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:

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

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:

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

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

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:

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

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

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:

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

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

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, $x = \dfrac{\beta_1}{\beta_2}$. 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:

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

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.

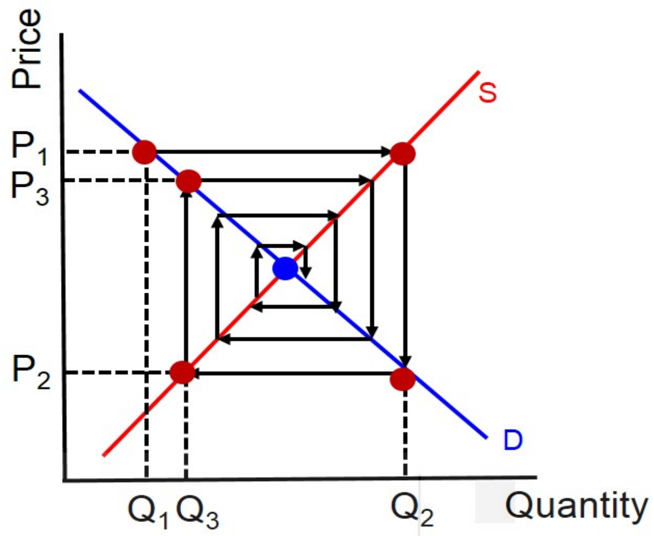

Fig. 4.1 Market Equilibrium

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'.

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

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.

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

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:

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

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

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:

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

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

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) ≠ 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:

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

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:

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

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 $\delta = 1$, 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 $\delta = 0$, 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:

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

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 $(v_t+\delta u_t)$ 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 $(1–\delta)$ 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 $0 < \gamma \le 1$. 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:

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

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

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:

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

$$Y_t = (\delta\gamma\beta_0^*) + (\delta\gamma\beta_1^*)X_t + \left[(1-\gamma) + (1-\delta))\right]Y_{t-1} - \left[(1-\gamma)(1-\delta)\right]Y_{t-2} +$$
$$\left[u_t - (1-\gamma)u_{t-1}\right] \qquad (1.36)$$

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.