frac{\beta_1}{\beta_2}\right)^t = \left(\frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1}\right) + \left\{P_0 - \left(\frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1}\right)\right\}\left(\frac{\beta_1}{\beta_2}\right)^t \qquad (4.9_b)$$

Given that $\beta_1 > 0$ and $\beta_2 < 0$ i.e. the supply function slopes upwards and the demand function slopes downwards, we have, $\left(\frac{\beta_1}{\beta_2}\right)$ is negative. Further, $\left(\frac{\beta_1}{\beta_2}\right)^t$ will alternate in its sign i.e. it will be negative in odd-numbered periods and positive in even-numbered periods. This means with the normal-shaped demand and supply curves, the Cobweb will always produce a two-period oscillation. The actual price (P_t) will be alternately above and below the equilibrium price ($\overline{P}$). These oscillations will either converge to or diverge from $\overline{P}$. In the demand-supply curve diagrams, we generally measure quantity along the horizontal axis and price along the vertical axis. The slopes of the supply and demand curves would therefore be the reciprocals of the slopes of the supply and demand functions (i.e. β_1 and β_2). Even if the supply and demand functions have their normal slopes (i.e. positive and negative respectively), three possible outcomes emerge viz. convergent, divergent and neither convergent nor divergent. The convergence solution is obtained if the demand curve is flatter than the supply curve. The divergent solution will emerge in the opposite case i.e. if supply curve is flatter than the demand curve. A special case is when both the demand and supply curves are equally steep. In such cases, $\left|\frac{\beta_1}{\beta_2}\right| = 1$.

Now, suppose that initially the supply of the good concerned has fallen short of the equilibrium amount because of some disturbance like drought or flood. Let the initial supply be Q_1 as shown in Fig.4.1. This amount would be demanded in the initial period if the price is P_1. At $P = P_1$, the consumers demand is not matched with the producer's supply. As a result, in period 2, the price will sharply drop from P_1 to P_2. In the Cobweb model, this period's price influences the next period's supply. Hence, the price P_1 of the initial period (period 1) determines the supply of period 2. In other words, the price P_1 induces the entrepreneurs to supply more output in period 2. This way, the

process continues indefinitely producing a Cobweb pattern. The fluctuation in prices will be higher than the equilibrium price $\overline{P}$ in one period and lower than $\overline{P}$ in the next period. It however converges to the equilibrium level at the point of intersection of demand and supply curves.

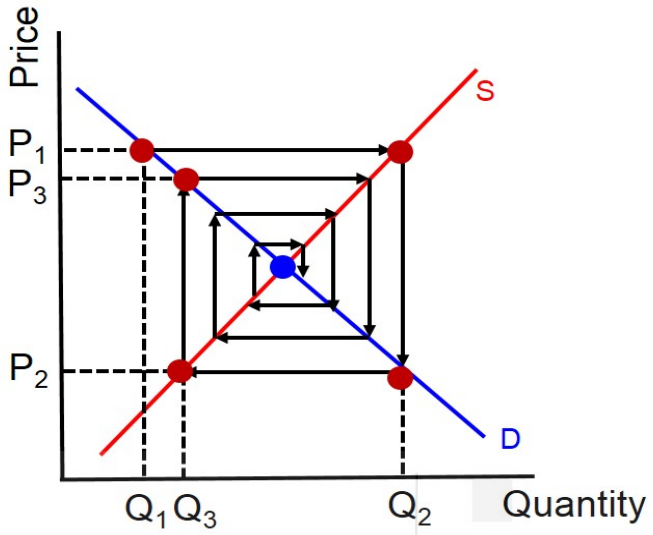

Fig. 4.1 Market Equilibrium

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) Define the term 'lags' with examples.

……

……

……

……

2) State the factors that contribute to the 'lag effect'.

……

……

……

……

4.3 SOLUTION TO DISTRIBUTED LAG MODELS

If the regression model includes not only the current but also the lagged values of the exogenous (or explanatory) variable, like in Equation (4.10) below, it is called a 'distributed lag model'.

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t \qquad (4.10)$$

If the model includes one or more lagged values of the dependent variable among its explanatory variables, as in Equation (4.11) below, it is called an auto-regressive model.

$$Y_t = \beta_0 + \beta_1 X_t + \gamma Y_{t-1} + u_t \qquad (4.11)$$

Auto-Regressive Distributed lag models are more dynamic and can include the lagged values of both the exogenous and the endogenous variables among the set of explanatory variables. Such dynamic models are like:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \gamma Y_{t-1} + u_t \qquad (4.12)$$

Therefore, a time series may be distributed or auto-regressive or both.

4.3.1 OLS Method

Consider a finite lag structure of a distributed lag model. Let us extend the equation (4.10) by assuming that the endogenous variable (Y) depends only on the values of X variables over s periods as follows.

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + ...\beta_s X_{t-s} + u_t \qquad (4.13)$$

Let us assume that the error term (u) follows the following properties:

$$u \approx N(0, \sigma^2)$$
$$E(u_i u_j) = 0, i \neq j$$
$$E(u_i X_j) = 0, j = 1,2....k$$

If we apply OLS to estimate the parameters of (4.13), two possible consequences could arise. Firstly, if the number of lags is large but the sample is small, we may face the problem of inadequate degrees of freedom. As a result, the standard statistical tests cannot be performed. Adding additional lags will reduce the sum of squares of the estimated residuals. In other words, such lags entail the estimation of additional coefficients and an associated loss of degrees of freedom. Moreover, the inclusion of extraneous coefficients reduces the forecasting ability of the fitted model. Secondly, the existence of multicollinearity may lead to erroneous results of OLS. This is very likely because successive values of the same variable (over time) generally create multicollinearity. To avoid these difficulties, various methods are suggested. The objective is to reduce the number of lagged variables meaningfully. This can either be done by constructing a new variable from the lagged values of the variable or by imposing restrictions on β's. Three types of lag schemes commonly suggested in this context are: (a) the declining lag scheme, (b) rectangular lag scheme and (c) an inverted V lag scheme.

In the declining lag scheme, the more recent values of X are given more weights. In the rectangular lag scheme all lagged values are assigned equal weights. This method assumes each past value of X has the same effect on Y. In the 'inverted V' scheme, the weights are initially increasing and

subsequently declining. Two of the most commonly used models for selection of optimal lag criteria are: (i) the Akaike Information Criterion (AIC) and (ii) the Schwartz Bayesian Criterion (SBC). You have studied about these in Unit 3 of this course. To recall, the AIC and SBC are estimated from the model following the rule: AIC = T. *ln*(sum of squared residuals) + 2*n* and SBC = T. *ln*(sum of squared residuals) + *n*. *ln*(T) [where, *n* = number of parameters estimated, T = number of usable observations].

One question on any estimated model is: how well does it fit the data? Adding additional lags necessarily reduces the sum of squares of the estimated residuals. Addition of such lags also entail the estimation of additional coefficients with further loss of degrees of freedom. The inclusion of extraneous coefficients also reduces the forecasting performance of the fitted model. The AIC and SBC criteria trades off a reduction in the sum of squares of the residuals for a more parsimonious (frugal) model. Thus, following points are to be kept in mind while using the AIC or SBC criteria for model selection.

- Some observations are lost if we estimate a model using lagged variables.
- We must compare the models (using AIC/SBC) keeping *T* fixed.
- Decreasing *T* has a direct effect of reducing the AIC and SBC.
- As the fit of the model improves, the AIC and SBC will approach to $-\infty$.
- Lagged model A is a better fit than lagged model B, if the AIC (or SBC) for A is smaller than for B.
- In using the criteria to compare alternative models, we must estimate them over the same sample period so that they are comparable.
- Increasing the number of regressors increases *n* but could have the effect of reducing the sum of squared residuals. Thus, if a regressor has no explanatory power, adding it to the model will cause both the AIC and SBC to increase. This is not desirable in choosing an appropriate model.

Though both the AIC and SBC criteria give the same results, SBC has superior large sample properties. However, the SBC will always select a more parsimonious model than the AIC since *lnT* generally exceeds 2! To verify this, let us assume that both the criteria give the same results. Then:

$$T.\ln(SSR) + 2n = T\ln(SSR) + n.\ln(T) \Rightarrow \ln(T) = 2 \Rightarrow T = e^2 = (2.71)^2 = 7.3$$

Hence, *T* the number of observations, is assumed to be larger than 7 in a time series econometrics.

4.3.2 Adhoc Estimation Method

The explanatory variables are assumed to be non stochastic and are uncorrelated to the error term. Hence, we can apply the OLS method to estimate the parameters. We must proceed sequentially to estimate the parameters. This means, to estimate (4.13), we first regress Y_t on X_t, then

regress Y_t on X_t and X_{t-1}, then regress Y_t on X_t, X_{t-1} and X_{t-2}, and so on. This sequential procedure stops when: (i) the regression coefficients of the lagged variables start becoming statistically insignificant and/or (ii) the coefficient of at least one of the variables changes sign. For two, three and four regressors, Equation (4.13) respectively takes the form:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t \qquad (4.14)$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + u_t \qquad (4.15)$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-3} + u_t \qquad (4.16)$$

Now, from among the Equations 4.14 to 4.16, we need to choose one that is best. For this, suppose β_3 is positive in Equation (4.15) but negative in Equation (4.16). Then, Equation (4.15) will be regarded as the best equation. Evidently, this method has its limitations. These are: (i) there is no *a priori* guide as to what is the maximum length of the lag. If the lag length is incorrectly specified, we will have to contend with the problem of misspecification errors. (ii) As one estimates successive lags, there are fewer degrees of freedom left. (iii) In economic time series data, successive lags tend to be highly correlated giving rise to multicollinearity. This leads to imprecise estimation i.e. the standard errors tend to be large in relation to the estimated coefficient. As a result, based on the routinely computed *t* ratios, we may wrongly decide that a lagged coefficient is statistically insignificant. In view of these limitations, the *ad hoc* estimation procedure cannot be recommended easily. Some prior or theoretical consideration must be brought to bear upon the procedure in order to make headway with the estimation problem.

4.3.3 Koyck's Approach

The Koyck's method considers a geometric lag scheme. The Koyck's distributed lag model assumes that the weights (viz. lag coefficients) are declining continuously following the pattern of a geometric progression. We begin with writing the equation (4.13) in a slightly different form as follows.

$$Y_t = \alpha_0 + \beta_o X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \ldots\ldots\ldots\ldots\ldots + u_t \qquad (4.16)$$

As before, we assume that the error term (u) has the following properties:

$$u \approx N(0, \sigma^2)$$
$$E(u_i, u_j) = 0, i \neq j$$
$$E(u_i, X_j) = 0, j = 1,2....k$$

Koyck's geometric lag scheme implies that the more recent values of X exert a greater influence on Y than the remote values of X. Let us verify how the lag coefficients of this model decline in the form of geometric progression. We have:

$$\beta_1 = \lambda\beta_0$$

$$\beta_2 = \lambda\beta_1 = \lambda(\lambda\beta_0) = \lambda^2\beta_0$$

$$\beta_3 = \lambda\beta_2 = \lambda(\lambda\beta_1) = \lambda^2\beta_1 = \lambda^2(\lambda\beta_0) = \lambda^3\beta_0 .$$

We can therefore generalise this as:

$$\beta_i = \lambda^i\beta_0 \quad 0 < \lambda < 1$$

We can write the sum of the regression coefficients of Equation (4.16) as:

$$\sum_{i=0}^{\infty}\beta_i = \beta_0 + \beta_1 + \beta_2 + ...$$
$$\Rightarrow \beta_0 + \lambda\beta_0 + \lambda^2\beta_0 + \lambda^3\beta_0 + ...$$
$$\Rightarrow \beta_0(1 + \lambda + \lambda^2 + \lambda^3 + ...)$$
$$\Rightarrow \beta_0\left(\frac{1}{1-\lambda}\right)$$

Note that Equation (4.16) can be written as:

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + ... + u_t$$

$$Y_t = \alpha_0 + \beta_0 X_t + \lambda\beta_0 X_{t-1} + \lambda^2\beta_0 X_{t-2} + \lambda^3\beta_0 X_{t-3} + ... + u_t \qquad (4.17)$$

Equation (4.17) is difficult to estimate because we have infinite number of parameters. The parameters are also non-linear. Hence, Koyck suggested a technique known as 'Koyck transformation'. Let us consider a one period lag for Equation (4.17) as:

$$Y_{t-1} = \alpha_0 + \beta_o X_{t-1} + \lambda\beta_0 X_{t-2} + \lambda^2\beta_0 X_{t-3} + \lambda^3\beta_0 X_{t-4} + ... + u_{t-1} \qquad (4.18)$$

Multiplying Equation (4.18) by λ, we get:

$$\lambda Y_{t-1} = \lambda\alpha_0 + \lambda\beta_0 X_{t-1} + \lambda^2\beta_0 X_{t-2} + \lambda^3\beta_0 X_{t-3} + \lambda^4\beta_0 X_{t-4} + ... + \lambda u_{t-1} \qquad (4.19)$$

Subtracting Equation (4.19) from (4.17), we get:

$$\begin{aligned} Y_t - \lambda Y_{t-1} &= (1-\lambda)\alpha_0 + \beta_o X_t + (u_t - \lambda u_{t-1}) \\ Y_t &= (1-\lambda)\alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + v_t \end{aligned} \qquad (4.20)$$

where, $v_t = (u_t - \lambda u_{t-1})$. A unique feature of Koyck model is that we started with an infinite distributed lag model but ended up with an auto regressive model with only three parameters λ, α_0, and β_0 to be estimated. Thus, in Koyck's geometric lag structure we have eliminated the two basic limitations of the distributed lag models. In the process, we have achieved: (i) maximum economy of degrees of freedom (since all the lagged X's have been substituted for a single variable, Y_{t-1}) and (ii) avoided multicollinearity to an extent (since Y_{t-1} will in generally be less correlated with X_t than the successive values of the latter).

The Koyck's auto-regressive distributed lag model also suffers from certain limitations. These are as follows.

i) In the new formulation (i.e. Equation 4.20), the error term v_t is auto-correlated while the error term of the original model (u_t) was serially independent. This can be verified as follows:

$$E(v_t, v_{t+1}) = E[(u_t - \lambda u_{t-1}).(u_{t+1} - \lambda u_t) = E[u_t u_{t+1} - \lambda u_t^2 - \lambda u_{t-1} u_{t+1} + \lambda^2 u_{t-1} u_t]$$

$E(v_t, v_{t+1}) = -\lambda E(u_t^2) = -\lambda \sigma_u^2 \neq 0$, since u is serially independent. Hence:

$$E[u_t, u_{t+1}] = E[u_{t-1}, u_{t+1}] = E[u_{t-1}, u_t] = 0$$

ii) The lagged variable Y_{t-1} is not independent of the error term v_t. This means: $E(v_t, Y_t) \neq 0$. This can also be verified as follows:

$$E[v_t, Y_t] = \frac{1}{n}\sum (v_t - \bar{v})(Y_t - \bar{Y})$$

$$= \frac{1}{n}\sum (u_t - \lambda u_{t-1})[\beta_0 (X_t - \bar{X}_t) + \beta_1 (X_{t-1} - \bar{X}_{t-1}) + ...u_t]$$

Since, $v_t = u_t - \lambda u_{t-1}$, therefore, $\bar{v}_t = \bar{u}_t - \lambda \bar{u}_{t-1} = 0 \because \bar{u} = 0$. From Equation 4.16), we can estimate the value of $(Y_t - \bar{Y}_t)$. Now, we expand the above expression as:

$$\beta_0 . \frac{1}{n}\sum u_t (X_t - \bar{X}_t) + \beta_1 . \frac{1}{n}\sum u_t (X_{t-1} - \bar{X}_{t-1}) + ... + \frac{1}{n}\sum u_t^2 -$$

$$\lambda\beta_0 . \frac{1}{n}\sum u_{t-1}(X_t - \bar{X}_t) ..... - \lambda . \frac{1}{n} . \sum u_{t-1} . u_t$$

By assumptions, we have:

$$u \approx N(0, \sigma^2)$$
$$E(u_i, u_j) = 0, i \neq j$$
$$E(u_i, X_j) = 0, j = 1,2....k$$

Therefore, $Cov(v, Y) \neq 0$ (since, $\sigma_u^2 = \frac{1}{n}\sum u_t^2$).

iii) The autocorrelation of v_t, superimposed on values of Y_{t-1}, renders the OLS estimates not only biased but also inconsistent even in large samples. The OLS estimates are therefore asymptotically biased i.e. bias in small samples due to $E(Y_{t-1}, v_t) \neq 0$ and does not vanish even as n tends to ∞. Hence, the estimates are inconsistent.

iv) The combined violation of two important assumptions of OLS creates a major problem. This concerns the power of Durbin-Watson d statistics in detecting autocorrelation. Since the asymptotic bias depends on λ (viz. autocorrelation coefficient), the bias will be positive if $\lambda > 0$. Further, the bias does not vanish even as n increases. Hence, the d statistic is biased towards 2 if Y_{t-1} appears as an explanatory variable in the right hand side of the equation.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) Distinguish between distributed lag model, autoregressive model and autoregressive distributed lag model.

……

……

……

……

……

……

2) What is an unique feature of Kyock's transformation?

……

……

……

……

……

4.4 ALTERNATIVE APPROACHES TO SOLVING DISTRIBUTED LAG MODELS

We discuss four approaches in this section viz. (i) Partial Adjustment Model (PAM), (ii) Adaptive Expectations Model, (iii) Blended Approach and (iv) Almon's Approach.

4.4.1 Partial Adjustment Model

To overcome the limitations of the Koyck model, Marc Nerlove developed a model known as 'partial adjustment (or stock adjustment) model'. Here, it is hypothesised that the investment in fixed capital is based on 'stock adjustment principle'. It is assumed that there is a desired level of capital stock, Y_t^*, so as to optimise without excess capacity or overworking the existing machinery. This Y_t^* is to be determined by the level of output X_t. Therefore, Nerlove specified this as:

$$Y_t^* = \beta_0 + \beta_1 X_t + u_t \qquad (4.21)$$

We cannot directly estimate Equation (4.21) by OLS because Y_t^* is not observable. Hence, we need to replace it by postulating some behavioural principle. The 'stock adjustment principle' implies a behavioural pattern as follows. The actual change to be realised in the capital stock in any one period can only be a fraction of the desired change. This is because of the gestation period involved in all investment projects with the administrative

and financial systems required to be mobilised and geared up (*constraints*). Therefore, the adjustment of capital to the desired level can only be gradual. This 'gradual adjustment process' can be expressed in an alternative form as:

$$Y_t - Y_{t-1} = \delta\left(Y_t^* - Y_{t-1}\right) + v_t, \quad 0 < \delta \leq 1, \tag{4.22}$$

Here, $Y_t - Y_{t-1}$ is the actual change in capital stock. This means the realized investment in period *t*, $\left(Y_t^* - Y_{t-1}\right)$, is the change in capital stock with δ as the adjustment coefficient. Therefore, substituting the value of Y_t^* of Equation (4.21) in equation (4.22), we get: $Y_t - Y_{t-1} = \delta\left(\beta_0 + \beta_1 X_t + u_t - Y_{t-1}\right) + v_t$. This can be re-arranged as:

$$Y_t = (\delta\beta_0) + (\delta\beta_1)X_t + (1-\delta)Y_{t-1} + (\delta u_t + v_t) \tag{4.23}$$

Equation (4.23) can be interpreted as follows. The capital stock in any one period *t* depends partly on the level of output X_t in that period and partly on the initial level of investment Y_{t-1}. This postulates that the actual change in capital stock in any given time period *t* is some fraction δ of the desired change for that period. If δ =1, it means that the actual capital stock is equal to the desired stock i.e. the actual stock adjusts to the desired stock instantaneously. If δ = 0, it means that nothing changes since actual stock at time *t* is the same as the observed stock in the previous time period. Therefore, it is reasonable to assume $0 < \delta \leq 1$ since the adjustment to the desired capital stock is incomplete till the *constraints* are sorted out. This is the reason why the formulation is described as a 'partial adjustment model (PAM)'. Equation (4.23) is a gross investment function. The net investment can be stated as:

$$Y_t - Y_{t-1} = \Delta K = I_t = (\delta\beta_0) + (\delta\beta_1)X_t - \delta Y_{t-1} + (v_t + \delta u_t) \tag{4.24}$$

We can relate the Koyck's autoregressive model with the Nerlove's partial adjustment model. There are many similarities between the two. These can be stated as follows:

- Both the models consider the same variables Y_t, X_t and Y_{t-1}. Further, the error term in PAM does not involve any autoregressive scheme in the *u*'s as in the case of Koyck's.
- Since in PAM, the disturbance term has no direct connection with its own previous values, we can assume that the new error term (v_t+δu_t) is not auto-correlated.
- However, in PAM, if the error term appears with serial correlation, using the *d* statistic, we need to evolve some special estimation method (since the OLS fails).
- The PAM is less complicated than the Koyck's approach.
- In the PAM, the coefficient (*1–δ*) of the lagged *Y* has a clear economic meaning since it involves δ as the adjustment coefficient. Information

about the value of δ can be obtained from the firms. This enables the application of a mixed estimation procedure.

- The long-run and the short-run demand for capital stock are represented by the Equations (4.21) and (4.23) respectively. This is a special characteristic of PAM missing in the Koyck's distributed model. Once we estimate the short-run model [viz. Equation (4.23)] and get the estimate of δ, we can derive the long-run function.

4.4.2 Adaptive Expectations Model

Adaptive expectation model, developed by Cagan (1956), is an improvement of the Koyck's autoregressive model. Its general form is:

$$Y_t = \beta_0 + \beta_1 X_t^* + u_t \qquad (4.25)$$

The value of Y in any period t is taken to depend on the expected level of X_t (X_t^*). But since X_t^* is not directly observable, it is hypothesised that expectations concerning its value are formed on the adaptive principle stated as:

$$X_t^* - X_{t-1}^* = \gamma\left(X_t - X_{t-1}^*\right) \qquad (4.26_a)$$

where γ is the expectation coefficient such that $0 < \gamma \leq 1$. The value of γ being in the range of 0 to 1 means that the expectations are adaptive. This can also be stated as 'current expectations are formed based on previous expectations'. This is in the light of both the actual achievements and experience. Expectations are reformulated in each period, in the light of current expectations, as $X_t^* - X_{t-1}^*$. The change is a fraction of the difference between the currently achieved (or observed) value of the variable X_t and the previous expectations X_{t-1}^*. A gap between the actual and expected levels of output is assumed to exist. Hence, the current expectations X_t^* are partly determined by past expectations X_{t-1}^*, and partly by the adapted expectations in the light of the current experience. This means:

$$X_t^* = X_{t-1}^* + \gamma\left(X_t - X_{t-1}^*\right) \qquad (4.26_b)$$

Substituting for the unobservable variable X_t^* in the original model [viz. Equation (4.25)], we get:

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

$$\Rightarrow X_t^* = -\frac{\beta_0}{\beta_1} + \frac{Y_t}{\beta_1} - \frac{u_t}{\beta_1} \qquad (4.27)$$

Taking the lagged value by one period, we get:

$X_{t-1}^* = -\frac{\beta_0}{\beta_1} + \frac{Y_{t-1}}{\beta_1} - \frac{u_{t-1}}{\beta_1}$. Substituting X_t^* and X_{t-1}^* in the adaptive expectations Equation (4.26_a) we get:

$$\left(-\frac{\beta_0}{\beta_1}+\frac{Y_t}{\beta_1}-\frac{u_t}{\beta_1}\right)-\left(-\frac{\beta_0}{\beta_1}+\frac{Y_{t-1}}{\beta_1}-\frac{u_{t-1}}{\beta_1}\right)=\gamma\left[X_t-\left(-\frac{\beta_0}{\beta_1}+\frac{Y_{t-1}}{\beta_1}-\frac{u_{t-1}}{\beta_1}\right)\right]$$

$$\Rightarrow -\beta_0+Y_t-u_t+\beta_0-Y_{t-1}+u_{t-1}=\gamma\beta_1X_t+\gamma\beta_0-\gamma Y_{t-1}+\gamma u_{t-1}$$

$$\Rightarrow Y_t=(\gamma\beta_0)+(\gamma\beta_1)X_t+(1-\gamma)Y_{t-1}+\left[u_t-(1-\gamma)u_{t-1}\right] \quad (4.28)$$

Thus, we again arrive at an expression which contains the same variables like Y_t, X_t and Y_{t-1} as in Koyck's model and Nerlove's partial adjustment model. The adaptive expectation model is, however, appealing because it allows for expectations to be accommodated. But its error term suffers from the same defect as the error term of Koyck's model i.e. it is auto-correlated in the *u*'s of the original model. Therefore, we have the same estimation difficulties as in Koyck's model. However, this model has become popular because it can deal with 'expectations' whose importance is getting more and more recognised in monetary economics and consumption hypotheses. During rapid inflation, the quantity demanded is determined by the expected price. Similarly, as per Friedman's permanent income hypothesis, the level of consumption is determined by the expected income. Now, the expected variables are '*ex ante*' which are not observable. A variant of the 'adaptive expectations' model, often used in practice, includes X_{t-1} instead of X_t. This replacement is done because when expectations are formed in period *t*, the current level of *X* (viz. X_t), is usually unknown. Hence, we may replace it by X_{t-1}, the most recent available information on *X*. Such replacement implies the behavioural rule driven by expectation. Therefore, we now have the revised equation as:

$$X_t^*-X_{t-1}^*=\gamma\left(X_{t-1}-X_{t-1}^*\right) \quad (4.29)$$

4.4.3 Blended Approach

A blended approach is one which includes both the principles of 'partial adjustment' and 'adaptive expectation'. In the 'partial adjustment', Y_t is replaced by its desired level Y_t^* as: $Y_t^*=a_0+a_1X_t+u_t^*$. In the case of 'adaptive expectations', X_t is replaced by its 'expected value', X_t^* as: $Y_t=c_0+c_1X_t^*+u_t^*$.

Note that in both the versions, considerations about future levels of *X* and *Y* are involved. Therefore, combining both the desire and expectations in one single equation we get:

$$Y_t^*=\beta_0^*+\beta_1^*X_t^* \quad (4.30)$$

Equation (4.30) is based on the two behavioural principles stated as one. This is stated as 'the desired level of *Y* depending on the expected level of *X*'. Therefore, we can state the two behavioural rules as:

$$Y_t-Y_{t-1}=\delta\left(Y_t^*-Y_{t-1}\right)+u_t \quad 0<\delta\leq 1 \quad (4.31)$$

$$X_t^* - X_{t-1}^* = \gamma\left(X_t - X_{t-1}^*\right) \quad 0<\gamma\leq 1 \qquad (4.32)$$

In (4.32), we do not add the error term. This is a theoretical modelling which states that economic agents will adapt their expectations in the light of the past experience and in particular they will learn from their mistakes. It is therefore believed that Equation (4.32) is truly an equilibrium relation with no scope for adding an error term. However, the adjustment mechanism [viz. coefficient of expectation (γ)] can be imperfect and may require the disturbance term. We proceed to solve (4.31) for Y_t^* as follows.

$$\delta Y_t^* = Y_t - Y_{t-1} + \delta Y_{t-1} - u_t \Rightarrow Y_t^* = \left(\frac{1}{\delta}\right)Y_t + \left(\frac{\delta-1}{\delta}\right)Y_{t-1} - \left(\frac{1}{\delta}\right)u_t \qquad (4.33)$$

Substituting this value of Y_t^* in equation (4.30), we get:

$Y_t^* = \left(\frac{1}{\delta}\right)Y_t + \left(\frac{\delta-1}{\delta}\right)Y_{t-1} - \left(\frac{1}{\delta}\right)u_t = \beta_0^* + \beta_1^* X_t^*$. Solving for X_t^* we get:

$$X_t^* = \left(\frac{1}{\beta_1^*\delta}\right)Y_t + \left(\frac{\delta-1}{\beta_1^*\delta}\right)Y_{t-1} - \left(\frac{1}{\beta_1^*\delta}\right)u_t - \frac{\beta_0^*}{\beta_1^*} \qquad (4.34)$$

Taking the lagged value by one period, we have:

$$X_{t-1}^* = \left(\frac{1}{\beta_1^*\delta}\right)Y_{t-1} + \left(\frac{\delta-1}{\beta_1^*\delta}\right)Y_{t-2} - \left(\frac{1}{\beta_1^*\delta}\right)u_{t-1} - \frac{\beta_0^*}{\beta_1^*} \qquad (4.35)$$

Now, substituting the values of X_t^* and X_{t-1}^* (as in Equations 4.34 and 4.35) in Equation (4.32), we get:

$$X_t^* - X_{t-1}^* = \gamma\left(X_t - X_{t-1}^*\right) \Rightarrow X_t^* - X_{t-1}^* + \gamma X_{t-1}^* = \gamma X_t \Rightarrow X_t^* - (1-\gamma).X_{t-1}^* = \gamma X_t$$

$$\left(\frac{1}{\beta_1^*\delta}\right)Y_t + \left(\frac{\delta-1}{\beta_1^*\delta}\right)Y_{t-1} - \left(\frac{1}{\beta_1^*\delta}\right)u_t - \frac{\beta_0^*}{\beta_1^*} - \left(\frac{1-\gamma}{\beta_1^*\delta}\right)Y_{t-1} - \left(\frac{(1-\gamma)(\delta-1)}{\beta_1^*\delta}\right).Y_{t-2} + \left(\frac{1-\gamma}{\beta_1^*\delta}\right)u_{t-1}$$
$$+ \left(\frac{(1-\gamma)\beta_0^*}{\beta_1^*}\right) = \gamma X_t$$

Multiplying both sides by $\delta\beta_1^*$ and re-arranging we get:

$$Y_t + (\delta-1).Y_{t-1} - u_t - \delta\beta_0^* - (1-\gamma).Y_{t-1} - (1-\gamma)(\delta-1).Y_{t-2} + (1-\gamma).u_{t-1} + \delta\beta_0^*(1-\gamma)$$
$$= \delta\beta_1^*\gamma X_t$$

$$Y_t = (\delta\gamma\beta_0^*) + (\delta\gamma\beta_1^*)X_t + \left[(1-\gamma) + (1-\delta)\right)\right]Y_{t-1} - \left[(1-\gamma)(1-\delta)\right]Y_{t-2} + \left[u_t - (1-\gamma)u_{t-1}\right] \qquad (4.36)$$

How does (4.36) differ from the 'partial adjustment' and the 'adaptive expectation' models? This contains the additional variable Y_{t-2} but the γ and δ parameters appear symmetrically. It is therefore impossible to obtain separate

estimates of each one of them. However, we can obtain estimates for (γ+δ) and ($\gamma\delta$) and thereby of β_0^* and β_1^*.

4.4.4 Almon's Approach

Almon proposed the following method for estimating the parameters of the lagged exogenous variables. The lagged model here is assumed to be finite including only the exogenous lagged variables as follows:

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + ...\beta_s X_{t-s} + u_t \qquad (4.37)$$

Instead of attempting to directly estimate all the β's [which are (s+1) in number] by applying the OLS to the Equation (4.37), Almon proposed an 'indirect method'. This is done by assuming that β's in the lagged model can be approximated by some polynomial function such that $\beta \approx f(z)$. The function $f(z)$ is unknown if we do not make any prior assumptions about its form. The assumption made therefore is that the function *f(z)* may be approximated by a polynomial in *z* of the r^{th} order as:

$$f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + ...\alpha_r z^r \qquad (4.38)$$

Such a function $f(z)$ yields approximate values of β*'s* if we know α's and the degree of the polynomial (*r*).

Check Your Progress 3 [answer within the space given in about 50-100 words]

1) What is an essential common feature and an important difference between the Koyck's and Nerlove's PAM?

   ……

   ……

   ……

   ……

2) In which respect is the 'adaptive expectations model' similar to the Koyck's and Nerlove's models?

   ……

   ……

   ……

   ……

3) In what respect is the blended model different from Koyck's and Nerlove's? What is a particularly notable feature of the blended model?

   ……

   ……

   ……

……

……

4.5 LET US SUM UP

The unit draws on various real life examples based on economic principles. This is done with a view to understanding how a regressand may respond to a regressor(s) with a time lag. There are two types of lagged models: distributed-lag and autoregressive. In the former, the current and lagged values of regressors are explanatory variables. In the latter, the lagged value(s) of the regressand appear as explanatory variables. A purely distributed-lag model can be estimated by OLS, but the problem of multicollinearity surfaces. This is due to the successive lagged values of a regressor tending to be correlated. Different dynamic models developed by Koyck, Nerlove, Cagan, Almon and a blended model (incorporating the partial adjustment and adaptive expectations) are discussed in the unit. A unique feature of the Koyck's model, adaptive expectations model and the partial adjustment model is that they all are autoregressive in nature. This is in the sense of the lagged value(s) of the regressand appearing as one of the explanatory variables. Auto-regressive character poses estimation challenges. This is because, if the lagged regressand is correlated with the error term, the OLS estimators of such models are not only biased but are also inconsistent. Bias and inconsistency are thus the case with both the Koyck and the adaptive expectation models. The partial adjustment model is different in this respect since it can be consistently estimated by OLS despite the presence of the lagged regressand. An alternative to the lagged regression models is the Almon polynomial distributed-lag model. This avoids the estimation problems associated with the autoregressive models. A major problem with the Almon approach, however, is that one must pre-specify both the lag length and the degree of the polynomial.

4.6 KEY WORDS

Auto-Regressive Model of Regression : An autoregressive model is one in which a value from a time series is regressed on previous values from that same time series. In this type of models, the dependent variable in the previous time period becomes the predictor.

Distributed Lag Models : This is a model for time series data in which a regression equation is used to predict the current values of a dependent variable based on both the current values of an explanatory variable and the lagged values of this explanatory variable.

Adaptive : Adaptive expectations is a hypothesized process by

Expectations Model which people form their expectations about what will happen in the future based on what has happened in the past. For instance, if inflation has been higher than expected in the past, people would revise expectations for the future.

4.7 SUGGESTED BOOKS FOR FURTHER READING

1) Kmenta J, Elements of Econometrics, Macmillan, New York.

2) Johnston J, Econometric Methods, McGraw-Hill.

3) Gujarati D, Porter D C and Gunasekhar S (2018). Basic Econometrics, 5th Edition, McGraw Hill Education (India) Pvt. Ltd.

4) G. S. Maddala and Kajal Lahiri (2009). Introduction to Econometrics, 4th Edition, Wiley.

5) Koutsoyiannis, A. (2001). Theory of Econometrics, Palgrave Macmillan Limited.

4.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Lags are time related delays between cause and effect. Examples are: relative income hypothesis of consumption, accelerator theory of investment, farmer's decision to produce a commodity based on last year's price for that commodity, etc.

2) (i) effectiveness of an economic policy, (ii) contractual constraints, (iii) technology and imperfect information, (iv) psychological reasons and (v) decision delay.

Check Your Progress 2

1) If the lags are there only in the exogenous variables, with both their current and lagged values appearing together as exogenous variables, then the models are called as 'distributed lag models'. If the lags are there only in the endogenous variables and they appear as exogenous regressors, then the model is called as 'autoregressive model'. If lags of both the exogenous and endogenous variables appear among the set of explanatory variables (as in Equation 4.12) then such models are called as 'autoregressive distributed lag models'.

2) The transformation results in finite number of parameters to be estimated.

Check Your Progress 3

1) (i) Both the models consider the same variables Y_t, X_t and Y_{t-1}. (ii) The error term in PAM does not involve any autoregressive scheme in the u's as in the case of Koyck's.

2) It is in respect of common variables Y_t, X_t and Y_{t-1}.

3) It contains X_t, Y_{t-1} and Y_{t-2}. Its two parameters γ and δ cannot be separately estimated but can be estimated as their sum or the product. The original parameters (β_0^* and β_1^*) need to be estimated from these.

UNIT 5 BINARY DEPENDENT VARIABLE MODELS*

Structure

5.0 OBJECTIVES

After reading this unit, you will be able to:

- specify the nature of variables in the 'binary dependent variable models';
- define a 'linear probability model (LPM)' with an illustration;
- discuss why the OLS method of estimation is not applicable in case of 'binary dependent variable models';
- state why the conventional R^2 is not of use in case of 'binary response models';
- indicate the form of a Logit model indicating why the OLS method of estimation cannot be applied here;

* Prof. Sushil Haldar, Jadavpur University.

- list the properties of a Logit model;
- outline the estimation procedure for the parameters of a Logit model;
- explain the interpretation of Logistic parameters;
- illustrate the interpretation of logistic parameters with an empirical exercise;
- write a short note on the application of Logit model for 'grouped data';
- describe the application of Probit model with an illustration for the estimation of its 'marginal effect';
- compute the 'marginal effect' in an estimated Probit model for a 'mean value'; and
- present the relationship between the regression coefficients of: (i) Logit model and Probit model and (ii) LPM and Logit model.

5.1 INTRODUCTION

In applied economics, we face situations where the 'dependent variable' is binary or dichotomous or qualitative. Some examples are as follows: Banks provide loans to the customers. Some customers may be defaulters who fail to re-pay their loans to the banks in time. Therefore, the borrowers can be divided into two groups: defaulters and non-defaulters. In medical science, a group of people of a fixed age cohort is diagnosed about the incidence of diabetes. Some persons may suffer from diabetes while others may not. Thus, two groups of persons emerge – diabetic and non-diabetic. Numerous other examples can be given like: vegetarian and non-vegetarian, smokers and non-smokers, drug addicted vs. non-drug addicted, owning a car or not, etc. In all these cases, the dependent variable is qualitative and can be quantified as 1 and 0. Now the question that arises is about the factors that influence the probability of being, say, a defaulter. Is it due to the age of the customer? Or is it due to the cultural characteristics? Or is it due to the marital status? Or is it due to the overall socio-economic profile? We can similarly raise many questions pertaining to the incidence of diabetes, smokers vs. non-smokers, vegetarian vs. non-vegetarian, etc. In most cases of regression analysis, we consider the dependent variable as quantitative or numeric whereas the independent variables may be either quantitative or qualitative or a mixture of both the types. In this Unit, we consider the dependent variable as not only qualitative but also as dichotomous. For instance, if the dependent variable is 'yes' or 'no' type, then we define the dependent variable as binary (viz. 1 or 0) and proceed.

5.2 NATURE OF VARIABLES

In two variable regression models, we try to establish an average mathematical relationship between the two variables. The dependent variable is termed in various ways (e.g. regressand or endogenous or response or explained) whereas, the independent variables are termed as regressors (or

explanatory or exogenous variables). Let us now suppose that we want to study the ownership of a car of an individual. Car being a luxury good, owning it depends on the income level. It may also depend on where the person resides i.e. in urban or rural areas. In spite of such factors, a person may still own a car or not. We therefore define as 1, if the person owns a car, 0 otherwise. The two variables regression equation can then be written as:

$$y_i = \alpha + \beta x_i + u_i \text{.....} \quad (5.1)$$

where, $y_i = 1$ if owning a car; $y_i = 0$ otherwise. Here, α and β are the parameters to be estimated given the level of income x_i. Now, is the standard 'method of least squares' applicable? What are the possible consequences if we apply the 'method of least squares'? These issues are dealt with in this section. In a model where y_i is quantitative, our objective is to estimate its expected or mean value given the values of the explanatory variables. The regressor(s) can either be quantitative or qualitative. If the regressor is qualitative [like locality of dwelling (rural or urban), gender (male or female), religion (Hindu or Non-Hindu), caste (backward or not-backward), etc.] we can estimate the influence of such categorical variables on the dependent variable by considering them as dummy variables. This means, both the dependent as well as independent variables may be simultaneously, dichotomous. In such cases, Equation (5.1) can be written as:

$$y_i = \alpha_0 + \alpha_1 D_i + u_i \quad (5.2)$$

where, $D_i = 1$ if the person resides in urban area, 0 otherwise. The dependent variable (viz. owning a car) remains the same. The parameter α_1 determines the influence of residential status on having a car or not. We can incorporate more variables on the right hand side. For instance, we can include both the income and residential status as follows:

$$y_i = \alpha + \beta x_i + \alpha_1 D_i + u_i \quad (5.3).$$

In models where y is qualitative, our objective is to find the probability of something happening (e.g. owning a car or owning of a house or belonging to any category or group). Such qualitative response regression models are known as 'probability models'. In this unit, we discuss three alternative approaches for estimating the parameters in such binary response models. These are: (i) The Linear Probability Model (LPM), (ii) The Logit Model and (iii) The Probit Model.

5.3 THE LINEAR PROBABILITY MODEL (LPM)

In order to explain the linear probability model (LPM), we consider a model like:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (5.4)$$

where, x_i is the income (independent variable) and y_i is the dependent variable referring to the ownership of a car. Note that Equation (5.4) has the conditional expectation of y_i given x_i i.e. $E(y_i|x_i)$. This is interpreted as the conditional probability that the event y_i will occur given x_i, that is 'probability of $(y_i|x_i)$. This is the reason why a model like (5.4) is also known as the 'Linear Probability Model (LPM)'. Hence, the expectation of owning a car given a certain level of income can also be written as:

$$E(y_i / x_i) = \beta_0 + \beta_1 x_i + u_i \qquad (5.5a)$$

Now, if p_i is the probability that y_i=1 and $1 - p_i$ is the probability that $y_i = 0$, the variable y_i has its probability distribution as in Table 5.1.

Table 5.1: Probability Distribution of y_i

y_i	**Probability**
0	$1 - p_i$
1	p_i
Total	**1**

Thus, y_i follows the Bernoulli Distribution. By applying mathematical expectation, we get:

$$E(y_i) = 0.(1 - p_i) + 1.p_i = p_i \quad \text{or}$$

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i = p_i \qquad (5.5b)$$

Extending the above to '*n*' independent trials, each with probability '*p*' for success and probability (*1–p*) for failure, and if *X* of these trials represent the number of successes, then we know that *X* follows the binomial distribution with mean '*np*' and variance '*npq*' where *q=(1–p)*. Since the probability p_i must lie between 0 and 1, we have the restriction:

$$0 \le E(y_i \mid x_i) \le 1 \qquad (5.6)$$

Equation (5.6) means that the conditional expectation (or conditional probability) of y_i lies between 0 and 1. Let us now examine whether the ordinary least squares (OLS) method for estimating the model (5.5) works. We will see that we encounter three specific types of problems viz. (i) non-normality of the disturbances u_i, (ii) heteroscedastic variances of the disturbances and (iii) non-fulfilment of conditional probability: $0 \le E(y_i|x_i$.

5.3.1 Non-Normality of Disturbances

Recall that OLS does not require the disturbances (u_i) to be normally distributed. However, we need to examine 'normality' for the purpose of

'statistical inference'. Note that like y_i, the disturbances u_i also can take only two values which follow Bernoulli distribution as shown in Table 5.2.

Table 5.2: Probability Distribution of u_i

y_i	u_i	Probability
When y_i=1	$1-\beta_0-\beta_1 x_i$	p_i
When y_i= 0	$-\beta_0-\beta_1 x_i$	$1-p_i$
Total	**-**	**1**

Hence, u_i also follows Bernoulli distribution. Therefore, as the sample size increases indefinitely, the OLS estimators of Equation (5.5) tend to be normally distributed. As a result, in large samples, the statistical inference of the LPM can be made by estimating the parameters of (5.5) by the OLS procedure (under the normality assumption). However, the non-normality continues to prevail as can be seen from Fig.5.1 in which vertically we measure y (which takes the two values 1 and 0) and horizontally we measure x. The expected value of y, or the conditional probability, is shown by the line $E[y|x]$. We can see from the Figure that the error term is not normally distributed since it has two possible values for given value of x i.e. one value given by A (when y = 1) and the other value given by B when y = 0. At point A, y = 1 and u_i = 'observed value – predicted value' or: $u_i = 1-\Pr[y \mid x] = 1-\beta_0-\beta_1 x_i$. Similarly, at point B, the value of y being zero, it yields, $u_i = 0-\beta_0-\beta_1 x_i$. The curve has no scope of becoming bell shaped. Hence, u_i is not normally distributed.

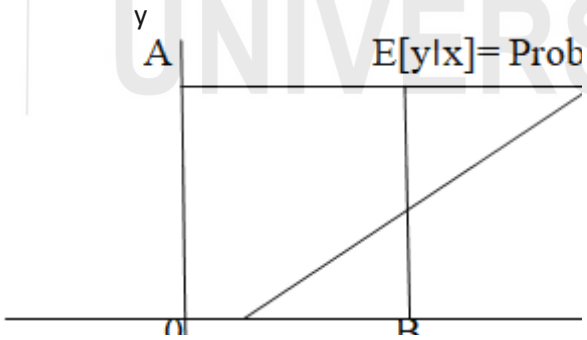

Fig. 5.1: Graphical Presentation of a LPM

5.3.2 Heteroscedastic Variances

Even if $E(u_i)=0$ and $\text{cov}(u_i,u_j)=0$ for $i \neq j$ (i.e. no serial correlation), it cannot be maintained that in the LPM the disturbances are homoscedastic. As statistical theory shows, for a Bernoulli distribution the theoretical mean and variance are p and $p(1-p)$ respectively (p being the probability of success). This shows that the variance is a function of the mean. Hence, the variance of error terms are heteroscedastic. This can also be formally seen. Since the probability distribution of 'u_i' is as in Table 5.2, we have:

$$Var(u_i) = (1-\beta_0-\beta_1 x_i)^2.p_i + (-\beta_0-\beta_1 x_i)^2.(1-p_i)$$
$$= \left[1-2(\beta_0+\beta_1 x_i)+(\beta_0+\beta_1 x_i)^2\right].p_i + (\beta_o+\beta_1 x_i)^2 - (\beta_0+\beta_1 x_i)^2.p_i$$
$$= p_i - 2(\beta_0+\beta_1 x_i).p_i + (\beta_0+\beta_1 x_i)^2.p_i + (\beta_o+\beta_1 x_i)^2 - (\beta_0+\beta_1 x_i)^2.p_i$$
$$= p_i - 2(\beta_0+\beta_1 x_i).p_i + (\beta_o+\beta_1 x_i)^2$$
$$= p_i.(1-p_i)$$

Since from (5.5b), $\beta_0 + \beta_1 x_i = p_i$ (5.7)

It follows from Equation (5.7) that the expected value of u_i is a function of x and hence not a constant. This is because 'the expected value of u_i is contained in the *Var* (u_i). This tells us that the *Var* (u_i) also depends on the values of x and hence is not homoscedastic. Because of this heteroscedasticity, the OLS estimators will be unbiased but inefficient i.e. they do not have the minimum variance property. One way to resolve the problem of heteroscedasticity is to transform the Equation (5.4) by dividing it throughout by $\sqrt{p_i(1-p_i)} = \sqrt{w_i}$.

By doing this, we get:

$$\frac{y_i}{\sqrt{w_i}} = \frac{\beta_0}{\sqrt{w_i}} + \frac{\beta_1 x_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \quad (5.8)$$

The error term in (5.8) will now be homoscedastic. Therefore, after estimating Equation (5.4), we can estimate Equation (5.8) by OLS. This is known as the 'weighted least squares (WLS)' method where w_i serves as the weights. But in practice, the true $E(y_i / x_i)$ is unknown. Hence, the weights w_i are also unknown. To estimate w_i, we can proceed as follows. Run the OLS regression in Equation (5.4) despite the presence of heteroscedasticity and obtain $\hat{y}_i = E(y_i \mid x_i)$. Then obtain $\hat{w}_i = \hat{y}_i(1-\hat{y}_i)$ as the estimate of w_i. Use this w_i to transform the initial equation and obtain the new transformed equation. We can now estimate the transformed equation by OLS (WLS).

5.3.3 Non-fulfilment of Probability Limits

Since $E(y_i / x_i)$ in the linear probability models measures the conditional probability of the event y_i occurring given x_i, it must necessarily lie between 0 and 1. Although this is true a priori, there is no guarantee that $\hat{y}_i$, the estimators of $E(y_i / x_i)$, will necessarily fulfill this condition. This is the real problem with the OLS estimation of the LPM. There are two ways of finding out whether the estimated $\hat{y}_i$ lie between 0 and 1. One is to estimate the LPM by the usual OLS method and find out whether the estimated $\hat{y}_i$ lie between 0 and 1. If some values are less than 0 (that is, negative), $\hat{y}_i$ is assumed to be zero for those cases; if they are greater than 1, they are

assumed to be 1. The second procedure is to devise an estimating technique that will guarantee that the estimated conditional probabilities $\hat{y}_i$ lie between 0 and 1. The Logit and Probit models (discussed later) will guarantee that the estimated probabilities will indeed lie between the logical limits of 0 and 1.

5.3.4 R^2 as Measure of Goodness of Fit

The conventional R^2 is of limited value in the Binary or dichotomous response models where corresponding to a given *X*, *Y* is either 0 or 1. Therefore, all the *Y* values will either lie along the *X* axis or along the line corresponding to 1 (as shown in Fig. 5.1). Therefore, the LPM is not expected to fit such a scatter well. But if a LPM is estimated in such a way that it will not fall outside the logical band 0–1, the conventional R^2 is likely to be much lower than 1 for such models. Therefore, in practical applications, R^2 is of little use in case of binary dependent models.

5.3.5 Limitations of the LPM and Corrective Measures

The LPM suffers from non-normality of u_i, heteroscedasticity of u_i, possibility of the predicted value of '*y*' lying outside the 0–1 range and a very low value of R^2. However, we can use WLS to resolve the heteroscedasticity problem and increase the sample size to minimise the non-normality problem. By resorting to restricted least-squares or mathematical programming techniques, we can even make the estimated probabilities lie in the 0–1 interval. But, a fundamental problem with the LPM is that it is not logically a very attractive model because it assumes that $p_i = E(y_i \mid x_i)$ increases linearly with x_i i.e. the marginal or incremental effect of *x* remains constant throughout. This is unrealistic. In reality, one would expect that p_i is nonlinearly related to x_i. This is because, at very low income a family will not own a car. But at a sufficiently high level of income, say x^*, it is most likely to own a car. Any increase in income beyond x^* will have little effect on the probability of owning a car. Thus, at both ends of the income distribution, the probability of owning a car will be virtually unaffected by a small increase in income (*x*). Therefore, what we need is a (probability) model that has these two features: (i) as x_i increases, $p_i = E(y_i \mid x_i)$ should also increase but not step outside the 0–1 interval, and (ii) the relationship between p_i and x_i is nonlinear i.e. "one which approaches 'zero' at slower and slower rates as x_i gets small" and "one which approaches 'one' also at slower and slower rate as x_i gets very large". Such a model is the Logit model discussed next.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) What is the nature of variables we will have in a 'binary dependent variable model'?

……

……

……

……

……

2) Can the method of OLS be applied for estimating the parameters of a 'binary dependent variable model'? Give reasons.

……

……

……

……

……

3) What method is applicable to overcome the problem of heteroscedasticity in estimating a binary dependent variable model like in Equation (5.4)?

……

……

……

……

……

4) Why is R^2 not a good measure of fitness of the model in case of binary dependent variable models?

……

……

……

……

5.4 THE LOGIT MODEL

The logistic regression, or the Logit model, overcomes the limitations of LPM in case of dummy dependent variable models. We continue with the conditional expectation of y as in the LPM. That is:

$$p_i = \beta_0 + \beta_1 x_i \qquad (5.9)$$

where x_i is income and $p_i = E(y_i \mid x_i)$ is the conditional probability of the individual owning a car. Now, consider the following representation of car ownership between income and probability of owning a car:

$$p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}} = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \qquad (5.10)$$

where, $z = \beta_0 + \beta_1 x_i$ so that if z tends to $-\infty$, p tends to zero and if z tends to ∞, p tends to 1. Hence, z_i ranges from $-\infty$ to $+\infty$ and p_i ranges between 0 and 1. Here, p_i is nonlinearly related to z_i. This is an important property of a logistic Equation like (5.10). Graphically, this is as shown in Fig. 5.2.

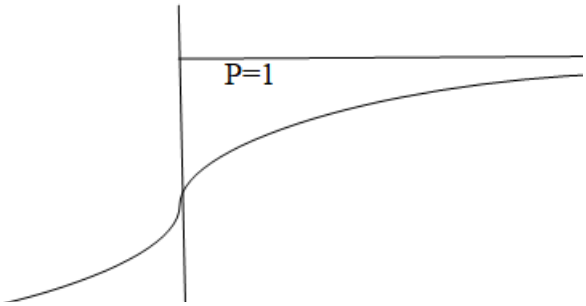

Fig. 5.2: Logistic Curve

The question now is, can we apply the OLS to estimate the parameters of the Equation (5.10)? The answer is 'no'. This is because p_i is nonlinear not only in x but also in the β's. This means we cannot use the OLS procedure to estimate its parameters. If p_i is the probability of owning a car, then $(1 - p_i)$ is the probability of not owning a car. Therefore, we consider the ratio of odds (or the 'odds ratio') as $\frac{p_i}{1-p_i}$ and consider estimating '$1 - p_i$'. This means, we estimate:

$$1 - p_i = 1 - \frac{e^{z_i}}{1+e^{z_i}}$$

$$1 - p_i = \frac{1+e^{z_i} - e^{z_i}}{1+e^{z_i}}$$

$$1 - p_i = \frac{1}{1+e^{z_i}} \quad \text{.....(5.11)}$$

Equation (5.11) gives us the probability of not owning a car. Therefore, we can write the 'odds ratio' as:

$$\frac{p_i}{1-p_i} = \frac{\frac{e^{z_i}}{1+e^{z_i}}}{\frac{1}{1+e^{z_i}}} = e^{z_i} \quad \text{.....(5.12)}$$

The ratio of the probabilities as in Equation (5.12) is known as 'odds ratio in favour of owning a car'. This is the 'ratio of the probability that an individual will own a car to the probability that he will not own a car'. Numerically, this means, if $p_i = 0.8$, the odds are '4 to 1' [since $p_i/(1 - p_i) = 0.8/0.2 = 4$] in favour of the individual owning a car. Now, taking the natural log of (5.12), we get:

$$L_i = \ln\left(\frac{p_i}{1-p_i}\right) = \ln e^{z_i} = z_i$$

$$L_i = z_i = \beta_0 + \beta_1 x_i \quad \text{.....(5.13)}$$

In (5.13), L_i is the log of odds ratio. It is linear in x_i. It is also linear in the parameters. L_i or (5.13) is called the 'Logit model'.

5.4.1 Properties of Logit Model

Can we directly estimate Equation (5.3) or L_i if $p_i = 0$ and 1? The answer is 'no'. This is because, if $p_i = 0$, Logit (L_i) is undefined. Likewise, if we p_i =1, L_i is infinity. This poses the question, what is the limiting value of L_i? Since, Z_i and L_i are both same, and Z lies between $-\infty < Z < +\infty$ (from the logistic curve), L_i also lie between $-\infty$ and $+\infty$. Therefore, as p_i goes from 0 to 1, the Logit L_i goes from $-\infty$ to $+\infty$. That is, although the probabilities lie between 0 and 1, the Logit is differently bounded. This can be explained as follows.

We have, $L_i = \ln(\frac{p_i}{1-p_i}) = \ln e^{z_i} = z_i$.

Hence: $\frac{p_i}{1-p_i} = e^{z_i}$ (5.14)

where, if $z_i \to -\infty, p_i \to 0$ and if $z_i \to \infty, p_i \to 1$. Thus, although L_i is linear in x_i, the probabilities themselves are not. If L_i i.e. the Logit is positive, it means when the value of regressor increases, the odds that the regressand equals 1 (i.e. y = 1 or some event of interest happens) increases. In order to explain this further, let us assume, that the value of L_i = 0.7. This means $\ln\frac{p}{1-p} = 0.7 \Rightarrow \frac{p}{1-p} = e^{0.7} = (2.71)^{0.7} = 2.009$. This is greater than 1. Hence, the odds in favour of 'happening of the event' increases as the regressor (x) increases. This can be interpreted in a different way. One can verify that the value of p = 0.667 as follows:

$$\ln\frac{p}{1-p} = 0.7 \Rightarrow \frac{p}{1-p} = e^{0.7} = (2.71)^{0.7} = 2.009$$

$$p = 2.009 - 2.009p$$
$$\Rightarrow p(1 + 2.009) = 2.009$$
$$\Rightarrow p = \frac{2.009}{3.009}$$
$$\Rightarrow p = 0.667$$

This means that the probability of the event's (y = 1) happening, is 66.7 percent, which is higher than 50 percent. Now, let us see how we can interpret the Logit, when it is negative, say L_i= -0.7? This means:

$\ln\frac{p}{1-p} = -0.7 \Rightarrow \frac{p}{1-p} = e^{-0.7} = (2.71)^{-0.7} = 0.497$. This is less than one. This means the odds in favour of the event's (y = 1) happening decreases as the regressor (x) increases. This can also be interpreted in probability terms. The

probability of the event's is happening 33.19 percent (as can be seen by the following calculations) if the regressor increases. Given that:

$$\frac{p}{1-p} = 0.497$$
$$\Rightarrow p = 0.497 - 0.497p$$
$$\Rightarrow p(1+0.497) = 0.497$$
$$\Rightarrow p = \frac{0.497}{1.497}$$
$$\Rightarrow p = 0.331$$

This is also less than 0.5 or 50 percent! Hence, if L is negative, the odds that the regressand equals 1 decreases as the value of x increases. To put it differently, the Logit becomes negative and increasingly large in magnitude as the odds ratio decreases from 1 to 0. It also becomes increasingly large and positive as the odds ratio increases from 1 to infinity. More formally, the interpretation of the Logit model given in (5.13) is as follows: β_1, the slope, measures the change in L_i for a unit change in x_i i.e. it tells us how the log-odds changes in favour of owning a car as income changes by a unit. The intercept β_0 is the value of the log odds in favour of owning a car if income is zero. Given a certain level of income, say, x^*, we can estimate the odds in favour of owning a car if β_0 and β_1 are known. Here, we have considered only one regressor but one can add as many regressors (including categorical) as may be dictated by the underlying theory.

5.4.2 Estimation of the Logit Model

Let us add the error term in Equation (5.13). We get:

$$L_i = \ln(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_i + u_i \quad (5.15)$$

In order to estimate the parameters, we need data. Data may be of two types i.e. either for a group or at the individual level. Here, we mainly consider the individual level data. If we have data at individual level, OLS estimation of the above equation is infeasible. This can be explained as follows. Consider our earlier case of owning a car. In terms of the data at the individual level, p_i = 1 if the individual owns a car and p_i = 0 if he does not own. But if we put these values directly into the Logit (L_i) we obtain, $L_i = \ln\left(\frac{1}{0}\right)$ which is infinite if the individual owns a car and $L_i = \ln\left(\frac{1}{0}\right)$ which is undefined for the case of individual not owning a car. These expressions are therefore meaningless. Hence, if we have data at the micro, or individual level, we cannot estimate (5.15) by the standard OLS method. In this situation, we have to resort to the 'maximum-likelihood (ML)' method of estimating the parameters. In our earlier logistic Equation (5.10), we assumed that the

probability of owning a car follows logistic equation. We cannot observe p_i, rather we can only observe $y = 1$ if the individual owns a car, 0 otherwise. Since, y_i is a Bernoulli random variable, we can write $\Pr(y_i = 1 \mid x) = p_i$ and $\Pr(y_i = 0 \mid x) = 1 - p_i$. Now, suppose we have a random sample of 'n' number of observations. Letting $f_i(y_i)$ denote the probability that $y_i = 1$ or 0; the joint probability of observing the '$n.y$' values i.e. $f(y_1, y_2 .....y_n)$ is given by:

$$f(y_1, y_2 .....y_n) = \prod_1^n f_i(y_i) = \prod_1^n p_i^{y_i} .(1 - p_i)^{1-y_i} \qquad (5.16)$$

We write the 'joint probability density function' as a product of individual density functions because each y_i is drawn independently and each y_i has the same logistic density function. The joint probability as given in equation (5.16) is known as the 'likelihood function'. In (5.16), we have skipped the constant term of the Bernoulli distribution (like ${}^N C_X$ in binomial distribution) because it is independent of response or outcome variable (like y_i in our case). Now, we take log of (5.16) and denote it as 'Log Likelihood Function (*LLF*)' to get:

$$\begin{aligned} LLF &= \sum_1^n \left[y_i . \ln p_i + (1 - y_i) . \ln(1 - p_i) \right] \\ &= \sum_1^n \left[y_i . \ln p_i - y_i . \ln(1 - p_i) + \ln(1 - p_i) \right] \\ &= \sum_1^n \left[y_i . \ln \left(\frac{p_i}{1 - p_i} \right) \right] + \sum_1^n \ln(1 - p_i) \\ &= \sum_1^n y_i \left(\beta_0 + \beta_1 x_i \right) - \sum_1^n \ln \left(1 + e^{(\beta_0 + \beta_1 x_i)} \right) \end{aligned} \qquad (5.17)$$

Thus, the values of $\ln \left(\frac{p_i}{1 - p_i} \right)$ and $\ln(1 - p_i)$ are estimated as $(\beta_0 + \beta_1 x_i)$ and $-\ln(1 + e^{\beta_0 + \beta_1 x_i})$ respectively. The objective is to maximise the LLF i.e. Equation (5.17) with respect to β_0 and β_1 using the values of x which are known. The resulting solutions become non-linear in the parameters because of the presence of $e^{\beta_0 + \beta_1 x}$ term. This is the reason why the OLS method cannot be applied for the Logit model. Once the values of the parameters are known, we can easily estimate the logistic equation.

5.4.3 Interpretation of Logistic Parameters

Once the Logit model is estimated, the most important part is to interpret the estimated parameters. For this, we will proceed to analyse the Logit coefficients, by continuing with the earlier example of ownership of car. Though one can include more explanatory variables in the Logit model, we restrict ourselves to considering only one independent variable (i.e. level of income). We further assume that the estimated Logit model is:

$$\hat{L} = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad (5.18)$$

where, $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated parameters. If income (say, x_i) is zero, then $\hat{L}$ is $\hat{\beta}_0$. This means the log of odds is $\hat{\beta}_0$ if income is zero. How do we interpret this? Now, assume that the value of $\hat{\beta}_0$ is 1. Then, the odds in favour of owing a car is: 2.71! This is higher than 1. What happens to log of odds or odds if the estimated value of $\hat{\beta}_0$ is –1? The odds in favour of owning a car is reduced to 0.369 if $\hat{\beta}_0$ becomes –1! This can be verified as follows: From (5.14), if income x_i is zero, then $\hat{L}$, the estimated L can be written as:

$\hat{L} = \ln\left(\frac{p}{1-p}\right) = \hat{\beta}_0$. Applying antilog, this becomes $\left(\frac{p}{1-p}\right) = e^{\hat{\beta}_0}$. We know the value of *e* is approximately 2.71. Given that $\hat{\beta}_0 = 1$, this means the odds in favour of happening the event is 2.71, which is greater than 1. In the same way, if we consider the value of $\hat{\beta}_0 = -1$, the odds is:

$$\left(\frac{p}{1-p}\right) = e^{\hat{\beta}_0} = (2.71)^{-1} = 0.369$$

which is less than 1. Again, how do we interpret the odds if $\hat{\beta}_0$ is zero? The odds of owning a car becomes 1 if the value of $\hat{\beta}_0 = 0$! We need to explain this. From (5.14), if income x_i is zero, then $\hat{L}$, the estimated L is written as:

$\hat{L} = \ln\left(\frac{p}{1-p}\right) = \hat{\beta}_0$. This can be written as: $\left(\frac{p}{1-p}\right) = e^{\hat{\beta}_0}$. The left hand side represents the odds ratio in favour of owning a car. The right hand expression is known since the value of *e* is 2.71 and $\hat{\beta}_0$ is known. If $\hat{\beta}_0 = 0$, the odds ratio in favour of owning a car is 1. This means $p = (1-p)$ i.e. the probability of owning a car is equal to the probability of not owning a car! This means, at a certain level of income, one may be indecisive or indifferent. Now, let us proceed to interpret the estimated coefficient of other parameter $\hat{\beta}_1$. If we differentiate (5.16) with respect to x_i, we obtain: $\frac{\partial \hat{L}}{\partial x_i} = \hat{\beta}_1$. This gives us the incremental increase in the log of odds in favour of owning a car due to increase in one additional unit of income (x_i). Like in the previous case, if we know the exact estimated value of $\hat{\beta}_1$, we can calculate the odds 'in favour of owning a car'. Let us take the estimated value of $\hat{\beta}_1 = 0.8$. The odds in favour of owning a car will be 2.22 (=$e^{0.8}$), which is greater than 1. What happens to the odds in favour of owning a car if $\hat{\beta}_1$ becomes –0.8? The odds in favour of owning a car will be 0.45 which is less than 1. Can we estimate the marginal effect given the exact estimated value of $\hat{\beta}_1$? Marginal effect means incremental increase in p due to one unit increase in explanatory variable (x). Mathematically, we can write: Marginal Effect = $\frac{\partial p_i}{\partial x_i}$. Differentiating

Equation (5.16) with respect to x_i and using the fact that $\left[\hat{\beta}_0 + \hat{\beta}_1 x_i = \ln\left(\frac{p}{1-p_i}\right)\right]$ which follows from the Equation (5.13), we get:

$$\frac{\partial \hat{L}}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\hat{\beta}_0 + \hat{\beta}_1 x_i = \ln\left(\frac{p}{1-p_i}\right)\right] = \hat{\beta}_1$$

$$\Rightarrow \frac{1}{\frac{p_i}{1-p_i}} \cdot \frac{\partial}{\partial x_i}\left[\frac{p_i}{1-p_i}\right] = \hat{\beta}_1$$

$$\Rightarrow \frac{1-p_i}{p_i}\left[\frac{(1-p_i)\cdot\frac{\partial p_i}{\partial x_i} - p_i \cdot\left(-\cdot\frac{\partial p_i}{\partial x_i}\right)}{(1-p_i)^2}\right] = \hat{\beta}_1$$

$$\Rightarrow \frac{\frac{\partial p_i}{\partial x_i}}{p_i(1-p_i)} = \hat{\beta}_1$$

$$\Rightarrow \frac{\partial p_i}{\partial x_i} = \hat{\beta}_1 . p_i(1-p_i) \qquad (5.19)$$

The above expression stands for 'marginal effect'. It basically captures the incremental change in the probability of owning a car ($y = 1$) due to one unit increase in x_i. But the major problem arises about the estimation of p_i. Theoretically, p_i should be n values (i = 1, 2, 3……n) but it is unrealistic to estimate n number of marginal effects. There should be only one p. Note that 'regression' basically captures average mathematical relationship between two variables as is our case here. Therefore, we have to estimate that value of p which corresponds to the 'mean predicted probability'. The unique value of p can be solved from the estimated Logit model as follows. From 5.13, we can write: $\hat{L} = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Replacing x_i by its mean value, $\bar{x} = \frac{\sum x}{n}$, the new estimated Logit becomes: $\hat{L} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$. Now, the right hand side is known. Let us set $\hat{\beta}_0 + \hat{\beta}_1 \bar{x} = C$. We then have:

$$\hat{L} = C \Rightarrow \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = C$$

$$\Rightarrow \frac{\hat{p}}{1-\hat{p}} = e^C = k$$

$$\Rightarrow \frac{\hat{p}}{1-\hat{p}} = k$$

$$\Rightarrow \hat{p} = \frac{k}{1+k} (\because \hat{p} = k(1-\hat{p}); \hat{p} = k - k\hat{p}; (1+k)\hat{p} = k; \hat{p} = k/1+k)$$

$$\Rightarrow \hat{p} = p^*$$

Thus, assuming, $e^C = k$, we obtained $p^* = \frac{k}{1+k}$. This is the value of 'mean predicted probability'. By doing this, instead of multiple values of p, we have estimated a single $p = p^*$. Thus, from Equation (5.19), the 'marginal effect' can be obtained by the following equation:

$$\text{Marginal Effect} = \frac{\partial p}{\partial x} = \hat{\beta}_1 . p^* (1 - p^*) \qquad (5.20)$$

Therefore, in order to estimate the 'marginal effect' we need two information: (i) estimated Logit model and (ii) the mean value of the regressor. A numerical example will be helpful to grasp the problem. Consider the example of ownership of car of some 25 individuals whose estimated Logit model is given by: $\hat{L} = \hat{\beta}_0 + \hat{\beta}_1 x_i = -0.75 + 0.01 . x_i$. For simplicity, we further assume that the mean of $x = \bar{x} = 100$. Then the Logit at the mean value is: $\hat{L} = -0.75 + 0.01 \times 100 = 0.25$. We know that $\hat{L} = \ln\left(\frac{p^*}{1-p^*}\right)$, where p^* is the 'mean predicted probability' evaluated at mean value of x. Thus, $\left(\frac{p^*}{1-p^*}\right) = 2.71^{0.25} = 1.28$. This yields $p^* = 0.56$ (since 1.28/(1 + 1.28) = 1.28/2.28 = 0.56). Now, using Equation (5.20), we can estimate the 'marginal effect' as:

Marginal Effect $= \frac{\partial p}{\partial x} = \hat{\beta}_1 . p^* (1 - p^*) = 0.01 \times 0.56 \times 0.44 = 0.024$. This means, the probability of owning a car on an average will increase by 2.4 percent due to one unit increase in the level of income. To illustrate further, let us now consider a practical example as in Table 5.3.

In the Table, if an individual owns a car, then it is denoted as $y = 1$, if he does not own, $y = 0$. The income is measured in Rupees (thousand per month). We run the Logit model and obtain the following result as in Table 5.4.

Table 5.3: Illustrative Data for Y and X

(X is in Rs. '000)

Individual	Y	X	Individual	Y	X
1	0	30	14	1	70
2	1	55	15	1	65
3	1	48	16	0	38
4	0	22	17	0	40
5	0	24	18	1	45
6	1	50	19	0	38

7	1	60	20	1	70
8	1	55	21	0	56
9	0	40	22	1	48
10	0	38	23	0	40
11	1	60	24	1	75
12	0	30	25	0	38
13	0	45			

Table 5.4: Results of Logistic Regression

Logistic regression Number of obs = 25

LR $chi^2(1) = 21.56$ Prob > $chi^2 = 0.0000$

Log likelihood = -6.5262 Pseudo $R^2 = 0.622$

Dep. Var.	Logit Coef.	Std. Err.	Z	P>|z|
Income(x)	0.293	0.1218	2.41	0.016
Constant	–13.76	5.6374	–2.44	0.015

The estimated Logit model is:

$$\hat{L} = -13.76 + 0.293x_i \qquad (5.21)$$

We can interpret the results of Equation (5.21) as follows. If income is zero, the value of log of odds ratio is '–13.76'. This means the odds of owing a car is almost zero (since it is 1.05×10^{-06}). The coefficient of x_i (viz. 0.293) is the log of odds of owning a car. It is positive. But we must estimate odds in favour of owning a car. The odds ratio in favour of owning a car is 1.33. This can be verified as follows:

The estimated Logit model is: $\hat{L} = -13.76 + 0.293x_i$.

Here, $\hat{L} = \ln\left(\frac{p}{1-p}\right) = -13.76 + 0.293x_i$. Differentiating this with respect to x_i, we get:

$$\frac{\partial \hat{L}}{\partial x_i} = 0.293$$

$$\Rightarrow \frac{p}{1-p} = e^{0.293} \Rightarrow \frac{p}{1-p} = (2.71)^{0.293}$$

$$\Rightarrow \frac{p}{1-p} = 1.33$$

This is higher than 1. Therefore, the odds in favour of owning a car increases as income increases. To calculate the marginal effect, we first estimate the mean of x which is $\bar{x} = 47.2$. By using this in Equation (5.21), we get:

$$\hat{L} = -13.76 + 0.293 \times 47.2 = 0.069$$

$$\Rightarrow \frac{p}{1-p} = e^{0.069}$$

$$\Rightarrow \frac{p}{1-p} = 2.71^{0.069}$$

$$\Rightarrow \frac{p}{1-p} = 1.071$$

$$\Rightarrow p = 0.517$$

Therefore, the mean predicted probability is $p = p^* = 0.517$. Now, using this estimate of 'p' in Equation (5.20), we can estimate the 'marginal effect' as:

$\frac{\partial p}{\partial x} = \hat{\beta}_1 . p^*(1-p^*) = 0.293 \times 0.517 \times 0.483 = 0.073$. This means that one unit increase in income (x) increases the probability of owning a car by 7.3 percent. The question now is: are the two coefficients statistically significant? The answer is 'yes'. The z values are high and both the parameters are found to be significant at 1 percent level (see the last two columns of Table 5.4). We have already discussed about R^2, as a measure of goodness of fit' in case of LPM. The R^2 is assumed to be least important in case of categorical dependent variable. There are various concepts of R^2 which are developed like: Pseudo R^2, McFadden R^2 and count R^2 . The count R^2 is the simplest form of measuring goodness of fit. It is defined as: $Count - R^2 = \frac{CP}{N}$, where CP = number of correct predictions and N = total number of observations.

5.4.4 Application of Logit Model for Grouped Data

Let us now consider the data on several individuals who form groups according to their income level and the number of individuals owning a car at each income level. Corresponding to each income level x_i, there are N_i individuals with n_i among them owning a car. Therefore, we can compute the probability of owning a car as:

$$\hat{p}_i = \frac{n_i}{N_i} \qquad (5.22)$$

This is the relative frequency. Hence, we can use it as an estimate of the true p_i corresponding to each x_i. If N_i is fairly large, $\hat{p}_i$ will be a reasonably good estimate of p_i. Using the estimated p_i, we can obtain the estimated Logit as:

$$\hat{L}_i = \ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad (5.23)$$

Equation (5.23) is expected to be a fairly good estimate of the 'true Logit' if the number of observations N_i at each x_i is reasonably large.

5.5 THE PROBIT MODEL

The Logit model uses the cumulative logistic function as shown in Equation (5.13). But this is not the only 'cumulative distribution/density function' (CDF) that one can use. In some applications, the normal CDF is used for 'limited dependent variable models'. We know that the probability density function (PDF) and the cumulative density function (CDF) of normal distribution are respectively given by:

$$PDF = f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-0.5\left(\frac{X-\mu}{\sigma}\right)^2} \quad CDF = F(x) = \int_{-\infty}^{X} \frac{1}{\sigma\sqrt{2\pi}}.e^{-0.5\left(\frac{X-\mu}{\sigma}\right)^2}$$

The econometric model based on the 'dummy dependent variable' that emerges from the 'normal CDF' is popularly known as the Probit model. We can explain the Probit model using the car ownership example of individuals. Owning a car gives some satisfaction, which is an unobserved or latent variable. We assume that the decision of the i^{th} individual to own a car depends on an unobserved utility level, U_i, which is determined by the level of income (x) that the individual possess. We further assume that U_i is positively and linearly related to x_i like:

$$U_i = \beta_0 + \beta_1 x_i \qquad (5.24)$$

An important question now is: 'how is the unobserved (latent) utility index (U_i) related to the actual decision of owning a car'? We further assume that there exists a cut-off or benchmark level of utility index, U_i^*, such that if the actual utility index (U_i) exceeds the cut-off level of utility index U_i^*, the individual will own a car. This means $y_i = 1$, if $U_i > U_i^*$, 0 otherwise. Given the assumption of normality, the probability that U_i^* is less than or equal to U_i can be computed from the standardised normal CDF as:

$$p_i = p(y = 1/x) = p(U_i^* \le U_i) = p(z_i \le \beta_0 + \beta_1 x_i) = F(\beta_0 + \beta_1 x_i) \qquad (5.25)$$

where $p(y \mid x)$ implies the probability of having a car given the income level, z_i is the standard normal variate (with zero mean and unit variance) and F is the standard normal CDF, which can be written as:

$$F(U_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0+\beta_1 x_i} e^{-\frac{z^2}{2}} dz \qquad (5.26)$$

Since, p represents the probability that an event will occur (which is here the probability of owning a car), it is measured by the area of the standard normal curve from $-\infty$ to U_i. Now, from Equation (5.25), we can write:

$$U_i = F^{-1}(U_i) = F^{-1}(P_i) = \beta_0 + \beta_1 x_i \qquad (5.27)$$

where F^{-1} is the inverse of normal CDF. Now, the data may be at the individual or group level. The same process as was discussed in the Logit model, needs to be applied here. The only difference is the CDF used. The parameters of β_0 and β_1 are to be estimated from the 'maximum likelihood method'. The 'marginal effect' of Probit model is obtained by differentiating the Equation (5.25) with respect to x_i i.e.:

$$\frac{\partial p_i}{\partial x_i} = F^{/}.\beta_1 \qquad (5.28)$$

where F$^{/}$ is the 'probability density function' of the standard normal variate $z_i = \beta_0 + \beta_1 x_i$. F$^{/}$is to be evaluated at z_i.

5.5.1 Empirical Example of Probit Model

We again consider the earlier example of car ownership as in Table 5.3. The estimated Probit model is reported in Table 5.5.

Table 5.5: Estimated Probit Model

Probit regression Number of observations=25

LR chi^2(1) =21.53 Prob> chi^2 =0.0000

Log likelihood = -6.5456442 Pseudo R^2 =0.6218

Dependent Var	Coefficient	Std. Err.	z	P>z
Income	0.1620	0.05848	2.77	0.006
Constant	-7.6841	2.782351	-2.76	0.006

The Probit coefficient is 0.1620. This means if income (x) increases, the actual utility index (U_i) exceeds the threshold level of Utility Index (U^*_i) with the probability of owning a car increasing by 16.2%. The intercept term is negative. This means that if income is zero, the utility index is also negative leading to 'zero' probability of owning a car. Now, the marginal effect can be estimated for any value of the level of income. Earlier we saw that there exists only one marginal effect of Logit model which is evaluated at mean value. But, in case of Probit model, one can estimate the 'marginal effect' for any positive value of x. This is done by estimating the probability density of the standardised normal curve.

5.5.2 Estimation of Marginal Effect in Probit Model

Suppose we want to evaluate the marginal effect, $\frac{\partial p_i}{\partial x_i}$ at the mean value of x i.e. $\bar{x}$ = 47.2. Using, 5.27, we can compute the marginal effect of the Probit model as:

$$\frac{\partial p_i}{\partial x_i} = F'.\beta_1 = f(-7.6841 + 0.1620 \times 47.2) \times 0.1620 = f(-0.03) \times 0.1620$$
$$= 0.4750 \times 0.162 = 0.076$$

In the above we have taken the value 0.475 because the value of $f(-0.03)$ corresponding to $P(-1.96 \leq Z \leq 0)$ is 0.475. The results of the 'marginal effect' means that for one unit increase in income (x) beyond the mean income ($\bar{x}$), the probability of owning a car will increase by 7.6 percent. This is more or less same as the Logit result (in case of Logit, we estimated the 'marginal effect' as 0.073). One can estimate the 'marginal effect' for any value of x in Probit model but we cannot do it in the Logit model. The Logit and Probit models are more or less similar. The sign of the coefficients are same. The only difference is related to numerical values. Empirically, one can establish the relationship between the two coefficients of Logit and Probit models by multiplying the Probit coefficient by 1.81. This gives us an approximate value of the Logit coefficient. This can be examined in our above car ownership example. The estimated Logit coefficient is 0.293 while the estimated Probit coefficient is 0.1620. You can verify that $0.293 \approx 1.81 \times 0.1620$. Therefore, in general, we can write this as:

$$\beta_1 \textbf{ of Logit} \approx \textbf{1.81} \times \beta_1 \textbf{ of Probit} \qquad (5.29)$$

There also exists a relationship between the coefficients of Logit model and LPM as follows:

$$\beta_1 \textbf{ of LPM} \approx \textbf{0.25} \times \beta_1 \textbf{ of Logit} \text{ [except for intercept]} \qquad (5.30)$$

$$\beta_1 \textbf{ of LPM} \approx \textbf{0.25} \times \beta_1 \textbf{ of Logit} + \textbf{0.5} \text{ [including intercept]} \qquad (5.31)$$

You can verify the above relationship by estimating the LPM which is not done here in this unit.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) Define the Logit model and state its properties.

.....

.....

.....

.....

.....

.....

.....

2) In the Logit model, to estimate the 'marginal effect' which two information are needed?

.....

.....

.....

.....

.....

3) What is a fundamental difference between the Logit model and the Probit model?

.....

.....

.....

.....

.....

4) Empirically, what is the relationship that exists between: (i) the Logit coefficients and the Probit coefficients and (ii) the Logit coefficients and the LPM coefficients?

.....

.....

.....

.....

.....

5) In Tables 5.4 and 5.5 we have used values of Z. But values of chi-square is also given. Write a comment on this.

.....

.....

.....

.....

.....

5.6 LET US SUM UP

Qualitative response regression model refers to models in which the response for the endogenous variable is not quantitative on an interval scale. There are three alternative models viz. the LPM, the Logit and the Probit to deal with

dummy dependent variable situations. The unit has explained these three models for situations of 'binary dependent variable models'. The LPM is the simplest one but it has serious limitations. The most important model is Logistic regression or Logit model. The limitations of LPM are overcome both in the Logit and the Probit models. There is no significant difference between Logit and Probit models. Their regression coefficients are related. This means, if we get a regression coefficient of Logit model, we can indirectly estimate the Probit regression coefficients. The regression coefficients of LPM and Logit model are also related. The unit has indicated these relationships.

5.7 KEY WORDS

Linear Probability Model : It is stated as in Equations (5.4) and (5.5) where dependent variable takes two values: 0 and 1. In this, we have to deal with situations on non-normality of disturbance terms, heteroscedastic variances of disturbance terms and non-fulfilment of conditional probability values being within the limits of '0' and '1'.

Logit Model : This is a model as stated in Equation (5.10). This is non linear both in x's and β. Hence, the OLS method of estimation does not work here. We can estimate the parameters by using the ML method as outlined in Equation (5.16). The test results of this model are based on standard normal Z values.

Probit Model : The form of the Probit model depends on the 'cumulative distribution function' (CDF). Usually, we work with the 'normal variate'. Hence, this is defined as: the econometric model based on 'dummy dependent variable' as emerges from the normal distribution's CDF. As in the Logit model, the test results of this model are also based on the Z values.

5.8 SUGGESTED BOOKS FOR FURTHER READING

1) Gujarati D., Porter D. C. and Gunasekhar S (2018). Basic Econometrics, 5th Edition, McGraw Hill Education (India) Pvt. Ltd.

2) G.S. Maddala,. (1992). Introduction to Econometrics, Second Edition, Macmillan, New York.

5.9 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) The word 'binary' restricts the dependent variable to be dichotomous. But among the independent variables we can have a normal variable like income or a dummy variable or a mixture of both. This means the regressors can be either quantitative or qualitative.

2) No. We encounter three specific type of problems viz. variances of disturbances are not homoscedastic, the probability values of conditional expectation is not between '0' and '1' and the error terms are not normally distributed, a property that we do not need for OLS but we need for drawing 'statistical inferences'.

3) We can apply the method of weighted least squares (WLS) by taking $\hat{w}_i = \hat{y}_i(1-\hat{y}_i)$ as the estimate of w_i.

4) This is because the conventional R^2 is likely to be much lower than 1 for such models.

Check Your Progress 2

1) A Logit model is stated as in Equation (5.10). We cannot straightaway apply the OLS method because p_i is nonlinear not only in x but also in the β's. But by considering the ratio as in Equation (5.12) and then taking logarithm on both sides of this equation, we get an equation like (5.13) which is linear in x_i. For estimation purposes, we have to work with L_i as stated in Equation (5.15). If data is available at individual level, we cannot estimate its parameters by the OLS method. For this reason, we take the help of the ML method to solve for its parameters.

2) (i) estimated Logit model and (ii) the mean value of the regressor.

3) One can estimate the 'marginal effect' for any value of x in Probit model but we cannot do it in the Logit model.

4) The relationships are approximations as can be obtained by Equations (5.29) to (5.31).

5) The likelihood-ratio (LR) test assesses the 'goodness of fit' of two competing statistical models based on the ratio of their likelihoods. Specifically, one is found by maximisation (over the entire parameter space) while another is found after imposing some constraint. If the constraint (i.e. the null hypothesis) is supported by the observed data, the two likelihoods should not differ by more than the sampling error. Thus, the likelihood-ratio test tests whether this ratio is significantly different from one, or equivalently, whether its natural logarithm is significantly different from zero. If the distribution of the likelihood ratio corresponding to a particular null (and the alternative hypothesis) can be

explicitly determined, then it can directly be used to form decision regions (i.e. to sustain or reject the null hypothesis). In most cases, however, the exact distribution of the likelihood ratio corresponding to specific hypotheses is very difficult to determine. However, following Wilk's theorem, we can argue that as n (sample size) tends to infinity, the test statistic asymptotically will follow the Chi Square distribution.

UNIT 6 SIMULTANEOUS EQUATION MODELS-I*

Structure

6.0 OBJECTIVES

After reading this unit, you will be able to:

- elucidate the meaning of the term 'simultaneous equation bias';
- show that the consequence of applying the OLS method of estimation in a situation of 'simultaneous equation bias' is to have inefficient estimates;
- analyse the case of 'simultaneous equation bias' in the 'Keynesian Model of Income Determination';
- state the meaning of the term 'identification';
- demonstrate how in the two situations of 'under and over identification' unique solution to regression models do not exist;
- prove that 'exact identification' is a sufficient condition under which the system of equations is uniquely determined;

* Prof. Sushil Haldar, Jadavpur University.

- explain the procedure of applying the 'order condition' of identification with suitable illustrations;
- describe the steps involved in applying the 'rank condition of identification' with an illustration; and
- verify that in case of the 'Keynesian Model of Income Determination' the 'rank and order conditions' are satisfied for the 'investment function' but not for the 'consumption function'.

6.1 INTRODUCTION

In two variable regression models, we assume one variable as endogenous (or dependent) and the other one as independent or explanatory. The explanatory variable is assumed to be strictly exogenous in nature. If this property holds, we can easily estimate the unknown parameters (viz. regression coefficients) by applying the method of least squares. If this can be attempted, the 'cause and effect relationship' is regarded as well established in the uni-directional sense i.e. it is considered to run from X to Y. However, in reality, we confront many problems that violate the above property of exogeneity of the independent variable. For instance, let us consider the relationship between health status and poverty. Poverty sometimes induces low level health status like higher morbidity or mortality (or disability). The converse is also true i.e. poor health indirectly leads to low level of income earning because of higher incidence of physical and mental disability. This also indirectly leads to poverty. This is true both at micro (at the individual level) as well as at the macro level. Now, if we define poverty as Y and health status as X, then we can write $Y = F(X)$ and $X = f(Y)$. Under such circumstances, we cannot apply the standard OLS method to determine the parameters. If applied, in disregard to the two way relationship, it gives biased and inconsistent estimate. This is because of the violation of the condition $Cov(X_i, u_i) = 0$. This means both X and Y are jointly determined endogenous variables. Hence, one way direction of causality is meaningless since there is clearly both way causality. Such situations with simultaneous impact from both sides, need to be treated as such i.e. for simultaneous impacts. In other words, in case of both way causality [viz. $Y = f(X)$ and $X = f(Y)$], we should not use the 'single equation approach'. We must instead use a multi-equation model, called the simultaneous equation model or the simultaneous systems approach. Such models include separate equations in which Y and X would appear as endogenous variables. Such a system of equations which describes the joint dependence of variables are called as 'system of simultaneous equations'. Given the nature of economic phenomena as cited above (viz. relationship between poverty and health status), it is almost certain that most of the economic situations would belong to a wider system of simultaneous equations. In this Unit, we first verify the consequences of the violation of the assumption of $Cov(X_i, u_i) = 0$ to note the implication in terms of what is called as the 'simultaneous equation bias'. The bias arises from the

application of classical least square to an equation belonging to a system of simultaneous relations. This is also alternatively termed as the 'presence of simultaneity' in regressions.

6.2 CONSEQUENCE OF SIMULTANEITY

In this section, we shall verify two aspects: (i) as a consequence of simultaneity, the OLS estimators are inefficient even in the case of a simple model of simultaneous regression equations and (ii) taking the particular case of 'Keynesian Model of Income Determination' we will verify, how due to simultaneity, the OLS estimates result in inconsistent estimates.

6.2.1 Simple Regression Models

As outlined above, when there is a joint dependence between Y and X, their relationship requires the treatment of a 'system of simultaneous equations'. This means, the explanatory variables appear as dependent variable in other equations of the system. Thus, for any particular equation the dependent variable is not independent of the explanatory variable(s). The consequence of the violation of the OLS's assumption E(X, u) = 0 is that the estimates are both biased and inconsistent. To verify this, let us consider a simple model of simultaneous equations as follows:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + u \\ X &= \alpha_0 + \alpha_1 Y + \alpha_2 Z + \upsilon \end{aligned} \qquad (6.1)$$

where $E(u) = E(v) = 0$; $E(u_i, u_j) = E(v_i, v_j) = 0$, $E(u^2) = \sigma_u^2$ and $E(v^2) = \sigma_v^2$. The model described by (6.1) is a 'simultaneous equation system' in which there are two endogenous variables (Y and X) and one exogenous variable Z. Now, substituting for Y from the first equation into the 2nd equation, we get:

$$\begin{aligned} & X = \alpha_0 + \alpha_1 (\beta_0 + \beta_1 X + u) + \alpha_2 Z + \upsilon \\ & \Rightarrow X = \frac{\alpha_0 + \beta_0 \alpha_1}{1 - \beta_1 \alpha_1} + \frac{\alpha_2}{1 - \beta_1 \alpha_1} . Z + \left(\frac{\alpha_1 u + \upsilon}{1 - \beta_1 \alpha_1} \right) \end{aligned} \qquad (6.2)$$

Note that from here afterwards, you need to fill-in some of the intermediate steps by working out separately yourself. These steps merely relate to transposing and simplifying to arrive at the equation in the next step. Equation (6.2) so arrived at shows that X is related to the disturbance terms u and v. Hence, X is not truly an exogenous variable as in the first equation of the system. It can also be further verified that the covariance of X and u is not zero. For this, let us write the covariance term by definition as: $Cov(X,u) = E\{[X - E(X)].[u - E(u)]\}$.

Now, since $E(u) = 0$, we have:

$$Cov(X,u) = E\{u[X - E(X)] \quad (6.3)$$

Thus, given that $X = \frac{\alpha_0 + \beta_0\alpha_1}{1-\beta_1\alpha_1} + \frac{\alpha_2}{1-\beta_1\alpha_1}.Z + \left(\frac{\alpha_1 u + \upsilon}{1-\beta_1\alpha_1}\right)$ and Z is exogenously determined, we have: $E(X) = \frac{\alpha_0 + \beta_0\alpha_1}{1-\beta_1\alpha_1} + \frac{\alpha_2}{1-\beta_1\alpha_1}.Z.$

Therefore, using (6.2) in (6.3), Cov (*X*,*u*) can be written as:

$$Cov(X,u) = E\left[\frac{u}{1-\alpha_1\beta_1}\{\alpha_0 + \alpha_1\beta_0 + \alpha_2 Z + \alpha_1 u + \upsilon - (\alpha_0 + \alpha_1\beta_0 + \alpha_2 Z)\}\right]$$

$$\Rightarrow E\left[\frac{u}{1-\alpha_1\beta_1}(\alpha_1 u + \upsilon)\right] \Rightarrow \frac{1}{1-\alpha_1\beta_1}E(\alpha_1 u^2 + u\upsilon) \Rightarrow \frac{\alpha_1}{1-\alpha_1\beta_1}.E(u^2) \neq 0$$

[since $E(u\upsilon) = 0$]. Therefore, if we apply the method of least squares to (6.1), the estimates of the coefficients will be 'biased and inconsistent'. Thus, in case of 'simultaneous equation bias', the unidirectional 'cause and effect' relationship will not hold even in the case of a set of 'simple linear regression'. To reiterate, this is due to some of the variables in one equation, acting as 'cause variables', also acting as 'effect variables' in the other equations of the system. If this is the situation in the case of simple linear regression model, we can imagine the nature of complicated 'systems of equations' we would have to deal with in a multiple regression model. Indeed, in real life situations, we do not have unidirectional relationships. We rather encounter two way relationships appearing as a system of simultaneous equations. That is, we have more than one linear regression equations with one variable acting as exogenous variable in one equation, but as endogenous variables in the other equations of the system. The consequence of such a situation, if we proceed with the method of OLS to estimate the parameter, is to have 'simultaneity bias'. Now, before we proceed further, let us see how simultaneity affects in the case of the Keynesian Model of Income Determination.

6.2.2 Keynesian Model of Income Determination

Let us now analyse for 'simultaneity bias' in a simple 'Keynesian model of income determination' defined as below:

$$C_t = \beta_0 + \beta_1 Y_t + u_t \quad 0 < \beta_1 < 1 \quad (6.4)$$

$$Y_t = C_t + I_t (= S_t) \quad (6.5)$$

where, C = consumption expenditure, Y = income, I = investment (assumed exogenous), S = savings, t = time, u = stochastic disturbance term, β_0 and β_1 are the parameters. In particular, the parameter β_1 is the marginal propensity to consume (MPC). From economic theory, we know that β_1 is expected to lie between 0 and 1. Equation (6.4) is the (stochastic) consumption function with a disturbance term. Equation (6.5) is the national income identity, signifying that the 'total income' is = total 'consumption expenditure' + total

'investment expenditure'. It is presumed that the total 'investment expenditure' and the total 'savings' are equal. In (6.4), consumption is determined by income whereas in (6.5) income depends on consumption. Therefore, both *C* and *Y* are endogenous i.e. simultaneously they are mutually dependent on one another. They are also correlated with the error term. Because of this, if we apply the OLS procedure, the estimation will be inconsistent and inefficient. Let us actually verify how this is so given that: $E(u_t) = 0,\ E(u_t^2) = \sigma^2,\ E(u_t, u_{t+j}) = 0$ and $Cov(I_t, u_t) = 0.$ We wish to verify whether Y_t and u_t are not correlated in (6.4) and $\hat{\beta}_1$ is a consistent estimator of β_1. To verify this, we substitute (6.4) into (6.5) to obtain:

$$Y_t = \beta_0 + \beta_1 Y_t + u_t + I_t \Rightarrow (1 - \beta_1)\, Y_t = \beta_0 + u_t + I_t \text{ or}$$

$$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t$$

Now, since $E(u_t) = 0,\ E[Y_t] = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t.$ Thus, $Y_t - E[Y_t] = \frac{u_t}{1-\beta_1}.$

Now, covariance between Y_t and u_t is:

$$Cov(Y_t, u_t) = E[Y_t - E(Y_t)][u_t - E(u_t)]$$

$$\Rightarrow Cov(Y_t, u_t) = E\left[\frac{u_t}{1-\beta_1}\right][u_t] \Rightarrow Cov(Y_t, u_t) = \frac{E(u_t^2)}{1-\beta_1}$$

$$\Rightarrow Cov(Y_t, u_t) = \frac{\sigma^2}{1-\beta_1} \qquad (6.6)$$

Hence, Y_t and U_t are correlated. This violates the assumption of the classical linear regression model that the disturbance term is independent (i.e. uncorrelated) with the explanatory variable. This means the OLS estimators, if applied, will not be consistent. Now, to show whether the OLS estimator $\hat{\beta}_1$ is a consistent estimators of β_1 or not, because of correlation between Y_t and u_t, we proceed as follows:

The estimated regression coefficient can be written as:

$$\hat{\beta}_1 = \frac{Cov(C,Y)}{V(Y)} = \frac{\sum (C_t - \bar{C})(Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2} \Rightarrow \hat{\beta}_1$$

$$= \frac{\sum \left[\beta_1 (Y_t - \bar{Y}) + u_i\right](Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2} = \beta_1 + \frac{\sum y_t u_t}{\sum y_t^2}$$

[since $C_t - \overline{C} = \beta_1(Y_t - \overline{Y}) + u_i$). and by defining, $Y_t - \overline{Y} = y_t$]. We therefore get the mathematical expectation of $\hat{\beta}_1$ as:

$$E(\hat{\beta}_1) = \beta_1 + E\left[\frac{\sum y_t u_t}{\sum y_t^2}\right] \qquad (6.7)$$

From (6.6) we have that 'the covariance between y_t and u_t is not equal to zero'. Hence, $E(\hat{\beta}_1) \neq \beta_1$. Therefore, the OLS estimates are inconsistent.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) Define the term 'simultaneity'. What are its consequences?

……

……

……

……

……

2) What is meant by 'simultaneity bias'?

……

……

……

……

……

3) What is the effect of 'simultaneity' in the classical Keynesian model of income determination?

……

……

……

……

……

6.3 IDENTIFICATION PROBLEM

A model is said to be identified if it has a unique statistical form enabling unique estimates of its parameters. Thus, identification is a problem of 'model formulation' and not 'estimation of its parameters'. By the term identification problem we mean "whether the empirical (numerical) estimates of the parameters of a 'structural equation' can be obtained from the estimation of the parameters of its 'reduced form' equations"? If this cannot

be done, then we say that there is a problem of identification i.e. the particular equation is unidentified or under-identified. Thus, if an equation is identified it may either be 'exactly identified' or 'over-identified'. It is said to be 'exactly identified' if unique numerical estimates of its 'structural parameters' can be obtained. It is said to be 'over identified' if more than one numerical value can be obtained for some of the parameters of the structural equations. The identification problem arises because, different sets of structural coefficients may be compatible with the same set of data. This can be explained as follows. Let us consider the standard demand and supply curves where *p* and *q* are the jointly dependent variables.

$$q_t^D = \alpha_0 + \alpha_1 p_t + u_{1t} \tag{6.8}$$

$$q_t^S = \beta_0 + \beta_1 p_t + u_{2t} \tag{6.9}$$

$$D = S \tag{6.10}$$

where $\alpha_1 < 0$, $\beta_1 > 0$, q_t^D and q_t^S are the quantity demanded and supplied respectively and $D = S$ is the market clearing condition. Any change in u_{1t} and u_{2t} affects the demand and supply curves respectively. This is in the sense that given the supply function, the demand function may shift upward or downward disturbing the 'equilibrium price and quantity combinations'. Likewise, we can have multiple combinations of price and quantity due to the shift in the supply function given the demand schedule.

We also know that given the demand function, the equilibrium price and quantity is subject to change due to changes in the 'disturbance term' of the supply function. Therefore, simultaneous dependence between p and q and u_{1t} and p_t in Equation (6.8) and u_{2t} and p_t in equation (6.9) leads us to the problem of simultaneity. This is due to the violation of 'independence' assumption. Thus, a regression of *q* on *p* as in Equation (6.8) would violate an important assumption of the CLRM, namely, there is 'no correlation between the explanatory variables and the disturbance term'. The same problem will arise if we run the regression of *q* on *p* as in Equation (6.9). In view of this, the 'identification problem' in the simultaneous equation systems needs to be discussed under three heads viz. (i) under-identification, (ii) over-identification and (iii) just or exact identification. Let us first discuss the case of 'under-identification'.

6.3.1 Under Identification

Consider the equation of 'market clearing condition' obtained by equating Equations (6.8) and (6.9), along with the corresponding equilibrium price, as follows:

$$\alpha_0 + \alpha_1 p_t + u_{1t} = \beta_0 + \beta_1 p_t + u_{2t} \tag{6.11}$$

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} = \Pi_0 + \upsilon_t \qquad (6.12)$$

where, $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$ and $\upsilon_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$. Substituting the value of p_t [from 6.12) in (6.8) and simplifying, we get the equilibrium value of quantity as:

$$q_t = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} = \Pi_1 + \omega_t \qquad (6.13)$$

where, $\Pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$ and $\omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$.

The error terms υ_t in (6.12) and ω_t in (6.13) are a linear combination of the original error terms u_1 and u_2. Equations (6.12) and (6.13) are the 'reduced form equations'. Now, our demand-supply model contains four structural constraints $\alpha_0, \alpha_1, \beta_0$ and β_1 [from (6.11)]. But there is no unique way of estimating these four parameters. This is because we have four unknown and only two reduced form equations. In other words, our models [as in (6.8) and (6.9)] are under-specified or under-identified.

6.3.2 Over Identification

In order to discuss the case of over identification, we assume that the demand for goods is determined by 'price, income and assets' and the supply of goods is determined by the 'price of current period and previous period'. The following are the demand and supply functions:

$$q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \quad \alpha_1 < 0,\ \alpha_2 > 0,\ \alpha_3 > 0 \qquad (6.14)$$

$$q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + u_{2t} \quad \beta_1 > 0,\ \beta_2 > 0 \qquad (6.15)$$

The market clearing condition i.e. $D = S$ gives us:

$$\alpha_0 + \alpha_1 p_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + u_{2t}$$

Solving for p_t, the equilibrium price is:

$$p_t = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2}{\alpha_1 - \beta_1}\right).I_t + \left(-\frac{\alpha_3}{\alpha_1 - \beta_1}\right)R_t + \left(\frac{\beta_2}{\alpha_1 - \beta_1}\right)p_{t-1} + \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)$$

$$\Rightarrow p_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + \Pi_3 p_{t-1} + \upsilon_t \qquad (6.16)$$

where, $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\Pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}$, $\Pi_2 = -\frac{\alpha_3}{\alpha_1 - \beta_1}$, $\Pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1}$,

$\upsilon_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$

The equilibrium quantity of goods is determined by inserting the equilibrium price in either (6.14) or (6.15). Using (6.14) for this, we get:

$$q_t = \alpha_0 + \alpha_1\left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_1\alpha_2}{\alpha_1 - \beta_1}\right)I_t + \left(-\frac{\alpha_1\alpha_3}{\alpha_1 - \beta_1}\right)R_t + \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\right)p_{t-1} + \alpha_1\left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right) + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$$

$$\Rightarrow q_t = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right)I_t + \left(-\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1}\right)R_t + \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\right)p_{t-1} + \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$$

$$\Rightarrow q_t = \Pi_4 + \Pi_5 I_t + \Pi_6 R_t + \Pi_7 p_{t-1} + \omega_t \qquad (6.17)$$

where, $\Pi_4 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$, $\Pi_5 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}$, $\Pi_6 = -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1}$, $\Pi_7 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$, $\omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$.

Under this situation, we have seven coefficients (Π_1 to Π_7) to be estimated from eight equations. This means the number of equations exceeds the number of unknown parameters. Consequently, a unique solution is not possible, because, notice that, the parameter β_1 can be estimated in two ways: $\beta_1 = \frac{\Pi_6}{\Pi_2}$ and $\beta_1 = \frac{\Pi_5}{\Pi_1}$. This situation is the opposite of the under identification problem, where we had less information about the demand and supply function.

6.3.3 Exact Identification

Let us consider the earlier supply function but with a new demand function where we assume that the quantity demanded is determined by both 'price and income (*I*)'. The demand and supply functions are:

$$q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} \quad \alpha_1 < 0, \alpha_2 > 0 \qquad (6.18)$$

$$q_t^S = \beta_0 + \beta_1 p_t + u_{2t} \quad \beta_1 > 0 \qquad (6.19)$$

The market clearing condition gives: $\alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} = \beta_0 + \beta_1 p_t + u_{2t}$

Solving for p_t, we get:

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \left(-\frac{\alpha_2}{\alpha_1 - \beta_1}\right)I_t + \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} = \Pi_0 + \Pi_1 I_t + \upsilon_t \qquad (6.20)$$

where, $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \Pi_1 = \left(-\frac{\alpha_2}{\alpha_1 - \beta_1}\right)., \upsilon_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$.

We can now substitute the value of 'p_t' in either the demand (6.18) or supply (6.19) function. Using (6.18) for this, we get the equilibrium value of quantity as:

$$q_t = \alpha_0 + \alpha_1.\left[\left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2}{\alpha_1 - \beta_1}\right)I_t + \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)\right] + \alpha_2.I_t + u_{it}$$

$$\Rightarrow q_t = \alpha_0 + \frac{\alpha_1\beta_0 - \alpha_0\alpha_1}{\alpha_1 - \beta_1} - \frac{\alpha_1\alpha_2}{\alpha_1 - \beta_1}.I_t + \frac{\alpha_1(u_{2t} - u_{1t})}{\alpha_1 - \beta_1} + \alpha_2 I_t + u_{1t}$$

$$\Rightarrow q_t = \left[\alpha_0 + \frac{\alpha_1\beta_0 - \alpha_0\alpha_1}{\alpha_1 - \beta_1}\right] + \alpha_2 I_t\left(1 - \frac{\alpha_1}{\alpha_1 - \beta_1}\right) + \left[u_{1t} + \alpha_1\left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)\right]$$

$$\Rightarrow q_t = \left[\frac{\alpha_0\alpha_1 - \alpha_0\beta_1 + \alpha_1\beta_0 - \alpha_0\alpha_1}{\alpha_1 - \beta_1}\right] + \alpha_2 I_t\left(\frac{\alpha_1 - \beta_1 - \alpha_1}{\alpha_1 - \beta_1}\right) + \left[\frac{\alpha_1 u_{1t} - \beta_1 u_{1t} + \alpha_1 u_{2t} - \alpha_1 u_{it}}{\alpha_1 - \beta_1}\right]$$

$$\Rightarrow q_t = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right).I_t + \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right).$$

$$\Rightarrow q_t = \Pi_2 + \Pi_3.I_t + \omega_t \qquad (6.21)$$

where,

$$\Pi_2 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right), \Pi_3 = \left(-\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right), \omega_t = \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$$

Therefore, Equations (6.20) and (6.21) are both reduced form equations. We can apply OLS for estimation of the parameters. But we still have five unknown parameters α_0, α_1, α_2, β_0 and β_1 but four equations i.e. two structural and two reduced forms. Therefore, unique solution is not possible to obtain. However, the parameters of supply function can be estimated by considering the following:

$$\Pi_2 - \beta_1\Pi_0 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right) - \beta_1.\left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) = \frac{\alpha_1\beta_0 - \alpha_0\beta_1 - \beta_0\beta_1 + \alpha_0\beta_1}{\alpha_1 - \beta_1}$$

$$= \frac{\beta_0(\alpha_1 - \beta_1)}{\alpha_1 - \beta_1} = \beta_0.$$

$\frac{\Pi_3}{\Pi_1} = \frac{-\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}}{-\frac{\alpha_2}{\alpha_1 - \beta_1}} = \beta_1$. Hence, we have two equations, $\Pi_2 - \beta_1\Pi_0 = \beta_0$ and $\frac{\Pi_3}{\Pi_1} = \beta_1$. The two parameters β_0 and β_1 associated with the supply function can be estimated. But we cannot estimate the parameters associated with the demand function. In order to identify the demand function, we need to introduce some more information into the supply function. Let us introduce one additional information in the supply function as: 'assume that supply is

also determined by the past period's price but the demand function remains the same as earlier'.

$$q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} \quad \alpha_1 < 0, \quad \alpha_2 > 0 \tag{6.22}$$

$$q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + u_{2t} \quad \beta_1 > 0, \ \beta_2 > 0 \tag{6.23}$$

Now, applying the market clearing condition ($D = S$) we get:

$\alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + u_{2t}$. We can now estimate the price as follows:

$$\alpha_1 p_t - \beta_1 p_t = (\beta_0 - \alpha_0) + \beta_2 p_{t-1} - \alpha_2 I_t + (u_{2t} - u_{1t})$$

$$\Rightarrow p_t = \left[\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right] + \left[-\frac{\alpha_2}{\alpha_1 - \beta_1}\right].I_t + \left[\frac{\beta_2}{\alpha_1 - \beta_1}\right].p_{t-1} + \left[\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right]$$

We can write the above as:

$$p_t = \Pi_0 + \Pi_1 . I_t + \Pi_2 . p_{t-1} + \upsilon_t \tag{6.24}$$

where, $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\Pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}$, $\Pi_2 = \frac{\beta_2}{\alpha_1 - \beta_1}$, $\upsilon_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$.

Substituting the equilibrium price in (6.22) we get the equilibrium value of output as:

$$q_t = \alpha_0 + \alpha_1 \left[\left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2}{\alpha_1 - \beta_1}\right).I_t + \left(\frac{\beta_2}{\alpha_1 - \beta_1}\right).p_{t-1} + \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)\right] + \alpha_2 I_t + u_{1t}$$

$$\Rightarrow q_t = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right) + \left(-\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right).I_t + \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\right).p_{t-1} + \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$$

$$\Rightarrow q_t = \Pi_3 + \Pi_4 . I_t + \Pi_5 p_{t-1} + \omega_t \tag{6.25}$$

where $\Pi_3 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$, $\Pi_4 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}$, $\Pi_5 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$,

$\omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$

The demand and supply functions in Equations (6.22) and 6.23) now consist of six structural coefficients viz. $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ and β_2 and there are six 'reduced form coefficients' viz. $\Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4$ and Π_5 from (6.24) and (6.25). Thus, we have six unknowns and six equations, therefore, the system is uniquely determined.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) Define the term 'identification'.

……

……

……

……

……

2) What is meant by the 'problem of identification'?

……

……

……

……

……

3) Which are the two forms that an equation identified can take? What is the distinction between the two?

……

……

……

……

……

4) Why does the problem of 'identification' basically arise?

……

……

……

……

……

……

5) Write the intermediate steps to arrive at: (i) (6.12) from (6.11) and (ii) (6.13).

……

……

……

……

6.4 CONDITIONS FOR IDENTIFICATION

There are two conditions, namely, 'order condition' and 'rank condition' useful to determine the identification status of equations in a simultaneous

equation system. The order condition is a necessary condition but not sufficient. The rank condition is a sufficient condition.

6.4.1 Order Condition

The term order refers to the rank of a matrix i.e. the number of independent rows and columns present in a matrix. We use this fact in our identification problem here. As a rule or a condition, it is based on a 'counting of the variables included and excluded from a particular equation'. It is a necessary but not sufficient condition for the identification of an equation. The Order condition may be stated as follows. Let K = number of total variables (both endogenous and exogenous or pre-determined) in the model, M = number of variables (both endogenous and exogenous or pre-determined) included in a particular equation, G = total number of equations (= total number of endogenous variables). Now, the order condition requires that '$K - M \geq (G - 1)$'. To explain the order condition further, let us consider our earlier demand supply equations, as stated in (6.8) and (6.9), as:

$$q_t^D = \alpha_0 + \alpha_1 p_t + u_{1t} \quad q_t^S = \beta_0 + \beta_1 p_t + u_{2t} \qquad (*)$$

In this system of equations, we have two endogenous variables, p and q. There is no exogenous variable. Therefore, $K–M = 2 – 2 = 0$; $G – 1 = 2 – 1 = 1$. Therefore, the order condition is not satisfied. Hence, the system is under-identified. Let us now consider the following demand and supply functions:

$$q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} \qquad \alpha_1 < 0,\ \alpha_2 > 0 \qquad (6.26)$$

$$q_t^s = \beta_0 + \beta_1 p_t + u_{2t} \qquad \beta_1 > 0 \qquad (6.27)$$

Here, p and q are endogenous variables and I is exogenous. Applying the order condition to the demand equation we have: $K – M = 3–3 = 0$, $G – 1 = 2 – 1 = 1$. Therefore, $K - M \geq (G - 1)$ is not satisfied. Thus, the demand function is unidentified. In the supply function, $K – M = 3 – 2 = 1$; $G – 1 = 2 – 1 = 1$. Thus, the supply function is just or exactly identified. But the system of equations taken together does not satisfy the order condition since out of two equations only one satisfies the order condition. Now, let us consider the following demand and supply functions:

$$q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + u_{1t} \quad \alpha_1 < 0, \quad \alpha_2 > 0 \qquad (6.28)$$

$$q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + u_{2t} \quad \beta_1 > 0,\ \beta_2 > 0 \qquad (6.29)$$

Here, p and q are endogenous variables and I and p_{t-1} are exogenous variables. Applying the order condition to the demand function, we have: $K – M = 4 – 3 = 1$; $G – 1 = 2 – 1 = 1$. The demand function is therefore just identified. Now, consider the supply function: $K – M = 4 – 3 = 1$; $G – 1 = 2 – 1 = 1$. The supply function is also 'just identified'. Let us now consider the following macro model.

$$C = a_1 + b_1 Y - c_1 T + d_1 R + u_1 \qquad (6.30)$$

$$I = a_2 + b_2Y + c_2R + u_2 \quad (6.31)$$

$$Y = C + I + G \quad (6.32)$$

$$M = a_3 + b_3Y + c_3R + d_3P + u_3 \quad (6.33)$$

$$Y = a_4 + b_4N + u_4 \quad (6.34)$$

$$N^D = a_5 + b_5W + c_5P + u_5 \quad (6.35)$$

$$N^S = a_6 + b_6W + c_6P + u_6 \quad (6.36)$$

The consumption and investment functions are shown by Equations (6.30) and (6.31) respectively. The macroeconomic identity is captured by Equation (6.32). The demand for money (or liquidity preference) is represented by Equation (6.33). The production of output or income is shown by Equation (6.34). The labour demand and labour supply functions are shown by Equations (6.35) and (6.36) respectively. The labour market equilibrium condition is $N^D = N^S$. Here we have seven endogenous variables: C, I, Y, M, N, W and P and three exogenous variables viz. T, R and G. Does the Equation (6.30) and (6.36) satisfy the order condition? In Equation (6.30), we have: $K - M = 10 - 4 = 6$; $G - 1 = 7 - 1 = 6$. Thus, the consumption function is 'just identified'. In the investment function (6.31), we have: $K - M = 10 - 3 = 7$; $G - 1 = 7 - 1 = 6$. Therefore, the investment function also satisfies the order condition. In the same way, we can check for the order condition for the rest of the equations of the system.

6.4.2 Rank Condition

The rank condition is both a necessary and sufficient condition. It allows a set of simultaneous equations (called the 'structural equations') of an economic system to be tested for identification. This is done by considering all the parameters (or coefficients) of a system of equations called the 'reduced form equations'. The 'reduced form equations' are obtained by a re-arrangement of the 'structural equations' in a certain order. The Rank condition states that: "in a system of G equations, any particular equation is identified 'if and only if' it is possible to construct at least one 'non-zero determinant' of order (G-1). Such a construction should be possible from the coefficients of the variables excluded from that particular equation but contained in the other equations of the model". This condition is called the Rank condition because it refers to the rank of the matrix of parameters of excluded variables. Note that the rank of a matrix is the order of the largest non-zero determinant which can be formed from the matrix. Thus, the rank condition may be alternatively stated as 'a sufficient condition for the identification of a relationship, if the rank of the matrix of parameters of all excluded variables (endogenous and exogenous/predetermined) from an equation is = (G–1)'. To verify this with an illustration, let us consider the following system of equations:

$$y_1 = 3y_2 - 2x_1 + x_2 + u_1 \quad (6.37)$$

$$y_2 = y_3 + x_3 + u_2 \quad (6.38)$$

$$y_3 = y_1 - y_2 - 2x_3 + u_3 \quad (6.39)$$

There are 'six' steps to be followed to verify or examine the 'rank condition'. The first step is to re-write the equations in such a way that in the new set of equations the 'dependent variables' will appear first followed by the 'independent variables'. In the above system of equations, y_1, y_2, y_3 are endogenous variables and x_1, x_2, x_3 are exogenous variables. Hence, the above system of equations can be re-written as:

$$-y_1 + 3y_2 + 0y_3 - 2x_1 + x_2 + 0x_3 + u_1 = 0$$
$$0y_1 - y_2 + y_3 + 0x_1 + 0x_2 + x_3 + u_2 = 0$$
$$y_1 - y_2 - y_3 + 0x_1 + 0x_2 - 2x_3 + u_3 = 0$$

The second step is to construct the 'Table of Coefficients' comprising the parameters of the new set of equations. We construct the 'Table of Coefficients', from the above three equations, ignoring the random disturbances (u), as follows.

Equations	All Variables (Endogenous and Exogenous)					
	y_1	y_2	y_3	x_1	x_2	x_3
1st Equation	-1	3	0	-2	1	0
2nd Equation	0	-1	1	0	0	1
3rd Equation	1	-1	-1	0	0	-2

The 3rd step is to strike out the row of coefficients of the equation which is being examined for identification. For instance, if we want to examine the identification of the 2nd equation of the model we should strike out the 2nd row of the above Table of coefficients. The 4th step is to strike out the columns in which a 'non-zero coefficient' of the equation being examined appears. By deleting the relevant row and columns we are left with the coefficients of variables not included in the particular equation, but contained in the other equations of the model. For an illustration, let us consider the 2nd equation for identification in the above system. We strike out the 2nd, 3rd and 6th columns of the Table to get the following:

Equations	All Variables (Endogenous and Exogenous/Predetermined)					
	y_1	y_2	y_3	x_1	x_2	x_3
1st Equation	-1	3	0	-2	1	0
2nd Equation	~~0~~	~~-1~~	~~1~~	~~0~~	~~0~~	~~1~~
3rd Equation	1	-1	-1	0	0	-2

Therefore, we have the following Table which consists of the parameters of excluded variables:

y_1	x_1	x_2
-1	-2	1
1	0	0

The 5th step is concerned with the formation of the determinant(s) of order $(G - 1)$ and their values. If at least one of these determinants is non-zero, the equation is said to be identified. If all the determinants of order $(G - 1)$ are zero, the equation is said to be 'under-identified'. Therefore, for the 2nd equation of the system, we have the following determinants of order $(G - 1) = (3 - 1) = 2$:

$$\begin{vmatrix} -1 & -2 \\ 1 & 0 \end{vmatrix} \neq 0 \qquad \begin{vmatrix} -2 & 1 \\ 0 & 0 \end{vmatrix} = 0 \qquad \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} \neq 0$$

Thus, from the above results, we can form two non-zero determinants of order $G - 1 = 3 - 1 = 2$. Hence, the 2nd equation of our system is identified.

The final step is to confirm whether it is over or exactly identified. In order to examine this we go for the order condition: $(K - M) \geq (G - 1)$. With this criterion, if the equality sign is satisfied i.e. if $(K - M) = (G - 1)$ then the equation is said to be 'exactly identified'. If the inequality sign holds i.e. if $(K - M) > (G - 1)$, then the equation is said to be 'over identified'.

In the case of 2nd equation, we have: $G = 3$, $K = 6$ and $M = 3$. The counting rule $(K - M) \geq (G - 1)$ gives $(6 - 3) > (3 - 1)$. Therefore, the 2nd equation of the system is 'over identified'. Now, using the same rule of rank and order conditions, one can examine the status of identification of Equations (1) and (3).

6.4.3 Illustration

Let us now consider the following simple Keynesian model of income determination. We examine the status of identification by applying the order and rank condition to each equation of the system below.

Consumption function: $C_t = a_0 + a_1 Y_t - a_2 T_t + u$

Investment function: $I_t = b_0 + b_1 Y_{t-1} + v$

Tax function: $T_t = c_0 + c_1 Y_t + w$

Macroeconomic Identity: $Y_t = C_t + I_t + G_t$

In this model, we have 4 equations and 4 endogenous variables (*C, I, Y, T*), two exogenous variables viz. lagged income (Y_{t-1}) and government expenditure (*G*). Therefore, mathematically the above macro model is complete (because the number of equations in the system is equal to the

number of dependent variables). Here, $G = 4$ and $K = 6$. Let us first take up the consumption function. For this equation $M = 2$. Hence, $K - M = 4 > G - 1 = 3$. Therefore, the order condition for identification is satisfied for the consumption function. The 'Table of Structural Coefficients' is constructed as follows:

Equations	All Variables (Endogenous and Exogenous/Predetermined)					
	C	Y	T	I	Y_{t-1}	G
1st Equation	-1	a_1	$-a_2$	0	0	0
2nd Equation	0	0	0	-1	b_1	0
3rd Equation	0	c_1	-1	0	0	0
4th equation	1	-1	0	1	0	1

We strike out the first row and first three columns of the Table to obtain the 'Table of Coefficients of Excluded Variables' as follows:

Equations	All Variables (Endogenous and Exogenous/Predetermined)					
	C	Y	T	I	Y_{t-1}	G
1st Equation	~~-1~~	~~a_1~~	~~$-a_2$~~	~~0~~	~~0~~	~~0~~
2nd Equation	0	0	0	-1	b_1	0
3rd Equation	0	c_1	-1	0	0	0
4th Equation	1	-1	0	1	0	1

Therefore, we have the following 'Table of Coefficients of Excluded Variables'

I	Y_{t-1}	G
-1	b_1	0
0	0	0
1	0	1

We have to evaluate the determinant of order $G - 1 = 3$ to see if at least one of them is $\neq 0$. However,

$\Delta = -1.\ (0 * 1 - 0 * 0) - 0.\ (b_1 * 1 - 0 * 0) + 1.(b_1 * 0 - 0 * 0) = 0.$

Consequently, we cannot form any non-zero determinant of order $(G - 1) = (4 - 1) = 3$. Therefore, the rank condition is violated. Hence, we conclude that the consumption function is 'not identified' though the function satisfies the order condition. Does the 'investment function' satisfy the order and rank condition? The investment function includes two variables. Hence, $M = 2$ and $(K - M) = 6 - 2 = 4$; $(G - 1) = 4 - 1 = 3$. Thus, $(K - M) \geq (G - 1)$ is fulfilled for investment function. Therefore, the investment function satisfies the order

condition. For verifying the 'rank condition', we delete the 2nd row and the 4th and 5th columns in the 'Table of Structural Coefficients' as follows:

Equations	All Variables (Endogenous and Exogenous/Predetermined)					
	C	Y	T	I	Y_{t-1}	G
1st Equation	-1	a_1	$-a_2$	0	0	0
2nd Equation	0	0	0	-1	b_1	0
3rd Equation	0	c_1	-1	0	0	0
4th Equation	1	-1	0	1	0	1

We now we form the 'Table of Coefficients of Excluded Variables' as follows:

C	Y	T	G
-1	a_1	$-a_2$	0
0	c_1	-1	0
1	-1	0	1

Notice that the value of the first 3 × 3 determinant of the parameters of excluded variables is:

$\Delta = -1.(c_1 * 0 - 1 * 1) + 1.(a_1 * -1 + a_2.c_1) = 1 - a_1 + a_2c_1 \neq 0$ provided $a_1 - a_2c_1 \neq 1$.

The Rank condition is satisfied since we can construct at least one non-zero determinant of order 3 = (G – 1). Therefore, we can conclude that the investment function satisfies both the order and rank conditions. You can now verify whether the tax function satisfies the order and rank conditions.

Check Your Progress 3 [answer within space given in about 50-100 words]

1) What is an 'order condition'? What is its use?

……

……

……

……

……

2) What is the 'rank condition'? How is it applied?

……

……

……

……

……

3) When do you say that an equation is 'exactly identified' or 'over identified'?

……

……

……

……

……

4) The following macroeconomic model is given which includes product market, labour market and money market.

The consumption function: $C = a_1 + b_1 Y - c_1 T + d_1 r + u_1$

Investment function: $I = a_2 + b_2 Y + c_2 r + u_2$

Macroeconomic Identity: $Y = C + I + G$

Money demand function: $M = a_3 + b_3 Y + c_3 r + d_3 P + u_3$

Production function: $Y = a_4 + b_4 N + u_4$

Labour demand function: $N^D = a_5 + b_5 W + c_5 P + u_5$

Labour supply function: $N^S = a_6 + b_6 W + c_6 P + u_6$

Check the order and rank condition in each equation of the above macroeconomic model.

6.5 LET US SUM UP

The problem of identification precedes the problem of estimation. The identification problem asks 'whether one can obtain unique numerical estimates of the structural coefficients from the estimated reduced form coefficients'? If this can be done, an equation in a system of simultaneous equations is said to be identified. If this cannot be done, that equation is said to be under-identified or unidentified. An identified equation can be just identified or over identified. In the former case, unique values of structural coefficients can be obtained; in the latter, there may be more than one value for one or more structural parameters. The identification problem arises because the same set of data may be compatible with different sets of structural coefficients i.e. different models. Thus, in the regression of price on quantity only, it is difficult to tell whether one is estimating the supply function or the demand function, because, price and quantity enter both the equations. To assess the identification of a structural equation, one may apply the technique of reduced-form equations. This technique expresses an endogenous variable solely as a function of expansion variables. However, this is a time-consuming procedure which can be avoided by resorting to either the order condition or the rank condition of identification. Although the order condition is easy to apply, it provides only a necessary condition for

identification. On the other hand, the rank condition is both a necessary and sufficient condition for identification. If the rank condition is satisfied, the order condition is satisfied too. However, the converse is not true. In practice, the order condition is generally adequate to ensure identification. But, it may be necessary to test for simultaneity explicitly.

6.6 KEY WORDS

Endogeneity : In a two variable regression model, we assume one variable as endogenous (or dependent) and the other one is independent (or explanatory). The explanatory variable is assumed to be strictly exogenous (or pre-determined) in nature. If this property does not hold, we cannot easily estimate the unknown parameters (viz. regression coefficients) by applying the method of least squares. This causes the problem of endogeneity because of simultaneous dependence of the two variables.

Simultaneous Equation Bias : The bias arising from the application of classical least square to an equation belonging to a system of simultaneous relations is called simultaneous equation bias. It arises from the dependence of the explanatory variables with the error term (u_i).

Problem of Identification : By the term identification problem, we basically mean whether numerical estimates of the parameters of a structural equation can be obtained from the estimated 'reduced form coefficients'. If this cannot be done, then we say that the particular equation is unidentified or under-identified.

Order Condition : This condition is based on a counting rule of the variables included and excluded from the particular equation. It is a necessary but not sufficient condition for the identification of an equation.

Rank Condition : The Rank condition states that in a system of G equations any particular equation is identified 'if and only if' it is possible to construct at least one non-zero determinant of order (G-1) from the coefficients of the variables excluded from that particular equation but contained in the other equations of the model. This condition is called Rank condition because it refers to the rank of the matrix of parameters of excluded variables.

6.7 SUGGESTED BOOKS FOR FURTHER READING

1) Greene W (2003). Econometric Analysis, 5th Edition, Pearson Education, Inc. India.

2) Gujarati D, Porter D C and Gunasekhar S (2018). Basic Econometrics, 5th Edition, McGraw Hill Education (India) Pvt. Ltd.

3) G S Maddala (1992). Introduction to Econometrics, Second Edition, Macmillan, New York

4) Koutsoyiannis A (2001). Theory of Econometrics, Palgrave Macmillan Limited.

6.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Simultaneity refers to a situation where the causation is not limited from *X*'s to *Y* but flows in both the directions. In the presence of simultaneity, the OLS estimators will not be efficient. This happens because of the violation of the condition required for the OLS estimates to be efficient viz. $Cov(X_i, u_i) = 0$.

2) When both *X* and *Y* are jointly determined, we encounter a situation called as 'simultaneity bias'. It means that if we disregard the endogeneity or dependency between the two variables and proceed to estimate the parameters by the OLS method, the resulting estimates will be questionable or biased. This feature is termed as 'simultaneity bias'.

3) OLS estimators are inconsistent and inefficient.

Check Your Progress 2

1) An equation is said to be identified if it has an unique form enabling the determination (or estimation) of its parameters uniquely.

2) It means that the parameters of the 'structural form of the equation' cannot be estimated uniquely by its 'reduced form' of equations. In other words, in the presence of the identification problem, the numerical estimates of the parameters of the equation will not be unique. In such situations, the equation is said to be 'under identified' or 'unidentified'.

3) An equation which is identified can either be 'exactly identified' or 'over identified'. It is said to be 'exactly identified' if unique numerical estimates of its 'structural parameters' can be obtained. It is said to be 'over identified' if more than one numerical value can be obtained for some of the parameters of the structural equation.

4) Basically, the problem of identification is a 'model specification problem'. Empirically, it means that different sets of 'structural coefficients' may be compatible with the same set of data.

5) We have $\alpha_0 + \alpha_1 p_t + u_{1t} = \beta_0 + \beta_1 p_t + u_{2t}$

$\Rightarrow (\alpha_1 - \beta_1)\, p_t = (\beta_0 - \alpha_0) + (u_{2t} - u_{1t})$

$\therefore p_t = (\beta_0 - \alpha_0)/(\alpha_1 - \beta_1) + (u_{2t} - u_{1t})/(\alpha_1 - \beta_1)$

Equation (6.8) is: $q_t^D = \alpha_0 + \alpha_1 p_t + u_{1t}$

$\Rightarrow q_t = \alpha_0 + \alpha_1[(\beta_0 - \alpha_0)/(\alpha_1 - \beta_1) + (u_{2t} - u_{1t})/(\alpha_1 - \beta_1)] + u_{1t}$

$= \alpha_0 + 1/(\alpha_1 - \beta_1)[\alpha_1 \beta_0 - \alpha_1 \alpha_0 + \alpha_1 u_{2t} - \alpha_1 u_{1t}] + u_{1t}$

$\therefore q_t (\alpha_1 - \beta_1) = \alpha_0\alpha_1 - \alpha_0 \beta_1 + \alpha_1 \beta_0 - \alpha_1\alpha_0 + \alpha_1 u_{2t} - \alpha_1 u_{1t} + \alpha_1 u_{1t} - \beta_1 u_{1t}$

$\therefore q_t = (\alpha_1 \beta_0 - \alpha_0 \beta_1)/(\alpha_1 - \beta_1) + (\alpha_1 u_{2t} - \beta_1 u_{1t})/(\alpha_1 - \beta_1)$

Check Your Progress 3

1) The 'order condition' is a necessary condition but not a sufficient condition. If is used to determine whether an equation in a system is 'identified' or not. It is based on the number of variables included and excluded in a model. The condition states that '$(K - M) \geq (G - 1)$' where K = total no. of variables in the model (i.e. the entire system of equations), M = No. of variables in the equation under examination for identification and G = total no. of equations in the system. Here, the word 'variables', includes both the dependent and the independent variables.

2) Rank condition is a sufficient condition. This subsumes the 'order condition'. This means, if rank condition is satisfied, then the order condition is automatically satisfied. Application of 'rank condition' has six steps. It involves the: (i) rearrangement of original set of equations (called as the 'structural equations') in to a new set of equations called as 'reduced form of equations', (ii) construction of the Table of Coefficients, (iii) striking off the row of coefficients for the particular equation under examination for satisfaction of 'rank condition', (iv) striking off the columns in which a non-zero entry in the equation under examination [struck off in step (iii)] figures, (v) working out the value of all those 'determinants' of the order 'G – 1' to verify whether at least one such determinant is $\neq 0$ and (vi) finally applying the 'order condition' for the equation being examined to determine whether that equation is 'over identified' or 'exactly identified'.

3) This depends upon the result of the 6th step in the application of 'rank condition' as detailed above. If the inequality is 'equal', then the

equation under examination is said to be 'just identified', if it is >, then the equation is said to be 'over identified'.

4) Follow Sub-section 6.4.3. Equilibrium condition in the labour market is attained when $N^D = N^S$. Similarly, the money market is clear if $M = M^S$. Here we have seven equations and seven endogenous variables like C, I, N, P, r, Y, W and three exogenous variables like G, T and M.

UNIT 7 SIMULTANEOUS EQUATIONS MODELS II*

Structure

7.0 OBJECTIVES

After reading this unit, you will be able to:

- distinguish between 'limited information' and 'full information' methods of estimating the parameters in a system of simultaneous equations;
- state the meaning of the term 'recursive models';
- explain why the OLS method of estimation is feasible to apply in case of 'recursive simultaneous equation systems';
- specify the conditions under which the method of 'indirect least squares (ILS)' is applied;
- bring out the difference between the 'structural form' and the 'reduced form' of equations with an illustration;
- describe the estimation procedure in the 'indirect least squares (ILS)' method;

* Prof. Sushil Haldar, Jadavpur University.

- list the criteria to be satisfied for the choice of a variable to be selected as an 'instrumental variable (IV)';
- enumerate the assumptions and properties of the IV method of estimation;
- elucidate the steps involved in applying the IV method with an illustration; and
- discuss the 2SLS method of estimating a system of simultaneous equations with its underlying assumptions and characteristics.

7.1 INTRODUCTION

Economic theories are built on sets, systems and causal relationships. Examples include: (i) market equilibrium, (ii) macroeconomic models and (iii) factor or commodity demand and supply equations. Our interest may be in a particular part of the system or the system as a whole. Depending on this, we would select the variables in the model. In such models, the interactions of the variables will have important implications for both the estimation and interpretation of the parameters.

In this context, there are two broad approaches for estimating the parameters of a simultaneous equations system. One is the limited information method. The other is the full information method. The Limited information method is applied for the 'single equations' of a simultaneous equations system. In the full information methods (also known as the systems method), all the equations of the system are estimated simultaneously. Theoretically, the systems method is ideal. But in practice, it is extremely difficult to estimate their parameters. This is because it may require solutions for models that are highly non-linear in parameters. They are hence more difficult to estimate. The existence of specification error may complicate the systems method. In contrast, the single equation methods of solving for simultaneous equations are less sensitive to specification error. This is in the sense that those parts of the system that are correctly specified may not be much affected by the errors in the specification in another part. For these reasons, we shall limit our discussion in the present unit only to the singe equation methods of estimation.

7.2 ORDINARY LEAST SQUARES (OLS) METHOD

A system of equations is said to be recursive (rather than simultaneous) if each of the endogenous variables can be determined sequentially (rather than requiring to be estimated jointly). To make the same point differently, a recursive model is one where the matrix of coefficients of the endogenous variables is triangular. By 'triangular' we mean, the matrix has zeros in all the cells above the diagonal. A recursive model is also one in which there is unidirectional dependency among the endogenous variables. In view of this

feature, the equations can be ordered in such a way that the first endogenous variable is determined only by exogenous variables, the second dependent variable is determined by the first endogenous variable and the exogenous variables, the third dependent variable is determined by only the first two endogenous variables and the exogenous variables, and so on. Therefore, there is no feedback for an endogenous variable located above in the casual chain to influence one that is lower in the causal chain. To see the nature of these models in this respect, let us consider the following system:

$$Y_{1t} = \gamma_{11} X_{1t} + u_{1t} \qquad (7.1)$$

$$Y_{2t} = \gamma_{21} X_{1t} + \beta_{21} Y_{1t} + u_{2t} \qquad (7.2)$$

$$Y_{3t} = \gamma_{31} X_{1t} + \beta_{31} Y_{1t} + \beta_{32} Y_{2t} + u_{3t} \qquad (7.3)$$

where, as usual, the *Y*'s and the *X*'s are the endogenous and exogenous variables respectively. From the structure of such a system, it is clear that there is a sequential (one-way) rather than a simultaneous (i.e. two-way) relationship between the endogenous variables. Thus, Y_1 affects Y_2, but Y_2 does not affect Y_1. Similarly, Y_1 and Y_2 influence Y_3 without, in turn, being influenced by Y_3. In other words, each equation exhibits a unilateral causal dependence. Hence, they are called recursive or triangular or causal models. A recursive model is thus a special case of a systems equation wherein the endogenous variables are determined one at a time in a sequence. Thus, the right-hand side of the equation for the first endogenous variable includes no endogenous variables but only the exogenous variables.

Can we apply the OLS method to estimate the parameters of such a recursive system of simultaneous Equations? In order to explain this, let us first consider the Equation 7.1. Since it contains only the exogenous variables on the right hand side, and since by assumption it is uncorrelated with the disturbance term u_1, this equation satisfies the critical assumption of the classical OLS viz. *Cov* (X_1, u_1) = 0. Therefore, OLS can be applied to this equation. Now consider the second Equation 7.2. It contains the endogenous variable Y_1 as an explanatory variable along with the non-stochastic (i.e. exogenous) X_1. Now, OLS can be applied provided *Cov* (Y_1, u_2) = 0. Is it so? The answer is 'yes'. This is because u_1, which affects Y_1, is by assumption uncorrelated with u_2. Therefore, for all empirical purposes, Y_1 is a predetermined variable insofar as Y_2 is concerned. Hence, one can proceed with the application of OLS. Carrying this argument further, we can also apply OLS to the 3rd equation (viz. 7.3) because both Y_1 and Y_2 are uncorrelated with u_3. Do we face any simultaneous problem in such a system of equations? The answer is 'no'. It is clear that there is no interdependence among the endogenous variables from the structure of the model. Therefore, when the system is recursive, we can apply the OLS method. Sometimes, recursive models are called causal models since each equation exhibits a unilateral causal dependence. The most important assumption of such a

model is zero contemporaneous correlation. This means, the disturbance terms (u's) follow zero contemporaneous correlation. That is:

$$Cov(u_{1t}, u_{2t}) = Cov(u_{1t,} u_{3t}) = Cov(u_{2t}, u_{3t}) = 0 \qquad (7.4)$$

Why is the alternative name for the recursive models 'triangular'? To see this, consider the following transformation of Equations (7.1), (7.2) and (7.3):

$$Y_{1t} + 0.Y_{2t} + 0.Y_{3t} - \gamma_{11}X_{1t} = u_{1t} \qquad (7.5_a)$$

$$-\beta_{21}Y_{1t} + Y_{2t} + 0.Y_{3t} - \gamma_{21}X_{it} = u_{2t} \qquad (7.5_b)$$

$$-\beta_{31}Y_{1t} - \beta_{32}Y_{2t} + Y_{3t} - \gamma_{31}X_{1t} = u_{3t} \qquad (7.5_c)$$

Let us arrange the coefficient of endogenous variables as follows:

Coefficients of → in Equation ↓	Y_1	Y_2	Y_3
(7.5_a)	1	0	0
(7.5_b)	$-\beta_{21}$	1	0
(7.5_c)	$-\beta_{31}$	$-\beta_{32}$	1

Notice that the entries above the main diagonal are zero and the diagonal elements are 1. This is the reason why the recursive models are also called 'triangular'.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) Distinguish between the terms 'limited information' and 'full information' methods of estimating the parameters in a system of simultaneous equations.

……

……

……

……

……

2) State the meaning of the term 'recursive models'.

……

……

……

……

……

3) Why is it that the OLS method of estimation becomes applicable in case of 'simultaneous systems' which are recursive?

……

……

……

……

……

7.3 INDIRECT LEAST SQUARES (ILS) METHOD

The Indirect Least Squares (ILS) method is an approach where the parameters of a simultaneous equations system are estimated by: (i) using the 'reduced of equations' and (ii) applying the ordinary least squares (OLS) method to the reduced form equations obtained in (i). Hence, the method of obtaining the estimates of the structural coefficients, from the OLS estimates of the reduced form coefficients, is an indirect approach. This is the reason why the approach is called as the 'indirect least squares (ILS) method'. Therefore, before proceeding further with the method of ILS, we must understand the concepts of structural and reduced form equations. Note that the method of ILS is applied in cases where the structural equations under consideration are just or exactly identified.

7.3.1 Structural and Reduced Form Equations

A structural model is a complete system of equations. This is in the sense that it describes the entire structure of the relationship between the economic variables. More specifically, structural equations express the endogenous variables as functions of: (i) other endogenous variables, (ii) exogenous variables and (iii) the disturbance (viz. error) term. The reduced form of equations of a structural model is a system of equations in which the endogenous variables are expressed as a function of: (i) the exogenous variables and (ii) the disturbance term. In other words, the reduced form equations essentially simplifies the system of equations. The simplification is achieved by presenting the new disturbance term as a composite term in which the disturbance term in the reduced form equations subsume the endogeneity effect within them.

Let us consider the following macroeconomic model of income determination for a closed economy comprising the equations for consumption (C_t), Investment (I_t) and the national income (Y) identity as follows:

$$C_t = a_0 + a_1 Y_t + u_{1t} \qquad (7.6)$$

$$I_t = b_0 + b_1 Y_t + b_2 Y_{t-1} + u_{2t} \qquad (7.7)$$

$$Y_t = C_t + I_t + G_t \qquad (7.8)$$

Here, we have three endogenous variables: C, I and Y and two exogenous variables: Y_{t-1} and G. Equations (7.6) and (7.7) are the 'structural equations' and equation (7.8) is an identity. The structural parameters are a_0, a_1, b_0, b_1 and b_2. Note that the structural parameters express the direct effect of each explanatory variable on the dependent variable. For instance, a_1 stands for the marginal effect, which is also the direct effect. Because of this marginal effect, if we differentiate the estimated equation of consumption (C) with respect to income (Y), we get the marginal effect (or the direct effect) of income on consumption as:

$$\frac{\delta \hat{C}}{\delta Y} = a_1$$

The indirect effects can only be computed by the solutions of the structural system and not by the individual structural parameters. Note that the factors not appearing in any function explicitly may have an indirect influence on the dependent variable of that function. Hence, in the above system, a change in consumption will affect investment indirectly through a change it results in income i.e. Y which is a determinant of investment. The effect of C on Y cannot be measured directly by any of the structural parameters. But it will be taken into account by the simultaneous solution of the system. In order to express the reduced forms of C, I and Y of the simultaneous equations system, we first substitute the value of C and I in Equation (7.8) to obtain:

$$Y_t = (a_0 + a_1 Y_t + u_{1t}) + (b_0 + b_1 Y_t + b_2 Y_{t-1} + u_2) + G_t$$

$$\Rightarrow Y_t - a_1 Y_t - b_1 Y_t = (a_0 + b_0) + b_2 Y_{t-1} + G_t + (u_{1t} + u_{2t})$$

$$\Rightarrow (1 - a_1 - b_1).Y_t = (a_0 + b_0) + b_2 Y_{t-1} + G_t + (u_{1t} + u_{2t})$$

$$\Rightarrow Y_t = \frac{(a_0 + b_0)}{(1 - a_1 - b_1)} + \frac{b_2}{(1 - a_1 - b_1)}.Y_{t-1} + \frac{G_t}{(1 - a_1 - b_1)} + \frac{(u_{1t} + u_{2t})}{(1 - a_1 - b_1)} \quad (7.9)$$

This is the reduced form of the 3rd structural Equation (7.8) which represents income. Here, income (Y) is expressed as a function of exogenous variables only. In the same way, we can estimate the reduced form equation for consumption (C). For this, we substitute the value of Y (as obtained from Equation 7.9) into the consumption function (Equation 7.6) to obtain:

$$C_t = a_0 + \frac{a_1(a_0 + b_0)}{(1 - a_1 - b_1)} + \frac{a_1 b_2}{(1 - a_1 - b_1)}.Y_{t-1} + \frac{a_1 G_t}{(1 - a_1 - b_1)} + \frac{a_1(u_{1t} + u_{2t})}{(1 - a_1 - b_1)} + u_{1t}$$

$$\Rightarrow C_t = \left(\frac{a_0 - a_0 a_1 - a_0 b_1 + a_0 a_1 + a_1 b_0}{1 - a_1 - b_1}\right) + \left(\frac{a_1 b_2}{1 - a_1 - b_1}\right).Y_{t-1} + \left(\frac{a_1}{1 - a_1 - b_1}\right)G_t +$$

$$\left(\frac{a_1 u_{1t} + a_1 u_{2t} + u_{1t} - a_1 u_{1t} - b_1 u_{1t}}{1 - a_1 - b_1}\right)$$

$$\Rightarrow C_t = \left(\frac{a_0(1-b_1)+a_1b_0}{1-a_1-b_1}\right)+\left(\frac{a_1b_2}{1-a_1-b_1}\right).Y_{t-1}+\left(\frac{a_1}{1-a_1-b_1}\right).G_t+\left(\frac{u_{1t}(1-b_1)+a_1u_{2t}}{1-a_1-b_1}\right). \quad (7.10)$$

Equation (7.10) is the reduced form equation of the consumption function in which consumption is expressed as a function of exogenous variables Y_{t-1} and G_t. Similarly, the reduced form equation of investment is obtained by substituting the value of Y (as in equation 7.9) into the structural equation of investment (as in Equation 7.7) to obtain:

$$I_t = b_0 + b_1\left[\frac{(a_0+b_0)}{(1-a_1-b_1)}+\frac{b_2Y_{t-1}}{(1-a_1-b_1)}+\frac{G_t}{(1-a_1-b_1)}+\frac{(u_{1t}+u_{2t})}{(1-a_1-b_1)}\right]+b_2Y_{t-1}+u_{2t}$$

$$\Rightarrow I_t = \left(b_0+\frac{a_0b_1+b_0b_1}{(1-a_1-b_1)}\right)+b_2Y_{t-1}\left(1+\frac{b_1}{(1-a_1-b_1)}\right)+\frac{b_1}{(1-a_1-b_1)}.G_t+\left(\frac{b_1(u_{1t}+u_{2t})}{(1-a_1-b_1)}+u_{2t}\right)$$

$$\Rightarrow I_t = \left(\frac{b_0-b_0a_1-b_0b_1+a_0b_1+b_0b_1}{(1-a_1-b_1)}\right)+b_2Y_{t-1}\left(\frac{1-a_1-b_1+b_1}{(1-a_1-b_1)}\right)+\frac{b_1}{(1-a_1-b_1)}.G_t+\left(\frac{b_1u_{1t}-a_1u_{2t}+u_2}{(1-a_1-b_1)}\right)$$

$$\Rightarrow I_t = \left(\frac{b_0-b_0a_1+a_0b_1}{(1-a_1-b_1)}\right)+b_2Y_{t-1}\left(\frac{1-a_1}{(1-a_1-b_1)}\right)+\frac{b_1}{(1-a_1-b_1)}.G_t+\left(\frac{b_1u_{1t}-a_1u_{2t}+u_2}{(1-a_1-b_1)}\right).... \quad (7.11)$$

Equation (7.11) is the reduced form equation of investment function in which the RHS only has exogenous variables. Note that the method of obtaining the reduced form equations, from the structural equations, as illustrated above, merely involves a process of substitution and transformation. It enables us to obtain the reduced form equations in which the RHS will have only the known exogenous variables.

7.3.2 Assumptions and Properties

The method of ILS is based on two assumptions. Firstly, the structural equations must be exactly identified. For this, we must apply both the rank and order conditions of identification. Secondly, the error terms of the reduced form equation must satisfy the assumptions of the OLS. These are: (a) normality with zero mean and constant variance, (b) error terms are serially independent and (c) the error terms are also independent of the exogenous variables. Note that the 'error terms' of the reduced form equations is a linear combination of the 'error terms' of the structural equations. This can be illustrated as follows. The error term of the consumption function of the structural equation is u_{1t} (from Equation 7.6). The 'error term' of the reduced form equation of the consumption function (as in Equation 7.10) is $\left(\frac{u_{1t}(1-b_1)+a_1u_{2t}}{1-a_1-b_1}\right)$. This shows that the error term of the reduced form equation of the consumption function is a linear

combination of the error terms of the structural equations of the consumption (u_{1t}) and investment function (u_{2t}) [a_1 and b_1 being constants]. The third assumption we need for the ILS method is that the exogenous variables of the model must not be perfectly collinear.

If the above three assumptions are fulfilled, then the estimates of the reduced form coefficients are 'best, linear and unbiased' for large samples. For small samples, the estimates of the structural parameters obtained from the reduced form equations are biased but consistent. This means that their bias tends to zero as the sample size increases.

7.3.3 Estimation Procedure

The ILS involves the following three steps in its estimation stage.

Step-1: We first obtain the reduced form equations. The reduced form equations are obtained from the structural equations in such a way that the dependent variable in each equation is a function solely of the exogenous variables.

Step-2: We apply the OLS to the reduced form equations individually. This is permissible since the explanatory variables in these equations are independent and hence uncorrelated with the stochastic disturbance terms. The estimates obtained are also consistent.

Step-3: We obtain the estimates of the original structural coefficients from the estimated reduced form coefficients obtained in step-2. Here, if an equation is exactly identified (which is in fact the very first assumption made for the ILS method), there is a one to one correspondence between the structural and reduced form coefficients. This means that one can obtain unique estimates of the former from the latter.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) Specify the approach in the 'ILS method of estimation'.

……

……

……

……

……

……

……

……

……

2) What is an essential difference between the 'structural equations' and their 'reduced form equations'?

……

……

……

……

……

3) What does the structural parameters basically represent?

……

……

……

……

……

4) In what way the 'indirect effect' of explanatory variables on the dependent variable can be measured?

……

……

……

……

……

5) What is the single most important advantage of working with the 'reduced form equations'?

……

……

……

……

……

6) State the three assumptions of the ILS method.

……

……

……

……

……

……

7) What are the properties or characteristics of the ILS estimates?

.....

.....

.....

.....

.....

7.4 INSTRUMENTAL VARIABLE (IV) METHOD

The method of instrumental variables (IVs) [popularly abbreviated as **IV**] is used to estimate the causal relationships. It is particularly applied when controlled experiments are not either feasible or when a treatment cannot be successfully administered to each and every unit. Intuitively, therefore, the method of IVs are useful when an explanatory variable of interest is correlated with the error term [in which case the ordinary least squares (OLS) are known to give biased estimates]. A valid instrument is one which induces changes in the explanatory variable but results in no independent effect on the dependent variable. This allows a researcher to uncover the causal effect of the explanatory variable on the dependent variable. The IV approach is a single equation method. This means it is applied on a system of equations one at a time. It is developed as a solution to the 'simultaneous equation bias' and is appropriate for over-identified systems. In principle, the IV method attains the reduction of the dependence of error term (u) with the explanatory variables, by using appropriate exogenous variables as instruments.

7.4.1 Selection

The major problem of IV method for applying it to solve a simultaneous equations system is the optimal choice of the instrument(s). An instrumental variable is an exogenous variable located somewhere in the system of simultaneous equations. It should satisfy the following conditions: (i) it must be strongly correlated with the endogenous variable which will be replaced in the structural equation, (ii) it must be truly exogenous and hence uncorrelated with the random error term (u) of the structural equation, (iii) it must be least correlated with the exogenous variables already included in the set of explanatory variables of the particular structural equation (as these variables will be serving as instruments for themselves), (iv) if more than one instrumental variable is used in the same structural equation, they must be least correlated with each other, and (v) the number of instrumental variables must not be more than the number of dependent variables in the particular structural equation.

7.4.2 Assumptions and Properties

There are two assumptions behind the IV method. One is that the instrumental variables chosen are appropriate. This is in the sense that they

are based on the knowledge of the independent variables in the other equations of the simultaneous system. The IVs chosen are used as multipliers of the structural equations. Hence, after multiplication, the new error terms must satisfy the standard assumptions of the OLS method. This is the second assumption made for the IV method. Note that as a part of the second assumption it is implied that there is no perfect collinearity with the exogenous variables ultimately included in the system of equations. If these assumptions are met, then the estimates of the structural parameters have the following two properties: (a) the estimates are biased in case of small samples but are consistent for large samples and (b) the estimates, although consistent, are not asymptotically efficient. This means that they do not possess the minimum variance property.

7.4.3 Estimation Procedure

The execution of the IV method involves the following two steps.

Step-1: Choose the appropriate instrumental variables to replace the endogenous variables appearing as explanatory variables on the right-hand side of the structural equation.

Step-2: Multiply both sides of the structural equation by the instrumental variable and sum up the equation over all sample observations. This step provides as many linear equations as there are unknown parameters. Hence, the solutions of these equations, are helpful to obtain unique estimates of the structural parameters.

Let us now consider an example to understand the application of the above two steps in practice. For this, we shall first take up the case of one explanatory variable. We know that poverty (Y) is caused by low health status (X). Now, although poverty is also caused by many other factors, let us in the first instance consider only these two variables. Then, the structural equation would contain only one explanatory variable X. This is evidently correlated with error term (u) because low health status is not the only determinant of poverty. Hence, application of the OLS method to the equation:

$$Y = \beta_0 + \beta_1 X + u \qquad (7.12)$$

will yield biased and inconsistent estimates [because low health status (X) is also caused by low level of income]. In order to avoid this difficulty, it suffices to include in some other part of the system of equations an exogenous variable like public expenditure on healthcare system (Z). Such a variable, used as an instrumental multiplier, fulfils the condition that Z is strongly correlated with X but not with u. Therefore, we can use Z as an instrumental variable for replacing X in our function. For this, we first write Equation (7.12) in deviations form to eliminate the intercept term β_0. We then get:

$$y = \beta_1 x + u \qquad (7.13)$$

where, $y = Y - \overline{Y}$, $x = X - \overline{X}$ and $u = u - \overline{u}$ with $\overline{u} = 0$ from the normality assumption of the error term. Now, multiplying (7.13) by $z = Z - \overline{Z}$ and summing over all the sample values, we get:

$$\sum zy = \beta_1 \sum zx + \sum zu \tag{7.14}$$

Now, since by assumption z and u are uncorrelated we have $\sum zu = 0$. Hence, the estimated coefficient $\beta_1^* = \dfrac{\sum zy}{\sum zx}$ serves as the estimated parameter by the IV method for the case of a single explanatory variable considered here. Let us now consider an example for applying the IV method when we have two explanatory variables. Let the two explanatory variables be X_1 and X_2. Since they are endogenous to the system i.e. they are not entirely alien to the system of equations, they are correlated with the error term. Note that the original equation is assumed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \tag{7.15}$$

We rewrite the Equation (7.15) in the 'deviation from the mean' form as:

$$y = \beta_1 x_1 + \beta_2 x_2 + u \tag{7.16}$$

Due to endogeneity, $E(X_1, u) \neq 0$ and $E(X_2, u) \neq 0$. Hence, in order to avoid least squares bias, it suffices to know that this equation is part of a larger system of simultaneous equations in which the exogenous variables Z_1 and Z_2 could be introduced provided they satisfy the conditions of instrumental variables. We can then use them as multipliers of the structural equation and sum up over all sample observations to get the following two equations (in deviation form):

$$\sum z_1 y = \beta_1 \sum z_1 x_1 + \beta_2 \sum z_1 x_2 + \sum z_1 u \tag{7.17}$$

$$\sum z_2 y = \beta_1 \sum z_2 x_1 + \beta_2 \sum z_2 x_2 + \sum z_2 u \tag{7.18}$$

Since, $Cov\ (Z_1, u) = Cov\ (Z_2, u) = 0$, the last term from both of the equations vanishes. We therefore are left with two equations and two unknowns, β_1 and β_2. Hence, using Cramer's rule, we can find unique solution to the parameters as:

$$\beta_1 = \frac{\left(\sum z_1 y\right)\left(\sum z_2 x_2\right) - \left(\sum z_2 y\right)\left(\sum z_1 x_2\right)}{\left(\sum z_1 x_1\right)\left(\sum z_2 x_2\right) - \left(\sum z_2 x_1\right)\left(\sum z_1 x_2\right)}$$

$$\beta_2 = \frac{\left(\sum z_2 y\right)\left(\sum z_1 x_1\right) - \left(\sum z_1 y\right)\left(\sum z_2 x_1\right)}{\left(\sum z_1 x_1\right)\left(\sum z_2 x_2\right) - \left(\sum z_2 x_1\right)\left(\sum z_1 x_2\right)}$$

Thus, the major issue is to choose the exogenous variables Z_1 and Z_2. There are some limitations of the IV Method. Firstly, the selection of the instrumental variable(s) remains arbitrary to some extent. Hence, the estimates will differ depending on the instrumental variables chosen. In other

words, there will be no unique estimates. Secondly, due to the complex causal mechanism in the simultaneous equation system, every exogenous variable, directly and indirectly, affects the endogenous variables of the system. The IV method ignores the influence of exogenous variables omitted from the system. Thirdly, each endogenous variable might be related to more than one exogenous variable. This means, the choice of a particular variable, as an instrumental variable to it, is an incomplete choice. In other words, since the exogenous variables are usually inter-correlated at least to an extent, the task of finding out an 'appropriate' instrumental variable is never attainable completely. Fourthly, the error term being unobservable, it is difficult to say with certainty whether the error term is independent of the instrumental variables. Given these imitations, the IV method qualifies as an easy to apply estimate which can be used for estimation with different sets of instrumental variables. This is particularly the case when one is doubtful as to which is an appropriate IV. It provides us with a choice from among the estimates to choose the one with the least variance property.

Check Your Progress 3 [answer within the space given in about 50-100 words]

1) Under what circumstances is the IV method of estimation preferable to apply?

……

……

……

……

……

2) State the characteristics of an IV. What does the method basically achieve?

……

……

……

……

3) Specify the two assumptions of the IV method.

……

……

……

4) State the properties of the IV method. What is a major limitation of this method?

……

……

……

……

……

5) What are the other limitations of the IV method?

……

……

……

……

……

6) Despite its limitations, how is the IV method yet quite useful?

……

……

……

……

……

7.5 TWO STAGE LEAST SQUARES (2 SLS) METHOD

The 2-stage least squares (2SLS) approach is an extension of the OLS method. It is applicable in case of over-identified models. It is used when the error term is correlated with the explanatory variables of an equation in the system of simultaneous equations. The method makes the following assumptions: (i) the disturbance term (u) of the original structural equations satisfy the stochastic assumptions of zero mean, constant variance and zero covariance, (ii) the error term of reduced form equation also satisfies the stochastic assumptions of zero correlation with all the exogenous variables in the system, (iii) the explanatory variables are not perfectly collinear, (iv) the specification of the model is correct so far as the exogenous variables are concerned and (v) the sample is large enough i.e. the number of observations exceed the number of exogenous variables in the structural system.

7.5.1 Characteristics

If the above assumptions are met, then the 2SLS estimates possess the following characteristics.

i) It is applied for each single equation of the simultaneous equations system;

ii) The 2SLS Method provides only one estimate per parameter whereas in ILS we may have multiple estimates of parameters.

iii) It is easy to apply in the simultaneous equation system.

iv) The 2SLS is designed to solve a problem of simultaneity in case of over identified model; however, it can be applied in case of exactly or just identified model also. Under this case, the ILS and 2SLS will yield same results.

v) If the R^2 in the first stage regressions are very low, the 2SLS estimates will be practically meaningless because we shall be replacing the original Y's in the 2nd stage regression by the estimated Y's from the 1st stage regression, which will essentially represent the disturbances in the first stage regressions.

vi) The estimates from 2SLS are consistent which means that as sample size increases, the estimates converge to true population values. However, the estimates may not satisfy the small sample properties. Therefore, the results obtained by applying the method of 2SLS to small sample and the inferences drawn from them should be interpreted with due caution.

vii) The method of 2SLS assumes knowledge of all the predetermined variables of the complete system of simultaneous equations. If the specification of these variables is not correct, the estimates of the parameters will not have the optimal properties.

viii) The 2SLS method is fairly simple compared to any standard methods of solutions of simultaneous equation problem.

7.5.2 Estimation Procedure

In order to explain the estimation process of 2SLS, we consider two endogenous, two exogenous variables and two structural equations as shown below:

$$Y_{1t} = \beta_0 + \beta_1 Y_{2t} + \gamma_1 X_{1t} + \gamma_2 X_{2t} + u_{1t} \qquad (7.17)$$

$$Y_{2t} = \beta_2 + \beta_3 Y_{1t} + u_{2t} \qquad (7.18)$$

Here, Y_1 and Y_2 are the endogenous variables; and X_1 and X_2 are the exogenous variables. The order condition suggests that the first Equation (7.17) is under-identified but the 2nd Equation (7.18) is over identified. Therefore, ILS cannot be applied. The 2SLS and IV Methods are the feasible options to estimate the parameters. There is a great similarity between IV and 2SLS Method; the two results give more or less same results. Here, we apply the 2SLS method with the following two stages:

Stage-1: There exists endogeneity between Y_1 and Y_2, therefore, $Cov(Y_1, u_2) \neq 0$. In order to get rid of the likely correlation between Y_1 and u_2, we regress Y_1 on all exogenous variables in the whole system. In the present case, this means regressing Y_1 on X_1 and X_2:

$$Y_{it} = \pi_0 + \pi_1 X_{1t} + \pi_2 X_{2t} + u_t \qquad (7.19).$$

The estimated Equation of (7.19) can be written as:

$$\hat{Y}_{it} = \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \hat{\pi}_2 X_{2t} \qquad (7.20)$$

where $\hat{Y}_{it}$ is an estimate of the mean value of Y conditional upon the fixed X's. It is interesting to note that Equation (7.19) is the reduced form regression because only the exogenous variables appear on the right hand side. We know the relationship between observed value (Y_{1t}), estimated or predicted value ($\hat{Y}_{1t}$) and the error term (u) as:

$$\hat{u}_t = Y_{1t} - \hat{Y}_{1t} \text{ or, } Y_{1t} = \hat{u}_t + \hat{Y}_{1t} \qquad (7.21)$$

This shows that the stochastic Y_1 consists of two parts $\hat{Y}_{1t}$, which is a linear combination of the non-stochastic X's and a random component $\hat{u}_t$. Basic principle of OLS states that $\hat{Y}_{1t}$ and $\hat{u}_t$ is uncorrelated.

Stage-2: In the 2nd stage, we insert the value of Y_{1t} (as obtained from Equation 7.21) in Equation (7.18), to get:

$$\begin{aligned} Y_{2t} &= \beta_2 + \beta_3(\hat{Y}_{1t} + \hat{u}_t) + u_{2t} \\ &= \beta_2 + \beta_3 \hat{Y}_{1t} + (u_{2t} + \beta_3 \hat{u}_t) \\ &= \beta_2 + \beta_3 \hat{Y}_{1t} + u_t^* \end{aligned} \qquad (7.22)$$

where, $u_t^* = u_{2t} + \beta_3 \hat{u}_t$

Now, if we compare Equation (7.22) with Equation (7.18), we see that they are very much similar in appearance with the difference in the error term and the value of Y_1. What is the advantage of Equation (7.22)? It can be shown that although Y_1 in the original (structural equation) is correlated (or likely to be correlated) with the disturbance term (u_2), $\hat{Y}_{1t}$ in Equation (7.22) is uncorrelated with u_t^* as sample size increases indefinitely. Therefore, OLS can be applied to Equation (7.22) from which we get give consistent estimates of the parameters of the structural equations.

7.6 LET US SUM UP

The ordinary least square (OLS) method is applied in case of full recursive model of simultaneous equation system. In this case, there is no need to examine the rank or order condition of identification of simultaneity. The ILS is applicable if the model is just identified. One must examine the rank and order conditions of identification before, to apply the ILS, IV and 2SLS methods. The IV and 2SLS are applied in case of over identified models of simultaneous equation system. Theoretically, the 2SLS may be considered as an extension of ILS and IV methods. The ILS, 2SLS and IV methods aim at the elimination of simultaneous equation biases or the 'problem of endogeneity' in the simultaneous equation system. In a simultaneous

equations system, each endogenous variable is associated with more than one exogeneous variables. Therefore, the choice of a particular variable as instrumental is extremely difficult. The selection to a certain extent is arbitrary. The 2SLS estimates are identical with the IV method, provided that we use as instruments the estimated values ($\hat{Y}_i$) of endogenous variables on the right side to replace the original *Y*'s (i) and (ii) the exogenous variables included in the function are as their own instruments.

The 2SLS method resembles the IV method of estimation in that the linear combination of the exogeneous variables serves as an instrument, or proxy, for the endogenous regressors. The feature of both ILS and 2SLS is that the estimates obtained are consistent i.e. as the sample size increases indefinitely, the estimates converge to their true population values. The estimates may not satisfy small-sample properties such as unbiasedness and minimum variance.

The method of 2SLS is more general than the IV method because it takes into account the influence on the dependent variable from all the exogeneous variables of the system. The method of IV however, chooses only a certain subset of the exogeneous variables (as instruments) and ignores the effects of the other exogenous variables. A major problem of 2SLS is the existence of specification errors.

7.7 KEY WORDS

Reduced Form of Equations : The reduced form method is a single equation method in that it is applied to one equation of a system at a time.

Recursive Model : A recursive model is a special case of an equation system where the endogenous variables are determined one at a time in sequence. Thus the right-hand side of the equation for the first endogenous variable includes no endogenous variables, but only the exogenous variables.

Indirect Least Squares Method : Indirect least squares is an approach where the coefficients in a simultaneous equations model are estimated from the reduced form model using the ordinary least squares approach. Once the coefficients are thus estimated, the structural form of the model is estimated using the reduced form coefficient in the next stage..

Instrumental Variable Method : Instrumental variables (IVs) are used to control for confounding and measurement error in observational studies. They allow for the possibility of making causal inferences with observational data. The IVs can adjust for both observed and unobserved confounding effects.

2SLS Method : Two-Stage least squares (**2SLS**) regression analysis is a statistical technique that is used in the analysis of structural equations. This technique is the extension of the OLS method. It is used when the dependent variable's error terms are correlated with the independent variables.

7.8 SUGGESTED BOOKS FOR FURTHER READING

1) G S Maddala, Kajal Lahiri (2009). Introduction to Econometrics, 4th Edition, Wiley.

2) Damodar N Gujarati, Dawn C Porter, and Sangeetha Gunasekar (2015). Basic Econometrics, 5th Edition, McGraw-Hill Education (India) Pvt Limited.

3) Koutsoyiannis A (2001). Theory of Econometrics, Palgrave Macmillan Limited.

7.9 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) They are the two broad approaches to solving the parameters of a simultaneous equations system. In the 'limited information' method, we proceed to estimate the parameters of each equation in the system individually. In the 'full information' method, we do the same 'jointly' or 'simultaneously'. Another essential distinction between the two approaches relate to the sensitivity of the 'specification errors' to the system of equations. The 'full information' methods are more sensitive than the 'limited information' methods in this regard.

2) A system of equations is said to be recursive (rather than simultaneous) if each of the endogenous variables can be determined sequentially (rather than requiring to be estimated jointly). Another way of distinguishing such models is to relate the direction of causation among the dependent variables. If Y_1, Y_2, ……, Y_n are the dependent variables, and if the direction of causation is from Y_1 to Y_2 and not from Y_2 to Y_1, from Y_1 and Y_2 to Y_3 and not from Y_3 to Y_1 and Y_2, and so on, the direction of causation, or dependency, is said to be uni-directional. This is an essential feature of recursive models. In such a model, if the coefficients of the endogenous variables are arranged in the form of a matrix, the diagonal entries will all be 1 and the elements above the diagonal entry will be zero. In view of this feature, recursive models are also said to be 'triangular'.

3) It is because, by assumption, the covariance between the *X'*s and *U'*s, *Y'*s and *U'*s, and between the *U'*s are all zero.

Check Your Progress 2

1) The ILS method requires that the original 'structural equations' be first expressed in a slightly different form known as its 'reduced form'. This step basically involves a process of substituting and transposing of the terms. On such a reduced form of equations, the OLS method is then applied to estimate its parameters. Hence the name 'ILS method' for its approach.

2) There are three determinants to the dependent variables in the structural equations viz. endogenous variables, exogenous variables and the error term. In the reduced form equations, these are reduced to two viz. exogenous variables and the error term. The reduced form equations are thus a simplified version in the sense that their error terms are composite i.e. they subsume the endogeneity effect within their error terms.

3) They express the direct effect of each explanatory variable on the dependent variable. They are the marginal effect of the exogenous (or the explanatory) variable on the endogenous (or the dependent) variable.

4) The indirect effects can be computed only by the solutions of the 'structural system' and not by the individual structural parameters.

5) In the 'structural equations' the dependent variables will be expressed both in terms of 'the dependent variables' themselves along with the exogenous variables. In the reduced form equations only the exogenous variables will be there on the RHS. This is of course along with the error term which will be present in both the structural and the reduced form equations. The exogenous variables being independent and pre-determined, the reduced form equations are simpler to proceed with the application of OLS for estimating their parameters. With these estimates, we can go back to estimating all the structural parameters.

6) (i) structural equations must be exactly identified. (ii) the error terms of the reduced form equations must satisfy the assumptions of OLS viz. normality with zero mean and constant variance, absence of serial correlation and the error terms are independent of exogenous variables. (iii) absence of multicollinearity among the regressors or the exogenous variables.

7) For large samples, they are 'best, linear and unbiased'. For small samples they are biased but consistent.

Check Your Progress 3

1) It is useful when an explanatory variable of interest is correlated with the error term. It is appropriate for over-identified systems.

2) A valid instrument is one which induces changes in the explanatory variable but results in no independent effect on the dependent variable. The IV method attains the reduction of the dependence of error term (u) with the explanatory variables by using appropriate exogenous variables as instruments.

3) (i) it is appropriate in the sense that they are based on the knowledge of the other explanatory variables in the system, (ii) the new error terms, after performing the multiplication of the structural equation with the IV, must satisfy the assumptions of the OLS method.

4) For large samples, they are unbiased and consistent. A limitation is that, although they are consistent for large samples, they are not asymptotically efficient. This means, they do not possess the minimum variance property.

5) (i) choice of IV is arbitrary and hence there are no unique estimates, (ii) the IV method ignores the influence of exogenous variables omitted from the system, (iii) the choice of a particular variable, as an instrumental variable to it, is an incomplete choice since the exogenous variables are usually inter-correlated at least to an extent and (iv) the error term being unobservable, it is difficult to say with certainty whether the error term is independent of the instrumental variables.

6) By providing us with a choice from among the estimates (obtained by alternative variables used as IVs) to choose the one with the least variance property.