17

DYNAMIC ECONOMETRIC MODELS: AUTOREGRESSIVE AND DISTRIBUTED-LAG MODELS

In regression analysis involving time series data, if the regression model includes not only the current but also the lagged (past) values of the explanatory variables (the X's), it is called a **distributed-lag model.** If the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called an **autoregressive model.** Thus,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$$

represents a distributed-lag model, whereas

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

is an example of an autoregressive model. The latter are also known as **dynamic models** since they portray the time path of the dependent variable in relation to its past value(s).

Autoregressive and distributed-lag models are used extensively in econometric analysis, and in this chapter we take a close look at such models with a view to finding out the following:

1. What is the role of lags in economics?
2. What are the reasons for the lags?
3. Is there any theoretical justification for the commonly used lagged models in empirical econometrics?
4. What is the relationship, if any, between autoregressive and distributed-lag models? Can one be derived from the other?

5. What are some of the statistical problems involved in estimating such models?

6. Does a lead–lag relationship between variables imply causality? If so, how does one measure it?

17.1 THE ROLE OF "TIME," OR "LAG," IN ECONOMICS

In economics the dependence of a variable Y (the dependent variable) on another variable(s) X (the explanatory variable) is rarely instantaneous. Very often, Y responds to X with a lapse of time. Such a lapse of time is called a *lag*. To illustrate the nature of the lag, we consider several examples.

EXAMPLE 17.1

THE CONSUMPTION FUNCTION

Suppose a person receives a salary increase of $2000 in annual pay, and suppose that this is a "permanent" increase in the sense that the increase in salary is maintained. What will be the effect of this increase in income on the person's annual consumption expenditure?

Following such a gain in income, people usually do not rush to spend all the increase immediately. Thus, our recipient may decide to increase consumption expenditure by $800 in the first year following the salary increase in income, by another $600 in the next year, and by another $400 in the following year, saving the remainder. By the end of the third year, the person's annual consumption expenditure will be increased by $1800. We can thus write the consumption function as

$$Y_t = \text{constant} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + u_t \tag{17.1.1}$$

where Y is consumption expenditure and X is income.

Equation (17.1.1) shows that the effect of an increase in income of $2000 is spread, or distributed, over a period of 3 years. Models such as (17.1.1) are therefore called **distributed-lag models** because the effect of a given cause (income) is spread over a number of time periods. Geometrically, the distributed-lag model (17.1.1) is shown in Figure 17.1, or alternatively, in Figure 17.2.

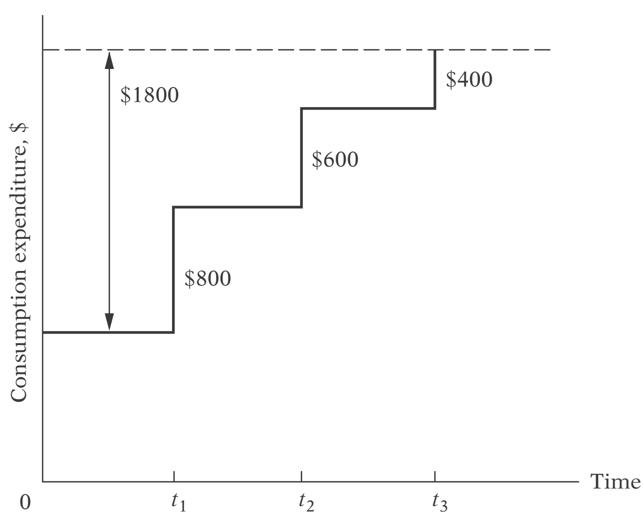

FIGURE 17.1 Example of distributed lags.

(*Continued*)

EXAMPLE 17.1 (*Continued*)

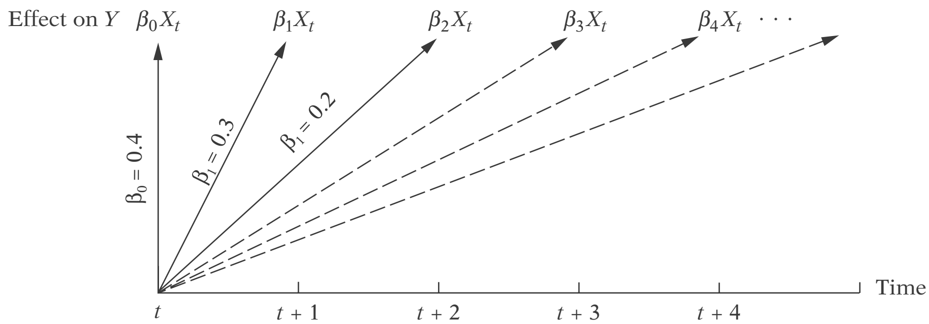

FIGURE 17.2
The effect of a unit change in X at time t on Y at time t and subsequent time periods.

More generally we may write

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \qquad \textbf{(17.1.2)}$$

which is a distributed-lag model with a finite lag of k time periods. The coefficient β_0 is known as the **short-run,** or **impact, multiplier** because it gives the change in the mean value of Y following a unit change in X in the same time period.[1] If the change in X is maintained at the same level thereafter, then, $(\beta_0 + \beta_1)$ gives the change in (the mean value of) Y in the next period, $(\beta_0 + \beta_1 + \beta_2)$ in the following period, and so on. These partial sums are called **interim,** or **intermediate, multipliers.** Finally, after k periods we obtain

$$\sum_{i=0}^{k} \beta_i = \beta_0 + \beta_1 + \beta_2 + \cdots + \beta_k = \beta \qquad \textbf{(17.1.3)}$$

which is known as the **long-run,** or **total, distributed-lag multiplier,** provided the sum β exists (to be discussed elsewhere).

If we define

$$\beta_i^* = \frac{\beta_i}{\sum \beta_i} = \frac{\beta_i}{\beta} \qquad \textbf{(17.1.4)}$$

we obtain "standardized" β_i. Partial sums of the standardized β_i then give the proportion of the long-run, or total, impact felt by a certain time period.

Returning to the consumption regression (17.1.1), we see that the short-run multiplier, which is nothing but the short-run marginal propensity to consume (MPC), is 0.4, whereas the long-run multiplier, which is the

[1]Technically, β_0 is the partial derivative of Y with respect to X_t, β_1 that with respect to X_{t-1}, β_2 that with respect to X_{t-2}, and so forth. Symbolically, $\partial Y_t / \partial X_{t-k} = \beta_k$.

long-run marginal propensity to consume, is $0.4 + 0.3 + 0.2 = 0.9$. That is, following a \$1 increase in income, the consumer will increase his or her level of consumption by about 40 cents in the year of increase, by another 30 cents in the next year, and by yet another 20 cents in the following year. The long-run impact of an increase of \$1 in income is thus 90 cents. If we divide each β_i by 0.9, we obtain, respectively, 0.44, 0.33, and 0.23, which indicate that 44 percent of the total impact of a unit change in X on Y is felt immediately, 77 percent after one year, and 100 percent by the end of the second year.

EXAMPLE 17.2

CREATION OF BANK MONEY (DEMAND DEPOSITS)

Suppose the Federal Reserve System pours \$1000 of new money into the banking system by buying government securities. What will be the total amount of bank money, or demand deposits, that will be generated ultimately?

Following the fractional reserve system, if we assume that the law requires banks to keep a 20 percent reserve backing for the deposits they create, then by the well-known multiplier process the total amount of demand deposits that will be generated will be equal to $\$1000[1/(1 - 0.8)] = \$5000$. Of course, \$5000 in demand deposits will not be created overnight. The process takes time, which can be shown schematically in Figure 17.3.

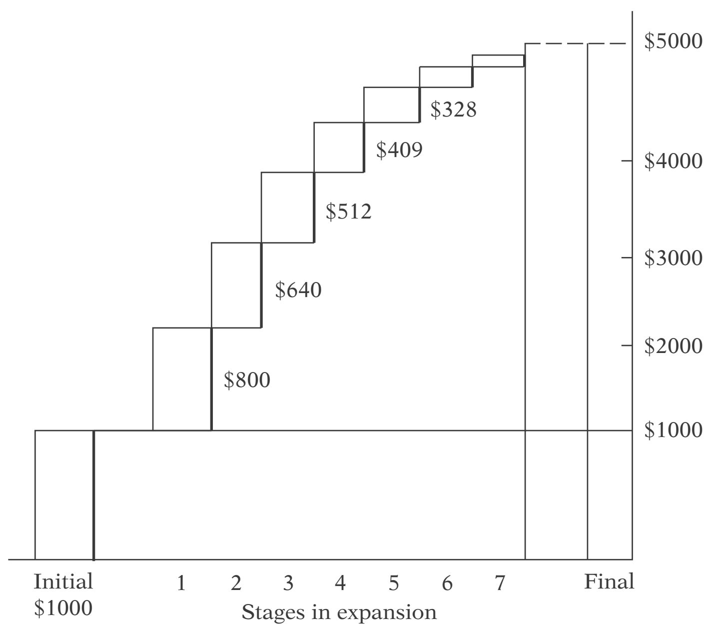

FIGURE 17.3
Cumulative expansion in bank deposits (initial reserve \$1000 and 20 percent reserve requirement).

EXAMPLE 17.3

LINK BETWEEN MONEY AND PRICES

According to the monetarists, inflation is essentially a monetary phenomenon in the sense that a continuous increase in the general price level is due to the rate of expansion in money supply far in excess of the amount of money actually demanded by the economic units. Of course, this link between inflation and changes in money supply is not instantaneous. Studies have shown that the lag between the two is anywhere from 3 to about 20 quarters. The results of one such study are shown in Table 17.1,[2] where we see the effect of a 1 percent change in the M1B money supply ($=$ currency $+$ checkable deposits at financial institutions) is felt over a period of 20 quarters. The long-run impact of a 1 percent change in the money supply on inflation is about 1 ($=\sum m_i$), which is statistically significant, whereas the short-run impact is about 0.04, which is not significant, although the intermediate multipliers seem to be generally significant. Incidentally, note that since P and M are both in percent forms, the m_i (β_i in our usual notation) give the elasticity of P with respect to M, that is, the percent response of prices to a 1 percent increase in the money supply. Thus, $m_0 = 0.041$ means that for a 1 percent increase in the money supply the short-run elasticity of prices is about 0.04 percent. The long-term elasticity is 1.03 percent, implying that in the long run a 1 percent increase in the money supply is reflected by just about the same percentage increase in the prices. In short, a 1 percent increase in the money supply is accompanied in the long run by a 1 percent increase in the inflation rate.

TABLE 17.1 ESTIMATE OF MONEY-PRICE EQUATION: ORIGINAL SPECIFICATION

Sample period: 1955–I to 1969–IV: $m_{21} = 0$

$$\dot{P} = \underset{(0.395)}{-0.146} + \sum_{i=0}^{20} m_i \dot{M}_{-i}$$

	Coeff.	$\lvert t \rvert$		Coeff.	$\lvert t \rvert$		Coeff.	$\lvert t \rvert$
m_0	0.041	1.276	m_8	0.048	3.249	m_{16}	0.069	3.943
m_1	0.034	1.538	m_9	0.054	3.783	m_{17}	0.062	3.712
m_2	0.030	1.903	m_{10}	0.059	4.305	m_{18}	0.053	3.511
m_3	0.029	2.171	m_{11}	0.065	4.673	m_{19}	0.039	3.338
m_4	0.030	2.235	m_{12}	0.069	4.795	m_{20}	0.022	3.191
m_5	0.033	2.294	m_{13}	0.072	4.694	$\sum m_i$	1.031	7.870
m_6	0.037	2.475	m_{14}	0.073	4.468	**Mean lag**	10.959	5.634
m_7	0.042	2.798	m_{15}	0.072	4.202			
$\bar{R}^2$	0.525							
se	1.066							
D.W.	2.00							

Notation: $\dot{P} =$ compounded annual rate of change of GNP deflator
$\dot{M} =$ compounded annual rate of change of M1B

Source: Keith M. Carlson, "The Lag from Money to Prices," *Review,* Federal Reserve Bank of St. Louis, October 1980, Table 1, p. 4.

[2]Keith M. Carlson, "The Lag from Money to Prices," *Review,* Federal Reserve Bank of St. Louis, October, 1980, Table 1, p. 4.

EXAMPLE 17.4

LAG BETWEEN R&D EXPENDITURE AND PRODUCTIVITY

The decision to invest in research and development (R&D) expenditure and its ultimate payoff in terms of increased productivity involve considerable lag, actually several lags, such as, ". . . the lag between the investment of funds and the time inventions actually begin to appear, the lag between the invention of an idea or device and its development up to a commercially applicable stage, and the lag which is introduced by the process of diffusion: it takes time before all the old machines are replaced by the better new ones."[3]

EXAMPLE 17.5

THE *J* CURVE OF INTERNATIONAL ECONOMICS

Students of international economics are familiar with what is called the *J curve,* which shows the relationship between trade balance and depreciation of currency. Following depreciation of a country's currency (e.g., due to devaluation), initially the trade balance deteriorates but eventually it improves, assuming other things are the same. The curve is as shown in Figure 17.4.

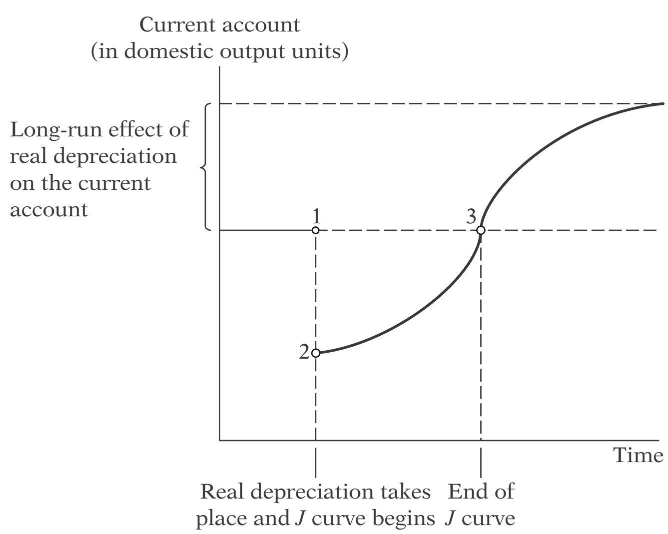

FIGURE 17.4 The *J* curve.

Source: Paul R. Krugman and Maurice Obstfeld, *International Economics: Theory and Practice,* 3d ed., Harper Collins, New York, 1994, p. 465.

[3]Zvi Griliches, "Distributed Lags: A Survey," *Econometrica,* vol. 36, no. 1, January 1967, pp. 16–49.

EXAMPLE 17.6

THE ACCELERATOR MODEL OF INVESTMENT

In its simplest form, the acceleration principle of investment theory states that investment is proportional to changes in output. Symbolically,

$$I_t = \beta(X_t - X_{t-1}) \qquad \beta > 0 \tag{17.1.5}$$

where I_t is investment at time t, X_t is output at time t, and X_{t-1} is output at time $(t - 1)$.

The preceding examples are only a sample of the use of lag in economics. Undoubtedly, the reader can produce several examples from his or her own experience.

17.2 THE REASONS FOR LAGS[4]

Although the examples cited in Section 17.1 point out the nature of lagged phenomena, they do not fully explain why lags occur. There are three main reasons:

1. Psychological reasons. As a result of the force of habit (inertia), people do not change their consumption habits immediately following a price decrease or an income increase perhaps because the process of change may involve some immediate disutility. Thus, those who become instant millionaires by winning lotteries may not change the lifestyles to which they were accustomed for a long time because they may not know how to react to such a windfall gain immediately. Of course, given reasonable time, they may learn to live with their newly acquired fortune. Also, people may not know whether a change is "permanent" or "transitory." Thus, my reaction to an increase in my income will depend on whether or not the increase is permanent. If it is only a nonrecurring increase and in succeeding periods my income returns to its previous level, I may save the entire increase, whereas someone else in my position might decide to "live it up."

2. Technological reasons. Suppose the price of capital relative to labor declines, making substitution of capital for labor economically feasible. Of course, addition of capital takes time (the gestation period). Moreover, if the drop in price is expected to be temporary, firms may not rush to substitute capital for labor, especially if they expect that after the temporary drop the price of capital may increase beyond its previous level. Sometimes, imperfect knowledge also accounts for lags. At present the market for personal computers is glutted with all kinds of computers with varying features and prices. Moreover, since their introduction in the late 1970s, the prices of most personal computers have dropped dramatically. As a result, prospective consumers for the personal computer may hesitate to buy until they have

[4]This section leans heavily on Marc Nerlove, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities,* Agricultural Handbook No. 141, U.S. Department of Agriculture, June 1958.

had time to look into the features and prices of all the competing brands. Moreover, they may hesitate to buy in the expectation of further decline in price or innovations.

3. Institutional reasons. These reasons also contribute to lags. For example, contractual obligations may prevent firms from switching from one source of labor or raw material to another. As another example, those who have placed funds in long-term savings accounts for fixed durations such as 1 year, 3 years, or 7 years, are essentially "locked in" even though money market conditions may be such that higher yields are available elsewhere. Similarly, employers often give their employees a choice among several health insurance plans, but once a choice is made, an employee may not switch to another plan for at least 1 year. Although this may be done for administrative convenience, the employee is locked in for 1 year.

For the reasons just discussed, lag occupies a central role in economics. This is clearly reflected in the short-run–long-run methodology of economics. It is for this reason we say that short-run price or income elasticities are generally smaller (in absolute value) than the corresponding long-run elasticities or that short-run marginal propensity to consume is generally smaller than long-run marginal propensity to consume.

17.3 ESTIMATION OF DISTRIBUTED-LAG MODELS

Granted that distributed-lag models play a highly useful role in economics, how does one estimate such models? Specifically, suppose we have the following distributed-lag model in one explanatory variable:[5]

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + u_t \qquad \textbf{(17.3.1)}$$

where we have not defined the length of the lag, that is, how far back into the past we want to go. Such a model is called an **infinite (lag) model,** whereas a model of the type (17.1.2) is called a **finite (lag) distributed-lag model** because the length of the lag k is specified. We shall continue to use (17.3.1) because it is easy to handle mathematically, as we shall see.[6]

How do we estimate the α and β's of (17.3.1)? We may adopt two approaches: (1) ad hoc estimation and (2) a priori restrictions on the β's by assuming that the β's follow some systematic pattern. We shall consider ad hoc estimation in this section and the other approach in Section 17.4.

Ad Hoc Estimation of Distributed-Lag Models

Since the explanatory variable X_t is assumed to be nonstochastic (or at least uncorrelated with the disturbance term u_t), X_{t-1}, X_{t-2}, and so on, are nonstochastic, too. Therefore, in principle, the ordinary least squares (OLS) can

[5]If there is more than one explanatory variable in the model, each variable may have a lagged effect on Y. For simplicity only, we assume one explanatory variable.

[6]In practice, however, the coefficients of the distant X values are expected to have negligible effect on Y.

be applied to (17.3.1). This is the approach taken by Alt[7] and Tinbergen.[8] They suggest that to estimate (17.3.1) one may proceed *sequentially;* that is, 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 the regression coefficients of the lagged variables start becoming statistically insignificant and/or the coefficient of at least one of the variables changes signs from positive to negative or vice versa. Following this precept, Alt regressed fuel oil consumption Y on new orders X. Based on the quarterly data for the period 1930–1939, the results were as follows:

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

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

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

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

Alt chose the second regression as the "best" one because in the last two equations the sign of X_{t-2} was not stable and in the last equation the sign of X_{t-3} was negative, which may be difficult to interpret economically.

Although seemingly straightforward, ad hoc estimation suffers from many drawbacks, such as the following:

1. There is no a priori guide as to what is the maximum length of the lag.[9]

2. As one estimates successive lags, there are fewer degrees of freedom left, making statistical inference somewhat shaky. Economists are not usually that lucky to have a long series of data so that they can go on estimating numerous lags.

3. More importantly, in economic time series data, successive values (lags) tend to be highly correlated; hence multicollinearity rears its ugly head. As noted in Chapter 10, multicollinearity leads to imprecise estimation; that is, the standard errors tend to be large in relation to the estimated coefficients. As a result, based on the routinely computed t ratios, we may tend to declare (erroneously), that a lagged coefficient(s) is statistically insignificant.

4. The sequential search for the lag length opens the researcher to the charge of **data mining.** Also, as we noted in Section 13.4, the nominal and true level of significance to test statistical hypotheses becomes an important issue in such sequential searches [see Eq. (13.4.2)].

In view of the preceding problems, the ad hoc estimation procedure has very little to recommend it. Clearly, some prior or theoretical considerations must be brought to bear upon the various β's if we are to make headway with the estimation problem.

---

[7] F. F. Alt, "Distributed Lags," *Econometrica,* vol. 10, 1942, pp. 113–128.

[8] J. Tinbergen, "Long-Term Foreign Trade Elasticities," *Metroeconomica,* vol. 1, 1949, pp. 174–185.

[9] If the lag length, k, is incorrectly specified, we will have to contend with the problem of misspecification errors discussed in Chap. 13. Also keep in mind the warning about **data mining.**

17.4 THE KOYCK APPROACH TO DISTRIBUTED-LAG MODELS

Koyck has proposed an ingenious method of estimating distributed-lag models. Suppose we start with the infinite lag distributed-lag model (17.3.1). *Assuming that the β's are all of the same sign*, Koyck assumes that they decline geometrically as follows.[10]

$$\beta_k = \beta_0 \lambda^k \qquad k = 0, 1, \ldots \tag{17.4.1}$$ [11]

where λ, such that $0 < \lambda < 1$, is known as the *rate of decline*, or *decay*, of the distributed lag and where $1 - \lambda$ is known as the *speed of adjustment*.

What (17.4.1) postulates is that each successive β coefficient is numerically less than each preceding β (this statement follows since $\lambda < 1$), implying that as one goes back into the distant past, the effect of that lag on Y_t becomes progressively smaller, a quite plausible assumption. After all, current and recent past incomes are expected to affect current consumption expenditure more heavily than income in the distant past. Geometrically, the Koyck scheme is depicted in Figure 17.5.

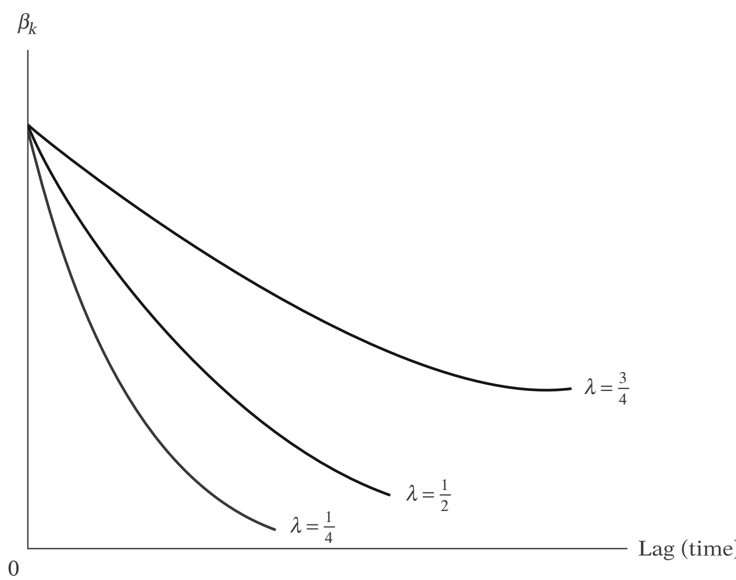

FIGURE 17.5 Koyck scheme (declining geometric distribution).

[10]L. M. Koyck, *Distributed Lags and Investment Analysis,* North Holland Publishing Company, Amsterdam, 1954.

[11]Sometimes this is also written as

$$\beta_k = \beta_0(1 - \lambda)\lambda^k \qquad k = 0, 1, \ldots$$

for reasons given in footnote 12.

As this figure shows, the value of the lag coefficient β_k depends, apart from the common β_0; on the value of λ. The closer λ is to 1, the slower the rate of decline in β_k, whereas the closer it is to zero, the more rapid the decline in β_k. In the former case, distant past values of X will exert sizable impact on Y_t, whereas in the latter case their influence on Y_t will peter out quickly. This pattern can be seen clearly from the following illustration:

λ	β_0	β_1	β_2	β_3	β_4	β_5	···	β_{10}
0.75	β_0	$0.75\beta_0$	$0.56\beta_0$	$0.42\beta_0$	$0.32\beta_0$	$0.24\beta_0$	···	$0.06\beta_0$
0.25	β_0	$0.25\beta_0$	$0.06\beta_0$	$0.02\beta_0$	$0.004\beta_0$	$0.001\beta_0$	···	0.0

Note these features of the Koyck scheme: (1) By assuming nonnegative values for λ, Koyck rules out the β's from changing sign; (2) by assuming $\lambda < 1$, he gives lesser weight to the distant β's than the current ones; and (3) he ensures that the sum of the β's, which gives the long-run multiplier, is finite, namely,

$$\sum_{k=0}^{\infty} \beta_k = \beta_0 \left(\frac{1}{1-\lambda} \right) \tag{17.4.2}$$ [12]

As a result of (17.4.1), the infinite lag model (17.3.1) may be written as

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \cdots + u_t \tag{17.4.3}$$

As it stands, the model is still not amenable to easy estimation since a large (literally infinite) number of parameters remain to be estimated and the parameter λ enters in a highly nonlinear form: Strictly speaking, the method of linear (in the parameters) regression analysis cannot be applied to such a model. But now Koyck suggests an ingenious way out. He lags (17.4.3) by one period to obtain

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_0 \lambda X_{t-2} + \beta_0 \lambda^2 X_{t-3} + \cdots + u_{t-1} \tag{17.4.4}$$

He then multiplies (17.4.4) by λ to obtain

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \beta_0 \lambda^3 X_{t-3} + \cdots + \lambda u_{t-1} \tag{17.4.5}$$

[12]This is because

$$\sum \beta_k = \beta_0(1 + \lambda + \lambda^2 + \lambda^3 + \cdots) = \beta_0 \left(\frac{1}{1-\lambda} \right)$$

since the expression in the parentheses on the right side is an infinite geometric series whose sum is $1/(1-\lambda)$ provided $0 < \lambda < 1$. In passing, note that if β_k is as defined in footnote 11, $\sum \beta_k = \beta_0(1-\lambda)/(1-\lambda) = \beta_0$ thus ensuring that the weights $(1-\lambda)\lambda^k$ sum to one.

Subtracting (17.4.5) from (17.4.3), he gets

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + (u_t - \lambda u_{t-1}) \qquad \textbf{(17.4.6)}$$

or, rearranging,

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t \qquad \textbf{(17.4.7)}$$

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

The procedure just described is known as the **Koyck transformation.** Comparing (17.4.7) with (17.3.1), we see the tremendous simplification accomplished by Koyck. Whereas before we had to estimate α and an infinite number of β's, we now have to estimate only three unknowns: α, β_0, and λ. Now there is no reason to expect multicollinearity. In a sense multicollinearity is resolved by replacing $X_{t-1}, X_{t-2}, \ldots,$ by a single variable, namely, Y_{t-1}. But note the following features of the Koyck transformation:

1. We started with a distributed-lag model but ended up with an autoregressive model because Y_{t-1} appears as one of the explanatory variables. This transformation shows how one can "convert" a distributed-lag model into an autoregressive model.

2. The appearance of Y_{t-1} is likely to create some statistical problems. Y_{t-1}, like Y_t, is stochastic, which means that we have a stochastic explanatory variable in the model. Recall that the classical least-squares theory is predicated on the assumption that the explanatory variables either are nonstochastic or, if stochastic, are distributed independently of the stochastic disturbance term. Hence, we must find out if Y_{t-1} satisfies this assumption. (We shall return to this point in Section 17.8.)

3. In the original model (17.3.1) the disturbance term was u_t, whereas in the transformed model it is $v_t = (u_t - \lambda u_{t-1})$. The statistical properties of v_t depend on what is assumed about the statistical properties of u_t, for, as shown later, if the original u_t's are serially uncorrelated, the v_t's are serially correlated. Therefore, we may have to face up to the serial correlation problem in addition to the stochastic explanatory variable Y_{t-1}. We shall do that in Section 17.8.

4. The presence of lagged Y violates one of the assumptions underlying the Durbin–Watson d test. Therefore, we will have to develop an alternative to test for serial correlation in the presence of lagged Y. One alternative is the **Durbin *h* test,** which is discussed in Section 17.10.

As we saw in (17.1.4), the partial sums of the standardized β_i tell us the proportion of the long-run, or total, impact felt by a certain time period. In practice, though, the **mean** or **median lag** is often used to characterize the nature of the lag structure of a distributed lag model.

The Median Lag

The median lag is the time required for the first half, or 50 percent, of the total change in Y following a unit sustained change in X. For the Koyck model, the median lag is as follows (see exercise 17.6):

$$\text{Koyck model: Median lag} = -\frac{\log 2}{\log \lambda} \tag{17.4.8}$$

Thus, if $\lambda = 0.2$ the median lag is 0.4306, but if $\lambda = 0.8$ the median lag is 3.1067. Verbally, in the former case 50 percent of the total change in Y is accomplished in less than half a period, whereas in the latter case it takes more than 3 periods to accomplish the 50 percent change. But this contrast should not be surprising, for as we know, the higher the value of λ the lower the speed of adjustment, and the lower the value of λ the greater the speed of adjustment.

The Mean Lag

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

$$\text{Mean lag} = \frac{\sum_0^\infty k\beta_k}{\sum_0^\infty \beta_k} \tag{17.4.9}$$

which is simply the weighted average of all the lags involved, with the respective β coefficients serving as weights. In short, it is a **lag-weighted average** of time. For the Koyck model the mean lag is (see exercise 17.7)

$$\text{Koyck model: Mean lag} = \frac{\lambda}{1-\lambda} \tag{17.4.10}$$

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

From the preceding discussion it is clear that the median and mean lags serve as a summary measure of the speed with which Y responds to X. In the example given in Table 17.1 the mean lag is about 11 quarters, showing that it takes quite some time, on the average, for the effect of changes in the money supply to be felt on price changes.

EXAMPLE 17.7

PER CAPITA PERSONAL CONSUMPTION

This example examines per capita personal consumption expenditure (PPCE) in relation to per capita disposable income (PPDI) in the United States for the period 1970–1999, all data in chained 1996 dollars. As an illustration of the Koyck model, consider the data given in Table 17.2. Regression of PPCE on PPDI and lagged PPCE gave the following results:

$$\begin{aligned}\widehat{\text{PPCE}}_t &= -1242.169 \quad + \ 0.6033\text{PPDI}_t + \ 0.4106\text{PPCE}_{t-1} \\ \text{se} &= \quad (402.5784) \quad (0.1502) \qquad (0.1546) \\ t &= \quad (-3.0855) \quad (4.0155) \qquad (2.6561) \\ & R^2 = 0.9926 \qquad d = 1.0056 \qquad \text{Durbin } h = 5.119\end{aligned} \tag{17.4.11}$$

Note: The calculation of Durbin h is discussed in Section 17.10.

If we assume that this model resulted from a Koyck-type transformation, λ is 0.4106. The median lag is:

$$-\frac{\log(2)}{\log\lambda} = -\frac{\log(2)}{\log(0.4106)} = 0.7786$$

and the mean lag is:

$$\frac{\lambda}{1-\lambda} = \frac{0.4106}{0.5894} = 0.6966$$

In words, it seems that PPCE adjusts to PPDI within a relatively short time.

TABLE 17.2 PPCE AND PPDI, 1970–1999

Observation	PPCE	PPDI	Observation	PPCE	PPDI
1970	11,300	12,823	1985	16,020	18,229
1971	11,581	13,218	1986	16,541	18,641
1972	12,149	13,692	1987	16,398	18,870
1973	12,626	14,496	1988	17,463	19,522
1974	12,407	14,268	1989	17,760	19,833
1975	12,551	14,393	1990	17,899	20,058
1976	13,155	14,873	1991	17,677	19,919
1977	13,583	15,256	1992	17,989	20,318
1978	14,035	15,845	1993	18,399	20,384
1979	14,230	16,120	1994	18,910	20,709
1980	14,021	16,063	1995	19,294	21,055
1981	14,069	16,265	1996	19,727	21,385
1982	14,105	16,328	1997	20,232	21,838
1983	14,741	16,673	1998	20,989	22,672
1984	15,401	17,799	1999	21,901	23,191

Notes: PPCE = per capita personal consumption expenditure, in 1996 dollars.
PPDI = per capita personal disposable income, in 1996 dollars.
Source: Economic Report of the President, 2001, Table B-31, p. 311.

17.5 RATIONALIZATION OF THE KOYCK MODEL: THE ADAPTIVE EXPECTATIONS MODEL

Although very neat, the Koyck model (17.4.7) is ad hoc since it was obtained by a purely algebraic process; it is devoid of any theoretical underpinning. But this gap can be filled if we start from a different perspective. Suppose we postulate the following model:

$$Y_t = \beta_0 + \beta_1 X_t^* + u_t \tag{17.5.1}$$

where Y = demand for money (real cash balances)
X^* = equilibrium, optimum, expected long-run or normal rate of interest
u = error term

Equation (17.5.1) postulates that the demand for money is a function of *expected* (in the sense of anticipation) rate of interest.

Since the expectational variable X^* is not directly observable, let us propose the following hypothesis about how expectations are formed:

$$X_t^* - X_{t-1}^* = \gamma(X_t - X_{t-1}^*) \qquad \textbf{(17.5.2)}^{13}$$

where γ, such that $0 < \gamma \le 1$, is known as the **coefficient of expectation.** Hypothesis (17.5.2) is known as the **adaptive expectation, progressive expectation,** or **error learning** hypothesis, popularized by Cagan[14] and Friedman.[15]

What (17.5.2) implies is that "economic agents will adapt their expectations in the light of past experience and that in particular they will learn from their mistakes."[16] More specifically, (17.5.2) states that expectations are revised each period by a fraction γ of the gap between the current value of the variable and its previous expected value. Thus, for our model this would mean that expectations about interest rates are revised each period by a fraction γ of the discrepancy between the rate of interest observed in the current period and what its anticipated value had been in the previous period. Another way of stating this would be to write (17.5.2) as

$$X_t^* = \gamma X_t + (1 - \gamma)X_{t-1}^* \tag{17.5.3}$$

[13]Sometimes the model is expressed as

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

[14]P. Cagan, "The Monetary Dynamics of Hyperinflations," in M. Friedman (ed.), *Studies in the Quantity Theory of Money,* University of Chicago Press, Chicago, 1956.

[15]Milton Friedman, *A Theory of the Consumption Function,* National Bureau of Economic Research, Princeton University Press, Princeton, N.J., 1957.

[16]G. K. Shaw, *Rational Expectations: An Elementary Exposition,* St. Martin's Press, New York, 1984, p. 25.

which shows that the expected value of the rate of interest at time t is a weighted average of the actual value of the interest rate at time t and its value expected in the previous period, with weights of γ and $1-\gamma$, respectively. If $\gamma = 1$, $X_t^* = X_t$, meaning that expectations are realized immediately and fully, that is, in the same time period. If, on the other hand, $\gamma = 0$, $X_t^* = X_{t-1}^*$, meaning that expectations are static, that is, "conditions prevailing today will be maintained in all subsequent periods. Expected future values then become identified with current values."[17]

Substituting (17.5.3) into (17.5.1), we obtain

$$\begin{aligned} Y_t &= \beta_0 + \beta_1[\gamma X_t + (1-\gamma)X_{t-1}^*] + u_t \\ &= \beta_0 + \beta_1\gamma X_t + \beta_1(1-\gamma)X_{t-1}^* + u_t \end{aligned} \tag{17.5.4}$$

Now lag (17.5.1) one period, multiply it by $1-\gamma$, and subtract the product from (17.5.4). After simple algebraic manipulations, we obtain

$$\begin{aligned} Y_t &= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + u_t - (1-\gamma)u_{t-1} \\ &= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + v_t \end{aligned} \tag{17.5.5}$$

where $v_t = u_t - (1-\gamma)u_{t-1}$.

Before proceeding any further, let us note the difference between (17.5.1) and (17.5.5). In the former, β_1 measures the average response of Y to a unit change in X^*, the equilibrium or long-run value of X. In (17.5.5), on the other hand, $\gamma\beta_1$ measures the average response of Y to a unit change in the actual or observed value of X. These responses will not be the same unless, of course, $\gamma = 1$, that is, the current and long-run values of X are the same. In practice, we first estimate (17.5.5). Once an estimate of γ is obtained from the coefficient of lagged Y, we can easily compute β_1 by simply dividing the coefficient of X_t $(=\gamma\beta_1)$ by γ.

The similarity between the adaptive expectation model (17.5.5) and the Koyck model (17.4.7) should be readily apparent although the interpretations of the coefficients in the two models are different. Note that like the Koyck model, the adaptive expectations model is autoregressive and its error term is similar to the Koyck error term. We shall return to the estimation of the adaptive expectations model in Section 17.8 and to some examples in Section 17.12. Now that we have sketched the adaptive expectations (AE) model, how realistic is it? It is true that it is more appealing than the purely algebraic Koyck approach, but is the AE hypothesis reasonable? In favor of the AE hypothesis one can say the following:

> It provides a fairly simple means of modelling expectations in economic theory whilst postulating a mode of behaviour upon the part of economic agents which

[17] Ibid., pp. 19–20.

> seems eminently sensible. The belief that people learn from experience is obviously a more sensible starting point than the implicit assumption that they are totally devoid of memory, characteristic of static expectations thesis. Moreover, the assertion that more distant experiences exert a lesser effect than more recent experience would accord with common sense and would appear to be amply confirmed by simple observation.[18]

Until the advent of the **rational expectations (RE) hypothesis,** initially put forward by J. Muth and later propagated by Robert Lucas and Thomas Sargent, the AE hypothesis was quite popular in empirical economics. The proponents of the RE hypothesis contend that the AE hypothesis is inadequate because it relies solely on the past values of a variable in formulating expectations,[19] whereas the RE hypothesis assumes, "that individual economic agents use *current available* and *relevant* information in forming their expectations and do not rely purely upon past experience."[20] In short, the RE hypothesis contends that "expectations are 'rational' in the sense that they efficiently incorporate *all* information available at the time the expectation is formulated"[21] and not just the past information.

The criticism directed by the RE proponents against the AE hypothesis is well-taken, although there are many critics of the RE hypothesis itself.[22] This is not the place to get bogged down with this rather heady material. Perhaps one could agree with Stephen McNees that, "At best, the adaptive expectations assumption can be defended only as a 'working hypothesis' proxying for a more complex, perhaps changing expectations formulation mechanism."[23]

EXAMPLE 17.8

EXAMPLE 17.7 REVISITED

If we consider the model given in Eq. (17.4.11), as generated by the adaptive expectations mechanism (i.e., PPCE as a function of expected PPDI), then γ, the expectations coefficient can be obtained from (17.5.5) as: $\gamma = 1 - 0.4106 = 0.5894$. Then, following the preceding discussion about the AE model, we can say that about 59 percent of the discrepancy between actual and expected PPCE is eliminated within a year.

---

[18]Ibid., p. 27.

[19]Like the Koyck model, it can be shown that, under AE, expectations of a variable are an exponentially weighted average of past values of that variable.

[20]G. K. Shaw, op. cit., p. 47. For additional details of the RE hypothesis, see Steven M. Sheffrin, *Rational Expectations,* Cambridge University Press, New York, 1983.

[21]Stephen K. McNees, "The Phillips Curve: Forward- or Backward-Looking?" *New England Economic Review,* July–August 1979, p. 50.

[22]For a recent critical appraisal of the RE hypothesis, see Michael C. Lovell, "Test of the Rational Expectations Hypothesis," *American Economic Review,* March 1966, pp. 110–124.

[23]Stephen K. McNees, op. cit., p. 50.

17.6 ANOTHER RATIONALIZATION OF THE KOYCK MODEL: THE STOCK ADJUSTMENT, OR PARTIAL ADJUSTMENT, MODEL

The adaptive expectation model is one way of rationalizing the Koyck model. Another rationalization is provided by Marc Nerlove in the so-called **stock adjustment** or **partial adjustment model (PAM).**[24] To illustrate this model, consider the **flexible accelerator model** of economic theory, which assumes that there is an *equilibrium*, *optimal*, *desired*, or *long-run* amount of capital stock needed to produce a given output under the given state of technology, rate of interest, etc. For simplicity assume that this desired level of capital Y_t^* is a linear function of output X as follows:

$$Y_t^* = \beta_0 + \beta_1 X_t + u_t \tag{17.6.1}$$

Since the desired level of capital is not directly observable, Nerlove postulates the following hypothesis, known as the **partial adjustment,** or **stock adjustment, hypothesis:**

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1}) \qquad \textbf{(17.6.2)}^{25}$$

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

Since $Y_t - Y_{t-1}$, the change in capital stock between two periods, is nothing but investment, (17.6.2) can alternatively be written as

$$I_t = \delta(Y_t^* - Y_{t-1}) \tag{17.6.3}$$

where I_t = investment in time period t.

Equation (17.6.2) postulates that the actual change in capital stock (investment) 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 stock of capital is equal to the desired stock; that is, actual stock adjusts to the desired stock instantaneously (in the same time period). However, if $\delta = 0$, it means that nothing changes since actual stock at time t is the same as that observed in the previous time period. Typically, δ is expected to lie between these extremes since adjustment to the desired stock of capital is likely to be

---

[24] Marc Nerlove, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities,* op. cit.

[25] Some authors do not add the stochastic disturbance term u_t to the relation (17.6.1) but add it to this relation, believing that if the former is truly an equilibrium relation, there is no scope for the error term, whereas the adjustment mechanism can be imperfect and may require the disturbance term. In passing, note that (17.6.2) is sometimes also written as

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

incomplete because of rigidity, inertia, contractual obligations, etc.—hence the name **partial adjustment model.** Note that the adjustment mechanism (17.6.2) alternatively can be written as

$$Y_t = \delta Y_t^* + (1 - \delta)Y_{t-1} \quad \textbf{(17.6.4)}$$

showing that the observed capital stock at time t is a weighted average of the desired capital stock at that time and the capital stock existing in the previous time period, δ and $(1 - \delta)$ being the weights. Now substitution of (17.6.1) into (17.6.4) gives

$$\begin{aligned} Y_t &= \delta(\beta_0 + \beta_1 X_t + u_t) + (1 - \delta)Y_{t-1} \\ &= \delta\beta_0 + \delta\beta_1 X_t + (1 - \delta)Y_{t-1} + \delta u_t \end{aligned} \quad \textbf{(17.6.5)}$$

This model is called the **partial adjustment model (PAM).**

Since (17.6.1) represents the long-run, or equilibrium, demand for capital stock, (17.6.5) can be called the *short-run* demand function for capital stock since in the short run the existing capital stock may not necessarily be equal to its long-run level. Once we estimate the short-run function (17.6.5) and obtain the estimate of the adjustment coefficient δ (from the coefficient of Y_{t-1}), we can easily derive the long-run function by simply dividing $\delta\beta_0$ and $\delta\beta_1$ by δ and omitting the lagged Y term, which will then give (17.6.1).

Geometrically, the partial adjustment model can be shown as in Figure 17.6.[26] In this figure Y^* is the desired capital stock and Y_1 the current actual capital stock. For illustrative purposes assume that $\delta = 0.5$. This implies that the firm plans to close half the gap between the actual and the

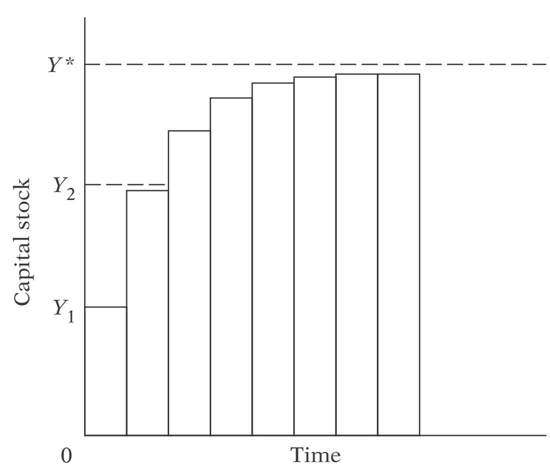

FIGURE 17.6 The gradual adjustment of the capital stock.

[26]This is adapted from Figure 7.4 from Rudiger Dornbusch and Stanley Fischer, *Macroeconomics,* 3d ed., McGraw-Hill, New York, 1984, p. 216.

desired stock of capital each period. Thus, in the first period it moves to Y_2, with investment equal to $(Y_2 - Y_1)$, which in turn is equal to half of $(Y^* - Y_1)$. In each subsequent period it closes half the gap between the capital stock at the beginning of the period and the desired capital stock Y^*.

The partial adjustment model resembles both the Koyck and adaptive expectation models in that it is autoregressive. But it has a much simpler disturbance term: the original disturbance term u_t multiplied by a constant δ. But bear in mind that although similar in appearance, the adaptive expectation and partial adjustment models are conceptually very different. The former is based on uncertainty (about the future course of prices, interest rates, etc.), whereas the latter is due to technical or institutional rigidities, inertia, cost of change, etc. However, both of these models are theoretically much sounder than the Koyck model.

Since in appearance the adaptive expectations and partial adjustment models are indistinguishable, the γ coefficient of 0.5894 of the adaptive expectations model can also be interpreted as the δ coefficient of the stock adjustment model if we assume that the latter model is operative in the present case (i.e., it is the desired or expected PPCE that is linearly related to the current PDPI).

The important point to keep in mind is that since Koyck, adaptive expectations, and stock adjustment models—apart from the difference in the appearance of the error term—yield the same final estimating model, one must be extremely careful in telling the reader which model the researcher is using and why. Thus, researchers must specify the theoretical underpinning of their model.

*17.7 COMBINATION OF ADAPTIVE EXPECTATIONS AND PARTIAL ADJUSTMENT MODELS

Consider the following model:

$$Y_t^* = \beta_0 + \beta_1 X_t^* + u_t \tag{17.7.1}$$

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

Since both Y_t^* and X_t^* are not directly observable, one could use the partial adjustment mechanism for Y_t^* and the adaptive expectations model for X_t^* to arrive at the following estimating equation (see exercise 17.2):

$$\begin{aligned} Y_t &= \beta_0\delta\gamma + \beta_1\delta\gamma X_t + [(1-\gamma) + (1-\delta)]Y_{t-1} \\ &\quad - (1-\delta)(1-\gamma)Y_{t-2} + [\delta u_t - \delta(1-\gamma)u_{t-1}] \\ &= \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + v_t \end{aligned} \tag{17.7.2}$$

*Optional.

where $\nu_t = \delta[u_t - (1 - \gamma)u_{t-1}]$. This model too is autoregressive, the only difference from the purely adaptive expectations model being that Y_{t-2} appears along with Y_{t-1} as an explanatory variable. Like Koyck and the AE models, the error term in (17.7.2) follows a moving average process. Another feature of this model is that although the model is linear in the α's, it is nonlinear in the original parameters.

A celebrated application of (17.7.1) has been Friedman's permanent income hypothesis, which states that "permanent" or long-run consumption is a function of "permanent" or long-run income.[27]

The estimation of (17.7.2) presents the same estimation problems as the Koyck's or the AE model in that all these models are autoregressive with similar error structures. In addition, (17.7.2) involves some nonlinear estimation problems that we consider briefly in exercise 17.10, but do not delve into in this book.

17.8 ESTIMATION OF AUTOREGRESSIVE MODELS

From our discussion thus far we have the following three models:

Koyck

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + (u_t - \lambda u_{t-1}) \tag{17.4.7}$$

Adaptive expectation

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1 - \gamma)Y_{t-1} + [u_t - (1 - \gamma)u_{t-1}] \tag{17.5.5}$$

Partial adjustment

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1 - \delta)Y_{t-1} + \delta u_t \tag{17.6.5}$$

All these models have the following common form:

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \nu_t \tag{17.8.1}$$

that is, they are all autoregressive in nature. Therefore, we must now look at the estimation problem of such models, because the classical least-squares may not be directly applicable to them. **The reason is twofold: the presence of stochastic explanatory variables and the possibility of serial correlation.**

Now, as noted previously, for the application of the classical least-squares theory, it must be shown that the stochastic explanatory variable Y_{t-1} is

[27]Milton Friedman, *A Theory of Consumption Function,* Princeton University Press, Princeton, N.J., 1957.

distributed independently of the disturbance term v_t. To determine whether this is so, it is essential to know the properties of v_t. If we assume that the original disturbance term u_t satisfies all the classical assumptions, such as $E(u_t) = 0$, $\text{var}(u_t) = \sigma^2$ (the assumption of homoscedasticity), and $\text{cov}(u_t, u_{t+s}) = 0$ for $s \neq 0$ (the assumption of no autocorrelation), v_t may not inherit all these properties. Consider, for example, the error term in the Koyck model, which is $v_t = (u_t - \lambda u_{t-1})$. Given the assumptions about u_t, we can easily show that v_t is serially correlated because

$$E(v_t v_{t-1}) = -\lambda\sigma^2 \qquad \textbf{(17.8.2)}^{28}$$

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

$$\text{cov}[Y_{t-1}, (u_t - \lambda u_{t-1})] = -\lambda\sigma^2 \qquad \textbf{(17.8.3)}$$

which is the same as (17.8.2). The reader can verify that the same holds true of the adaptive expectations model.

What is the implication of the finding that in the Koyck model as well as the adaptive expectations model the stochastic explanatory variable Y_{t-1} is correlated with the error term v_t? As noted previously, **if an explanatory variable in a regression model is correlated with the stochastic disturbance term, the OLS estimators are not only biased but also not even consistent; that is, even if the sample size is increased indefinitely, the estimators do not approximate their true population values.**[29] **Therefore, estimation of the Koyck and adaptive expectation models by the usual OLS procedure may yield seriously misleading results.**

The partial adjustment model is different, however. In this model $v_t = \delta u_t$, where $0 < \delta \leq 1$. Therefore, if u_t satisfies the assumptions of the classical linear regression model given previously, so will δu_t. Thus, OLS estimation of the partial adjustment model will yield consistent estimates although the estimates tend to be biased (in finite or small samples).[30] Intuitively, the reason for consistency is this: Although Y_{t-1} depends on u_{t-1} and all the

[28]
$$\begin{aligned} E(v_t v_{t-1}) &= E(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2}) \\ &= -\lambda E(u_{t-1})^2 \qquad \text{since covariances between } u\text{'s are zero by assumption} \\ &= -\lambda\sigma^2 \end{aligned}$$

[29]The proof is beyond the scope of this book and may be found in Griliches, op. cit., pp. 36–38. However, see Chap. 18 for an outline of the proof in another context. See also Asatoshi Maeshiro, "Teaching Regressions with a Lagged Dependent Variable and Autocorrelated Disturbances," *The Journal of Economic Education,* Winter 1996, vol. 27, no. 1, pp. 72–84.

[30]For proof, see J. Johnston, *Econometric Methods,* 3d ed., McGraw-Hill, New York, 1984, pp. 360–362. See also H. E. Doran and J. W. B. Guise, *Single Equation Methods in Econometrics: Applied Regression Analysis,* University of New England Teaching Monograph Series 3, Armidale, NSW, Australia, 1984, pp. 236–244.

previous disturbance terms, it is not related to the current error term u_t. Therefore, as long as u_t is serially independent, Y_{t-1} will also be independent or at least uncorrelated with u_t, thereby satisfying an important assumption of OLS, namely, noncorrelation between the explanatory variable(s) and the stochastic disturbance term.

Although OLS estimation of the stock, or partial, adjustment model provides consistent estimation because of the simple structure of the error term in such a model, one should not assume that it applies rather than the Koyck or adaptive expectations model.[31] The reader is strongly advised against doing so. A model should be chosen on the basis of strong theoretical considerations, not simply because it leads to easy statistical estimation. Every model should be considered on its own merit, paying due attention to the stochastic disturbances appearing therein. If in models such as the Koyck or adaptive expectations model OLS cannot be straightforwardly applied, methods need to be devised to resolve the estimation problem. Several alternative estimation methods are available although some of them may be computationally tedious. In the following section we consider one such method.

17.9 THE METHOD OF INSTRUMENTAL VARIABLES (IV)

The reason why OLS cannot be applied to the Koyck or adaptive expectations model is that the explanatory variable Y_{t-1} tends to be correlated with the error term v_t. If somehow this correlation can be removed, one can apply OLS to obtain consistent estimates, as noted previously. (*Note:* There will be some small sample bias.) How can this be accomplished? Liviatan has proposed the following solution.[32]

Let us suppose that we find a *proxy* for Y_{t-1} that is highly correlated with Y_{t-1} but is uncorrelated with v_t, where v_t is the error term appearing in the Koyck or adaptive expectations model. Such a proxy is called an **instrumental variable (IV).**[33] Liviatan suggests X_{t-1} as the instrumental variable for Y_{t-1} and further suggests that the parameters of the regression (17.8.1) can be obtained by solving the following normal equations:

$$\begin{aligned}\sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1 \sum X_t + \hat{\alpha}_2 \sum Y_{t-1}\\ \sum Y_t X_t &= \hat{\alpha}_0 \sum X_t + \hat{\alpha}_1 \sum X_t^2 + \hat{\alpha}_2 \sum Y_{t-1} X_t \qquad \textbf{(17.9.1)}\\ \sum Y_t X_{t-1} &= \hat{\alpha}_0 \sum X_{t-1} + \hat{\alpha}_1 \sum X_t X_{t-1} + \hat{\alpha}_2 \sum Y_{t-1} X_{t-1}\end{aligned}$$

[31]Also, as J. Johnston notes (op. cit., p. 350), "[the] pattern of adjustment [suggested by the partial adjustment model] . . . may sometimes be implausible."

[32]N. Liviatan, "Consistent Estimation of Distributed Lags," *International Economic Review,* vol. 4, January 1963, pp. 44–52.

[33]Such instrumental variables are used frequently in simultaneous equation models (see Chap. 20).

Notice that if we were to apply OLS directly to (17.8.1), the usual OLS normal equations would be (see Section 7.4)

$$\begin{aligned} \sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1 \sum X_t + \hat{\alpha}_2 \sum Y_{t-1} \\ \sum Y_t X_t &= \hat{\alpha}_0 \sum X_t + \hat{\alpha}_1 \sum X_t^2 + \hat{\alpha}_2 \sum Y_{t-1} X_t \\ \sum Y_t Y_{t-1} &= \hat{\alpha}_0 \sum Y_{t-1} + \hat{\alpha}_1 \sum X_t Y_{t-1} + \hat{\alpha}_2 \sum Y_{t-1}^2 \end{aligned} \tag{17.9.2}$$

The difference between the two sets of normal equations should be readily apparent. Liviatan has shown that the α's estimated from (17.9.1) are consistent, whereas those estimated from (17.9.2) may not be consistent because Y_{t-1} and $v_t\,[= u_t - \lambda u_{t-1}$ or $u_t - (1-\gamma)u_{t-1}]$ may be correlated whereas X_t and X_{t-1} are uncorrelated with v_t. (Why?)

Although easy to apply in practice once a suitable proxy is found, the Liviatan technique is likely to suffer from the multicollinearity problem because X_t and X_{t-1}, which enter in the normal equations of (17.9.1), are likely to be highly correlated (as noted in Chapter 12, most economic time series typically exhibit a high degree of correlation between successive values). The implication, then, is that although the Liviatan procedure yields consistent estimates, the estimators are likely to be inefficient.[34]

Before we move on, the obvious question is: How does one find a "good" proxy for Y_{t-1} in such a way that, although highly correlated with Y_{t-1}, it is uncorrelated with v_t? There are some suggestions in the literature, which we take up by way of an exercise (see exercise 17.5). But it must be stated that finding good proxies is not always easy, in which case the IV method is of little practical use and one may have to resort to maximum likelihood estimation techniques, which are beyond the scope of this book.[35]

Is there a test one can use to find out if the chosen instrument(s) is valid? Dennis Sargan has developed a test, dubbed the **SARG test,** for this purpose. The test is described in Appendix 17A, Section 17A.1.

17.10 DETECTING AUTOCORRELATION IN AUTOREGRESSIVE MODELS: DURBIN *h* TEST

As we have seen, the likely serial correlation in the errors v_t make the estimation problem in the autoregressive model rather complex: In the stock adjustment model the error term v_t did not have (first-order) serial correlation if the error term u_t in the original model was serially uncorrelated, whereas in the Koyck and adaptive expectation models v_t was serially

[34]To see how the efficiency of the estimators can be improved, consult Lawrence R. Klien, *A Textbook of Econometrics,* 2d ed., Prentice-Hall, Englewood Cliffs, N.J., 1974, p. 99. See also William H. Greene, *Econometric Analysis,* Macmillan, 2d ed., New York, 1993, pp. 535–538.

[35]For a condensed discussion of the ML methods, see J. Johnston, op. cit., pp. 366–371, as well as App. 4A and App. 15A.

correlated even if u_t was serially independent. The question, then, is: How does one know if there is serial correlation in the error term appearing in the autoregressive models?

As noted in Chapter 12, the Durbin–Watson d statistic may not be used to detect (first-order) serial correlation in autoregressive models, because the computed d value in such models generally tends toward 2, which is the value of d expected in a truly random sequence. In other words, if we routinely compute the d statistic for such models, there is a built-in bias against discovering (first-order) serial correlation. Despite this, many researchers compute the d value for want of anything better. Recently, however, Durbin himself has proposed a *large-sample* test of first-order serial correlation in autoregressive models.[36] This test is called the ***h* statistic.**

We have already discussed the Durbin *h test* in exercise 12.36. For convenience, we reproduce the h statistic (with a slight change in notation):

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\operatorname{var}(\hat{\alpha}_2)]}} \tag{17.10.1}$$

where n is the sample size, $\operatorname{var}(\hat{\alpha}_2)$ is the variance of the lagged $Y_t\ (= Y_{t-1})$ coefficient in (17.8.1), and $\hat{\rho}$ is an estimate of the first-order serial correlation ρ, first discussed in Chapter 12.

As noted in exercise 12.36, for large sample, Durbin has shown that, under the null hypothesis that $\rho = 0$, the h statistic of (17.10.1) follows the standard normal distribution. That is,

$$h_{\text{asy}} \sim N(0, 1) \tag{17.10.2}$$

where asy means asymptotically.

In practice, as noted in Chapter 12, one can estimate ρ as

$$\hat{\rho} \approx 1 - \frac{d}{2} \tag{17.10.3}$$

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

Let us illustrate the use of the h statistic with our Example 17.7. In this example, $n = 30$, $\hat{\rho} \approx (1 - d/2) = 0.4972$ (*note:* $d = 1.0056$), and $\operatorname{var}(\hat{\alpha}_2) = \operatorname{var}(\text{PPCE}_{t-1}) = (0.1546)^2 = 0.0239$. Putting these values in (17.10.1), we obtain:

$$h = 0.4972\sqrt{\frac{30}{1 - 30(0.0239)}} = 5.1191 \tag{17.10.4}$$

[36] J. Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Econometrica,* vol. 38, 1970, pp. 410–421.

Since this h value has the standard normal distribution under the null hypothesis, the probability of obtaining such a high h value is very small. Recall that the probability that a standard normal variate exceeds the value of ± 3 is extremely small. In the present example our conclusion, then, is that there is (positive) autocorrelation. Of course, bear in mind that h follows the standard normal distribution asymptotically. Our sample of 30 observations may not be necessarily large.

Note these features of the *h statistic*.

1. It does not matter how many X variables or how many lagged values of Y are included in the regression model. To compute h, we need consider only the variance of the coefficient of lagged Y_{t-1}.

2. The test is not applicable if $[n \operatorname{var}(\hat{\alpha}_2)]$ exceeds 1. (Why?) In practice, though, this does not usually happen.

3. Since the test is a large-sample test, its application in small samples is not strictly justified, as shown by Inder[37] and Kiviet.[38] It has been suggested that the Breusch–Godfrey (BG) test, also known as the Lagrange multiplier test, discussed in Chapter 12 is statistically more powerful not only in the large samples but also in finite, or small, samples and is therefore preferable to the h test.[39]

17.11 A NUMERICAL EXAMPLE: THE DEMAND FOR MONEY IN CANADA, 1979–I TO 1988–IV

To illustrate the use of the models we have discussed thus far, consider one of the earlier empirical applications, namely, the demand for money (or real cash balances). In particular, consider the following model.[40]

$$M_t^* = \beta_0 R_t^{\beta_1} Y_t^{\beta_2} e^{u_t} \tag{17.11.1}$$

where M_t^* = desired, or long-run, demand for money (real cash balances)
R_t = long-term interest rate, %
Y_t = aggregate real national income

For statistical estimation, (17.11.1) may be expressed conveniently in log form as

$$\ln M_t^* = \ln \beta_0 + \beta_1 \ln R_t + \beta_2 \ln Y_t + u_t \tag{17.11.2}$$

[37] B. Inder, "An Approximation to the Null Distribution of the Durbin–Watson Statistic in Models Containing Lagged Dependent Variables," *Econometric Theory,* vol. 2, no. 3, 1986, pp. 413–428.

[38] J. F. Kiviet, "On the Vigour of Some Misspecification Tests for Modelling Dynamic Relationships," *Review of Economic Studies,* vol. 53, no. 173, 1986, pp. 241–262.

[39] Gabor Korosi, Laszlo Matyas, and Istvan P. Szekely, *Practical Econometrics,* Ashgate Publishing Company, Brookfield, Vermont, 1992, p. 92.

[40] For a similar model, see Gregory C. Chow, "On the Long-Run and Short-Run Demand for Money," *Journal of Political Economy,* vol. 74, no. 2, 1966, pp. 111–131. Note that one advantage of the multiplicative function is that the exponents of the variables give direct estimates of elasticities (see Chap. 6).

Since the desired demand variable is not directly observable, let us assume the stock adjustment hypothesis, namely,

$$\frac{M_t}{M_{t-1}} = \left(\frac{M_t^*}{M_{t-1}}\right)^{\delta} \qquad 0 < \delta \le 1 \tag{17.11.3}$$

Equation (17.11.3) states that a constant percentage (why?) of the discrepancy between the actual and desired real cash balances is eliminated within a single period (year). In log form, Eq. (17.11.3) may be expressed as

$$\ln M_t - \ln M_{t-1} = \delta(\ln M_t^* - \ln M_{t-1}) \tag{17.11.4}$$

Substituting $\ln M_t^*$ from (17.11.2) into Eq. (17.11.4) and rearranging, we obtain

$$\ln M_t = \delta \ln \beta_0 + \beta_1 \delta \ln R_t + \beta_2 \delta \ln Y_t + (1-\delta) \ln M_{t-1} + \delta u_t \qquad \textbf{(17.11.5)}^{41}$$

which may be called the *short-run demand function* for money. (Why?)

As an illustration of the short-term and long-term demand for real cash balances, consider the data given in Table 17.3. These quarterly data pertain to Canada for the period 1979 to 1988. The variables are defined as follows: M [as defined by M1 money supply, Canadian dollars (C\$), millions], P (implicit price deflator, 1981 = 100), GDP at constant 1981 prices (C\$, millions) and R (90-day prime corporate rate of interest, %).[42] M1 was deflated by P to obtain figures for real cash balances. A priori, real money demand is expected to be positively related to GDP (positive income effect) and negatively related to R (the higher the interest rate, the higher the opportunity cost of holding money, as M1 money pays very little interest, if any).

The regression results were as follows[43]:

$$\begin{aligned}
\widehat{\ln M_t} &= 0.8561 - 0.0634 \ln R_t - 0.0237 \ln \text{GDP}_t + 0.9607 \ln M_{t-1}\\
\text{se} &= (0.5101) \quad (0.0131) \quad (0.0366) \quad (0.0414)\\
t &= (1.6782) \quad (-4.8134) \quad (-0.6466) \quad (23.1972)\\
& R^2 = 0.9482 \qquad d = 2.4582 \qquad F = 213.7234
\end{aligned} \qquad \textbf{(17.11.6)}^{43}$$

[41]In passing, note that this model is essentially nonlinear in the parameters. Therefore, although OLS may give an unbiased estimate of, say, $\beta_1\delta$ taken together, it may not give unbiased estimates of β_1 and δ individually, especially if the sample is small.

[42]These data are obtained from B. Bhaskar Rao, ed., *Cointegration for the Applied Economist,* St. Martin's Press, New York, 1994, pp. 210–213. The original data is from 1956-I to 1988-IV, but for illustration purposes we begin our analysis from the first quarter of 1979.

[43]Note this feature of the estimated standard errors. The standard error of, say, the coefficient of $\ln R_t$ refers to the standard error of $\widehat{\beta_1\delta}$, an estimator of $\beta_1\delta$. There is no simple way to obtain the standard errors of $\hat{\beta}_1$ and $\hat{\delta}$ individually from the standard error of $\widehat{\beta_1\delta}$, especially if the sample is relatively small. For large samples, however, individual standard errors of $\hat{\beta}_1$ and $\hat{\delta}$ can be obtained approximately, but the computations are involved. See Jan Kmenta, *Elements of Econometrics,* Macmillan, New York, 1971, p. 444.

TABLE 17.3 MONEY, INTEREST RATE, PRICE INDEX, AND GDP, CANADA

Observation	M1	*R*	*P*	GDP
1979–1	22,175.00	11.13333	0.77947	334,800
1979–2	22,841.00	11.16667	0.80861	336,708
1979–3	23,461.00	11.80000	0.82649	340,096
1979–4	23,427.00	14.18333	0.84863	341,844
1980–1	23,811.00	14.38333	0.86693	342,776
1980–2	23,612.33	12.98333	0.88950	342,264
1980–3	24,543.00	10.71667	0.91553	340,716
1980–4	25,638.66	14.53333	0.93743	347,780
1981–1	25,316.00	17.13333	0.96523	354,836
1981–2	25,501.33	18.56667	0.98774	359,352
1981–3	25,382.33	21.01666	1.01314	356,152
1981–4	24,753.00	16.61665	1.03410	353,636
1982–1	25,094.33	15.35000	1.05743	349,568
1982–2	25,253.66	16.04999	1.07748	345,284
1982–3	24,936.66	14.31667	1.09666	343,028
1982–4	25,553.00	10.88333	1.11641	340,292
1983–1	26,755.33	9.616670	1.12303	346,072
1983–2	27,412.00	9.316670	1.13395	353,860
1983–3	28,403.33	9.333330	1.14721	359,544
1983–4	28,402.33	9.550000	1.16059	362,304
1984–1	28,715.66	10.08333	1.17117	368,280
1984–2	28,996.33	11.45000	1.17406	376,768
1984–3	28,479.33	12.45000	1.17795	381,016
1984–4	28,669.00	10.76667	1.18438	385,396
1985–1	29,018.66	10.51667	1.18990	390,240
1985–2	29,398.66	9.666670	1.20625	391,580
1985–3	30,203.66	9.033330	1.21492	396,384
1985–4	31,059.33	9.016670	1.21805	405,308
1986–1	30,745.33	11.03333	1.22408	405,680
1986–2	30,477.66	8.733330	1.22856	408,116
1986–3	31,563.66	8.466670	1.23916	409,160
1986–4	32,800.66	8.400000	1.25368	409,616
1987–1	33,958.33	7.250000	1.27117	416,484
1987–2	35,795.66	8.300000	1.28429	422,916
1987–3	35,878.66	9.300000	1.29599	429,980
1987–4	36,336.00	8.700000	1.31001	436,264
1988–1	36,480.33	8.616670	1.32325	440,592
1988–2	37,108.66	9.133330	1.33219	446,680
1988–3	38,423.00	10.05000	1.35065	450,328
1988–4	38,480.66	10.83333	1.36648	453,516

Notes: M1 = C$, millions
P = implicit price deflator (1981 = 100)
R = 90-day prime corporate interest rate, %
GDP = C$, millions (1981 prices)
Source: Rao, op. cit., pp. 210–213.

The estimated short-run demand function shows that the short-run interest elasticity has the correct sign and that it is statistically quite significant, as its p value is almost zero. The short-run income elasticity is surprisingly negative, although statistically it is not different from zero. The coefficient of adjustment is $\delta = (1 - 0.9607) = 0.0393$, implying that only about 4 percent of the discrepancy between the desired and actual real cash balances is eliminated in a quarter, a rather slow adjustment.

To get back to the long-run demand function (17.11.2), all that needs to be done is to divide the short-run demand function through by δ (why?) and drop the $\ln M_{t-1}$ term. The results are:

$$\widehat{\ln M_t^*} = 21.7888 - 1.6132 \ln R_t - 0.6030 \ln \text{GDP} \qquad \textbf{(17.11.7)}$$[44]

As can be seen, the long-run interest elasticity of demand for money is substantially greater (in absolute terms) than the corresponding short-run elasticity, which is also true of the income elasticity, although in the present instance its economic and statistical significance is dubious.

Note that the estimated Durbin–Watson d is 2.4582, which is close to 2. This substantiates our previous remark that in the autoregressive models the computed d is generally close to 2. Therefore, we should not trust the computed d to find out whether there was serial correlation in our data. The sample size in our case is 40 observations, which may be reasonably large to apply the h test. In the present case, the reader can verify that the estimated h value is -1.5008, which is not significant at the 5 percent level, perhaps suggesting that there is no first-order autocorrelation in the error term.

17.12 ILLUSTRATIVE EXAMPLES

In this section we present a few examples of distributed lag models to show how researchers have used them in empirical studies.

EXAMPLE 17.9

THE FED AND THE REAL RATE OF INTEREST

To assess the effect of M1 (currency + checkable deposits) growth on Aaa bond real interest rate measure, G. J. Santoni and Courtenay C. Stone[45] estimated, using monthly data, the following distributed lag model for the United States.

$$r_t = \text{constant} + \sum_{i=0}^{11} a_i \dot{M}_{t-i} + u_i \qquad \textbf{(17.12.1)}$$

where r_t = Moody's Index of Aaa bond yield minus the average annual rate of change in the seasonally adjusted consumer price index over the prior 36 months, which is used as the measure of real interest rate, and $\dot{M}_t$ = monthly M_1 growth.

According to the "neutrality of money doctrine," which states that real economic variables—such as output, employment, economic growth and the real rate of interest—are not influenced permanently by money growth and, therefore, are essentially unaffected by monetary policy. . . . Given this argument, the Federal

(*Continued*)

[44]Note that we have not presented the standard errors of the estimated coefficients for reasons discussed in footnote 43.

[45]"The Fed and the Real Rate of Interest," *Review,* Federal Reserve Bank of St. Louis, December 1982, pp. 8–18.

EXAMPLE 17.9 (*Continued*)

Reserve has no permanent influence over the real rate of interest whatsoever.[46]

If this doctrine is valid, then one should expect the distributed lag coefficients a_i as well as their sum to be statistically indifferent from zero. To find out whether this is the case, the authors estimated (17.12.1) for two different time periods, February 1951 to September 1979 and October 1979 to November 1982, the latter to take into account the change in the Fed's monetary policy, which since October 1979 has paid more attention to the rate of growth of the money supply than to the rate of interest, which was the policy in the earlier period. Their regression results are presented in Table 17.4. The results seem to support the "neutrality of money doctrine," since for the period February 1951 to September 1979 the current as well as lagged money growth had no statistically significant effect on the real interest rate measure. For the latter period, too, the neutrality doctrine seems to hold since $\sum a_i$ is not statistically different from zero; only the coefficient a_1 is significant, but it has the wrong sign. (Why?)

TABLE 17.4
INFLUENCE OF MONTHLY M1 GROWTH ON AN AAA BOND REAL INTEREST RATE MEASURE: FEBRUARY 1951 TO NOVEMBER 1982

$$r = \text{constant} + \sum_{i=0}^{11} a_i \dot{M}1_{t-1}$$

	February 1951 to September 1979		October 1979 to November 1982	
	Coefficient	$\lvert t\rvert$*	Coefficient	$\lvert t\rvert$
Constant	1.4885†	2.068	1.0360	0.801
a_0	−0.00088	0.388	0.00840	1.014
a_1	0.00171	0.510	0.03960†	3.419
a_2	0.00170	0.423	0.03112	2.003
a_3	0.00233	0.542	0.02719	1.502
a_4	−0.00249	0.553	0.00901	0.423
a_5	−0.00160	0.348	0.01940	0.863
a_6	0.00292	0.631	0.02411	1.056
a_7	0.00253	0.556	0.01446	0.666
a_8	0.00000	0.001	−0.00036	0.019
a_9	0.00074	0.181	−0.00499	0.301
a_{10}	0.00016	0.045	−0.01126	0.888
a_{11}	0.00025	0.107	−0.00178	0.211
$\sum a_i$	0.00737	0.221	0.1549	0.926
$\bar{R}^2$	0.9826		0.8662	
D-W	2.07		2.04	
RH01	1.27†	24.536	1.40†	9.838
RH02	−0.28†	5.410	−0.48†	3.373
NOB	344.		38.	
SER (= RSS)	0.1548		0.3899	

*$\lvert t\rvert$ = absolute t value.
†Significantly different from zero at the 0.05 level.
Source: G. J. Santoni and Courtenay C. Stone, "The Fed and the Real Rate of Interest," *Review,* Federal Reserve Bank of St. Louis, December 1982, p. 16.

[46]"The Fed and the Real Rate of Interest," *Review,* Federal Reserve Bank of St. Louis, December 1982, p. 15.

EXAMPLE 17.10

THE SHORT- AND LONG-RUN AGGREGATE CONSUMPTION FOR SRI LANKA, 1967–1993

Suppose consumption C is linearly related to permanent income X^*:

$$C_t = \beta_1 + \beta_2 X_t^* + u_t \quad \textbf{(17.12.2)}$$

Since X_t^* is not directly observable, we need to specify the mechanism that generates permanent income. Suppose we adopt the adaptive expectations hypothesis specified in (17.5.2). Using (17.5.2) and simplifying, we obtain the following estimating equation (cf. 17.5.5):

$$C_t = \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \quad \textbf{(17.12.3)}$$

where $\alpha_1 = \gamma\beta_1$
$\alpha_2 = \gamma\beta_2$
$\alpha_3 = (1 - \gamma)$
$v_t = [u_t - (1 - \gamma)u_{t-1}]$

As we know, β_2 gives the mean response of consumption to, say, a \$1 increase in permanent income, whereas α_2 gives the mean response of consumption to a \$1 increase in current income.

From annual data for Sri Lanka for the period 1967–1993 given in Table 17.5, the following regression results were obtained[47]:

$$\begin{aligned}\hat{C} = \;& 1038.403 \;+\; 0.4043X_t + \;0.5009C_{t-1}\\ \text{se} = \;& (2501.455) \quad (0.0919) \quad (0.1213) \qquad \textbf{(17.12.4)}\\ t = \;& (0.4151) \quad (4.3979) \quad (4.1293)\\ & R^2 = 0.9912 \quad d = 1.4162 \quad F = 1298.466\end{aligned}$$

where C = private consumption expenditure, and X = GDP, both at constant prices. We also introduced real interest rate in the model, but it was not statistically significant.

The results show that the short-run marginal propensity to consume (MPC) is 0.4043, suggesting that a 1 rupee increase in the current or observed real income (as measured by real GDP) would increase mean consumption by about 0.40 rupee. But if the increase in income is sustained, then eventually the MPC out of the permanent income will be $\beta_2 = \gamma\beta_2/\gamma = 0.4043/0.4991 = 0.8100$ or about 0.81 rupee. In other words, when consumers have had time to adjust to the 1 rupee change in income, they will increase their consumption ultimately by about 0.81 rupee.

TABLE 17.5 PRIVATE CONSUMPTION EXPENDITURE AND GDP, SRI LANKA

Observation	PCON	GDP	Observation	PCON	GDP
1967	61,284	78,221	1981	120,477	152,846
1968	68,814	83,326	1982	133,868	164,318
1969	76,766	90,490	1983	148,004	172,414
1970	73,576	92,692	1984	149,735	178,433
1971	73,256	94,814	1985	155,200	185,753
1972	67,502	92,590	1986	154,165	192,059
1973	78,832	101,419	1987	155,445	191,288
1974	80,240	105,267	1988	157,199	196,055
1975	84,477	112,149	1989	158,576	202,477
1976	86,038	116,078	1990	169,238	223,225
1977	96,275	122,040	1991	179,001	233,231
1978	101,292	128,578	1992	183,687	242,762
1979	105,448	136,851	1993	198,273	259,555
1980	114,570	144,734			

Notes: PCON = private consumption expenditure.
GDP = gross domestic product.
Source: See footnote 47.

(*Continued*)

[47]The data are obtained from the data disk in Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries,* Routledge, New York, 1998. The original data is from World Bank's World Tables.

EXAMPLE 17.10 (*Continued*)

Now suppose that our consumption function were

$$C_t^* = \beta_1 + \beta_2 X_t + u_t \tag{17.12.5}$$

In this formulation permanent or long-run consumption C_t is a linear function of the current or observed income. Since C_t^* is not directly observable, let us invoke the partial adjustment model (17.6.2). Using this model, and after algebraic manipulations, we obtain

$$\begin{aligned} C_t &= \delta\beta_1 + \delta\beta_2 X_t + (1-\delta)C_{t-1} + \delta u_t \\ &= \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \end{aligned} \tag{17.12.6}$$

In appearance, this model is indistinguishable from the adaptive expectations model (17.12.3). Therefore, the regression results given in (17.12.4) are equally applicable here. However, there is a major difference in the interpretation of the two models, not to mention the estimation problem associated with the autoregressive and possibly serially correlated model (17.12.3). The model (17.12.5) is the long-run, or equilibrium, consumption function, whereas (17.12.6) is the short-run consumption function. β_2 measures the long-run MPC, whereas $\alpha_2\ (= \delta\beta_2)$ gives the short-run MPC; the former can be obtained from the latter by dividing it by δ, the coefficient of adjustment.

Returning to (17.12.4), we can now interpret 0.4043 as the short-run MPC. Since $\delta = 0.4991$, the long-run MPC is 0.81. Note that the adjustment coefficient of about 0.50 suggests that in any given time period consumers only adjust their consumption one-half of the way toward its desired or long-run level.

This example brings out the crucial point that in appearance the adaptive expectations and the partial adjustment models, or the Koyck model for that matter, are so similar that by just looking at the estimated regression, such as (17.12.4), one cannot tell which is the correct specification. That is why it is so vital that one specify the theoretical underpinning of the model chosen for empirical analysis and then proceed appropriately. If habit or inertia characterizes consumption behavior, then the partial adjustment model is appropriate. On the other hand, if consumption behavior is forward-looking in the sense that it is based on expected future income, then the adaptive expectations model is appropriate. If it is the latter, then, one will have to pay close attention to the estimation problem to obtain consistent estimators. In the former case, the OLS will provide consistent estimators, provided the usual OLS assumptions are fulfilled.

17.13 THE ALMON APPROACH TO DISTRIBUTED-LAG MODELS: THE ALMON OR POLYNOMIAL DISTRIBUTED LAG (PDL)[48]

Although used extensively in practice, the Koyck distributed-lag model is based on the assumption that the β coefficients decline geometrically as the lag lengthens (see Figure 17.5). This assumption may be too restrictive in some situations. Consider, for example, Figure 17.7.

In Figure 17.7*a* it is assumed that the β's increase at first and then decrease, whereas in Figure 17.7*c* it is assumed that they follow a cyclical pattern. Obviously, the Koyck scheme of distributed-lag models will not work in these cases. However, after looking at Figure 17.7*a* and *c*, it seems that one can express β_i as a function of i, the length of the lag (time), and fit suitable curves to reflect the functional relationship between the two, as indicated in Figure 17.7*b* and *d*. This approach is precisely the one suggested by Shirley Almon. To illustrate her technique, let us revert to the finite distributed-lag model considered previously, namely,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \tag{17.1.2}$$

[48]Shirley Almon, "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica,* vol. 33, January 1965, pp. 178–196.

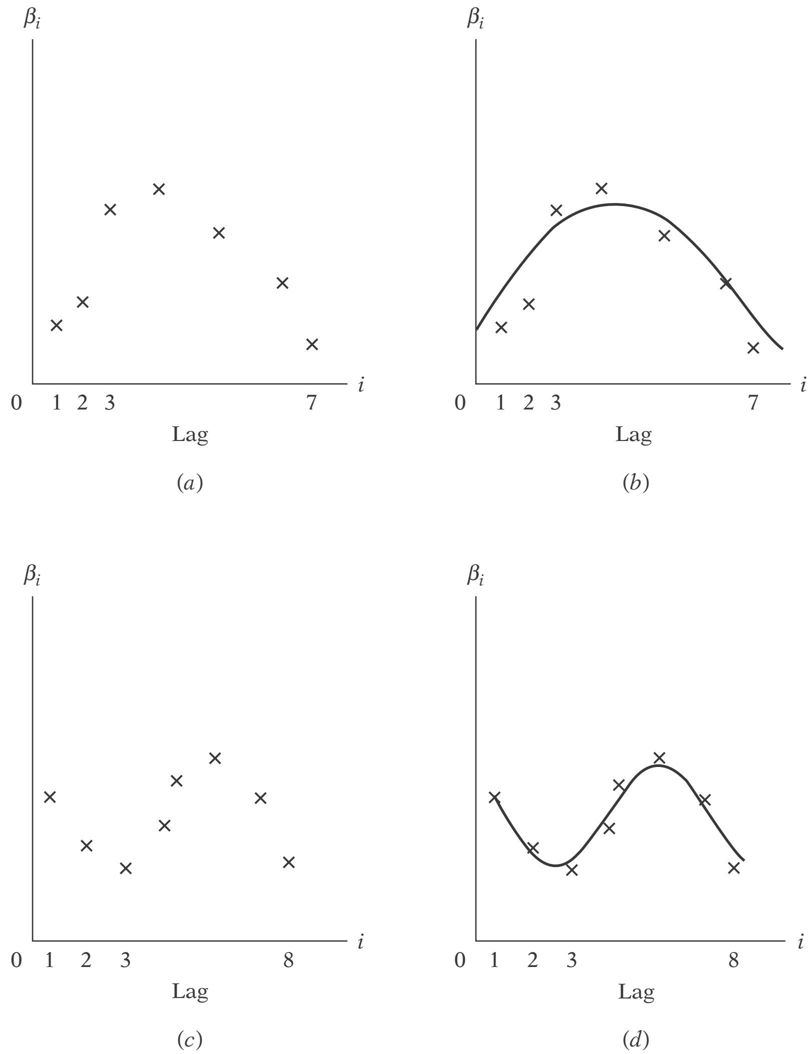

FIGURE 17.7 Almon polynomial-lag scheme.

which may be written more compactly as

$$Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t \quad \textbf{(17.13.1)}$$

Following a theorem in mathematics known as **Weierstrass' theorem,** Almon assumes that β_i can be approximated by a suitable-degree polynomial in i, the length of the lag.[49] For instance, if the lag scheme shown in

[49]Broadly speaking, the theorem states that on a finite closed interval any continuous function may be approximated uniformly by a polynomial of a suitable degree.

Figure 17.7*a* applies, we can write

$$\beta_i = a_0 + a_1 i + a_2 i^2 \tag{17.13.2}$$

which is a quadratic, or second-degree, polynomial in i (see Figure 17.7*b*). However, if the β's follow the pattern of Figure 17.7*c*, we can write

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 \tag{17.13.3}$$

which is a third-degree polynomial in i (see Figure 17.7*d*). More generally, we may write

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m \tag{17.13.4}$$

which is an mth-degree polynomial in i. It is assumed that m (the degree of the polynomial) is less than k (the maximum length of the lag).

To explain how the Almon scheme works, let us assume that the β's follow the pattern shown in Figure 17.7*a* and, therefore, the second-degree polynomial approximation is appropriate. Substituting (17.13.2) into (17.13.1), we obtain

$$\begin{aligned} Y_t &= \alpha + \sum_{i=0}^{k}(a_0 + a_1 i + a_2 i^2)X_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^{k} X_{t-i} + a_1 \sum_{i=0}^{k} i X_{t-i} + a_2 \sum_{i=0}^{k} i^2 X_{t-i} + u_t \end{aligned} \tag{17.13.5}$$

Defining

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^{k} X_{t-i} \\ Z_{1t} &= \sum_{i=0}^{k} i X_{t-i} \\ Z_{2t} &= \sum_{i=0}^{k} i^2 X_{t-i} \end{aligned} \tag{17.13.6}$$

we may write (17.13.5) as

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t \tag{17.13.7}$$

In the Almon scheme Y is regressed on the constructed variables Z, not the original X variables. Note that (17.13.7) can be estimated by the usual OLS procedure. The estimates of α and a_i thus obtained will have all the desirable statistical properties provided the stochastic disturbance term u satisfies the assumptions of the classical linear regression model. In this respect, the Almon technique has a distinct advantage over the Koyck method because, as we have seen, the latter has some serious estimation problems that result from the presence of the stochastic explanatory variable Y_{t-1} and its likely correlation with the disturbance term.

Once the a's are estimated from (17.13.7), the original β's can be estimated from (17.13.2) [or more generally from (17.13.4)] as follows:

$$\begin{aligned} \hat{\beta}_0 &= \hat{a}_0 \\ \hat{\beta}_1 &= \hat{a}_0 + \hat{a}_1 + \hat{a}_2 \\ \hat{\beta}_2 &= \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 \\ \hat{\beta}_3 &= \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 \\ &\cdots\cdots\cdots\cdots \\ \hat{\beta}_k &= \hat{a}_0 + k\hat{a}_1 + k^2\hat{a}_2 \end{aligned} \tag{17.13.8}$$

Before we apply the Almon technique, we must resolve the following practical problems.

1. The maximum length of the lag k must be specified in advance. Here perhaps one can follow the advice of Davidson and MacKinnon:

> The best approach is probably to settle the question of lag length first, by starting with a very large value of q [the lag length] and then seeing whether the fit of the model deteriorates significantly when it is reduced without imposing any restrictions on the shape of the distributed lag.[50]

This advice is in the spirit of Hendry's top-down approach discussed in Chapter 13. Remember that if there is some "true" lag length, choosing fewer lags will lead to the "omission of relevant variable bias," whose consequences, as we saw in Chapter 13, can be very serious. On the other hand, choosing more lags than necessary will lead to the "inclusion of irrelevant variable bias," whose consequences are less serious; the coefficients can be consistently estimated by OLS, although their variances may be less efficient.

One can use the *Akaike* or *Schwarz information criterion* discussed in Chapter 13 to choose the appropriate lag length. These criteria can also be used to discuss the appropriate degree of the polynomial in addition to the discussion in point 2.

2. Having specified k, we must also specify the degree of the polynomial m. Generally, the degree of the polynomial should be at least one more than the number of turning points in the curve relating β_i to i. Thus, in Figure 17.7*a* there is only one turning point; hence a second-degree polynomial will be a good approximation. In Figure 17.7*c* there are two turning points; hence a third-degree polynomial will provide a good approximation. A priori, however, one may not know the number of turning points, and therefore, the choice of m is largely subjective. However, theory may suggest a particular shape in some cases. In practice, one hopes that a fairly low-degree polynomial (say, $m = 2$ or 3) will give good results. Having chosen a particular value of m, if we want to find out whether a higher-degree polynomial will give a better fit, we can proceed as follows.

---

[50]Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics,* Oxford University Press, New York, 1993, pp. 675–676.

Suppose we must decide between the second- and third-degree polynomials. For the second-degree polynomial the estimating equation is as given by (17.13.7). For the third-degree polynomial the corresponding equation is

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + a_3 Z_{3t} + u_t \qquad \textbf{(17.13.9)}$$

where $Z_{3t} = \sum_{i=0}^{k} i^3 X_{t-i}$. After running regression (17.13.9), if we find that a_2 is statistically significant but a_3 is not, we may assume that the second-degree polynomial provides a reasonably good approximation.

Alternatively, as Davidson and MacKinnon suggest, "After q [the lag length] is determined, one can then attempt to determine d [the degree of the polynomial] once again starting with a large value and then reducing it."

However, we must beware of the problem of multicollinearity, which is likely to arise because of the way the Z's are constructed from the X's, as shown in (17.13.6) [see also (17.13.10)]. As shown in Chapter 10, in cases of serious multicollinearity, $\hat{a}_3$ may turn out to be statistically insignificant, not because the true a_3 is zero, but simply because the sample at hand does not allow us to assess the separate impact of Z_3 on Y. Therefore, in our illustration, before we accept the conclusion that the third-degree polynomial is not the correct choice, we must make sure that the multicollinearity problem is not serious enough, which can be done by applying the techniques discussed in Chapter 10.

3. Once m and k are specified, the Z's can be readily constructed. For instance, if $m = 2$ and $k = 5$, the Z's are

$$\begin{aligned}
Z_{0t} &= \sum_{i=0}^{5} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5}) \\
Z_{1t} &= \sum_{i=0}^{5} i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5}) \qquad \textbf{(17.13.10)} \\
Z_{2t} &= \sum_{i=0}^{5} i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5})
\end{aligned}$$

Notice that the Z's are linear combinations of the original X's. Also notice why the Z's are likely to exhibit multicollinearity.

Before proceeding to a numerical example, note the advantages of the Almon method. First, it provides a flexible method of incorporating a variety of lag structures (see exercise 17.17). The Koyck technique, on the other hand, is quite rigid in that it assumes that the β's decline geometrically. Second, unlike the Koyck technique, in the Almon method we do not have to worry about the presence of the lagged dependent variable as an explanatory variable in the model and the problems it creates for estimation. Finally, if a sufficiently low-degree polynomial can be fitted, the number of coefficients to be estimated (the a's) is considerably smaller than the original number of coefficients (the β's).

But let us re-emphasize the problems with the Almon technique. First, the degree of the polynomial as well as the maximum value of the lag is largely a subjective decision. Second, for reasons noted previously, the Z variables are likely to exhibit multicollinearity. Therefore, in models like (17.13.9) the estimated a's are likely to show large standard errors (relative to the values of these coefficients), thereby rendering one or more such coefficients statistically insignificant on the basis of the conventional t test. But this does not necessarily mean that one or more of the original $\hat{\beta}$ coefficients will also be statistically insignificant. (The proof of this statement is slightly involved but is suggested in exercise 17.18.) As a result, the multicollinearity problem may not be as serious as one might think. Besides, as we know, in cases of multicollinearity even if we cannot estimate an individual coefficient precisely, a linear combination of such coefficients (the **estimable function**) can be estimated more precisely.

EXAMPLE 17.11

ILLUSTRATION OF THE ALMON DISTRIBUTED-LAG MODEL

To illustrate the Almon technique, Table 17.6 gives data on inventories Y and sales X for the United States for the period 1954–1999.

For illustrative purposes, assume that inventories depend on sales in the current year and in the preceding 3 years as follows:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t \qquad \textbf{(17.13.11)}$$

Furthermore, assume that β_i can be approximated by a second-degree polynomial as shown in (17.13.2). Then, following (17.13.5), we may write

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t \qquad \textbf{(17.13.12)}$$

where

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^{3} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3}) \\ Z_{1t} &= \sum_{i=0}^{3} i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3}) \\ Z_{2t} &= \sum_{i=0}^{3} i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3}) \end{aligned} \qquad \textbf{(17.13.13)}$$

The Z variables thus constructed are shown in Table 17.6. Using the data on Y and the Z's, we obtain the following regression:

$$\begin{aligned} \hat{Y}_t &= 25{,}845.06 \quad + \ 1.1149Z_{0t} - \ 0.3713Z_{1t} - \ 0.0600Z_{2t} \\ \text{se} &= (6596.998) \quad (0.5381) \quad (1.3743) \quad (0.4549) \\ t &= (3.9177) \quad (2.0718) \quad (-0.2702) \quad (-0.1319) \\ & R^2 = 0.9755 \quad d = 0.1643 \quad F = 517.7656 \end{aligned} \qquad \textbf{(17.13.14)}$$

Note: Since we are using a 3-year lag, the total number of observations has been reduced from 46 to 43.

(*Continued*)

EXAMPLE 17.11 (*Continued*)

TABLE 17.6
INVENTORIES *Y* AND SALES *X*, U.S. MANUFACTURING, AND CONSTRUCTED *Z*'S

Observation	Inventory	Sales	Z_0	Z_1	Z_2
1954	41,612	23,355	NA	NA	NA
1955	45,069	26,480	NA	NA	NA
1956	50,642	27,740	NA	NA	NA
1957	51,871	28,736	106,311	150,765	343,855
1958	50,203	27,248	110,204	163,656	378,016
1959	52,913	30,286	114,010	167,940	391,852
1960	53,786	30,878	117,148	170,990	397,902
1961	54,871	30,922	119,334	173,194	397,254
1962	58,172	33,358	125,444	183,536	427,008
1963	60,029	35,058	130,216	187,836	434,948
1964	63,410	37,331	136,669	194,540	446,788
1965	68,207	40,995	146,742	207,521	477,785
1966	77,986	44,870	158,254	220,831	505,841
1967	84,646	46,486	169,682	238,853	544,829
1968	90,560	50,229	182,580	259,211	594,921
1969	98,145	53,501	195,086	277,811	640,003
1970	101,599	52,805	203,021	293,417	672,791
1971	102,567	55,906	212,441	310,494	718,870
1972	108,121	63,027	225,239	322,019	748,635
1973	124,499	72,931	244,669	333,254	761,896
1974	157,625	84,790	276,654	366,703	828,193
1975	159,708	86,589	307,337	419,733	943,757
1976	174,636	98,797	343,107	474,962	1,082,128
1977	188,378	113,201	383,377	526,345	1,208,263
1978	211,691	126,905	425,492	570,562	1,287,690
1979	242,157	143,936	482,839	649,698	1,468,882
1980	265,215	154,391	538,433	737,349	1,670,365
1981	283,413	168,129	593,361	822,978	1,872,280
1982	311,852	163,351	629,807	908,719	2,081,117
1983	312,379	172,547	658,418	962,782	2,225,386
1984	339,516	190,682	694,709	1,003,636	2,339,112
1985	334,749	194,538	721,118	1,025,829	2,351,029
1986	322,654	194,657	752,424	1,093,543	2,510,189
1987	338,109	206,326	786,203	1,155,779	2,688,947
1988	369,374	224,619	820,140	1,179,254	2,735,796
1989	391,212	236,698	862,300	1,221,242	2,801,836
1990	405,073	242,686	910,329	1,304,914	2,992,108
1991	390,905	239,847	943,850	1,389,939	3,211,049
1992	382,510	250,394	969,625	1,435,313	3,340,873
1993	384,039	260,635	993,562	1,458,146	3,393,956
1994	404,877	279,002	1,029,878	1,480,964	3,420,834
1995	430,985	299,555	1,089,586	1,551,454	3,575,088
1996	436,729	309,622	1,148,814	1,639,464	3,761,278
1997	456,133	327,452	1,215,631	1,745,738	4,018,860
1998	466,798	337,687	1,274,316	1,845,361	4,261,935
1999	470,377	354,961	1,329,722	1,921,457	4,434,093

Note: Y and *X* are in millions of dollars, seasonally adjusted.
Source: Economic Report of the President, 2001, Table B-57, p. 340. The *Z*'s are as shown in (17.13.13).

(*Continued*)

EXAMPLE 17.11 (*Continued*)

A brief comment on the preceding results. Of the three *Z* variables, only Z_0 is individually statistical significant at the 5 percent level, but the others are not, yet the *F* value is so high that we can reject the null hypothesis that collectively the *Z*'s have no effect on *Y*. As you can suspect, this might very well be due to multicollinearity. Also, note that the computed *d* value is very low. This does not necessarily mean that the residuals suffer from autocorrelation. More likely, the low *d* value suggests that the model we have used is probably mis-specified. We will comment on this shortly.

From the estimated *a*'s given in (17.13.3), we can easily estimate the original β's easily, as shown in (17.13.8). In the present example, the results are as follows:

$$\begin{aligned} \hat\beta_0 &= \hat a_0 = 1.1149 \\ \hat\beta_1 &= (\hat a_0 + \hat a_1 + \hat a_2) = 0.6836 \\ \hat\beta_2 &= (\hat a_0 + 2\hat a_1 + 4\hat a_2) = 0.1321 \\ \hat\beta_3 &= (\hat a_0 + 3\hat a_1 + 9\hat a_2) = -0.5394 \end{aligned} \tag{17.13.15}$$

Thus, the estimated distributed-lag model corresponding to (17.13.11) is:

$$\begin{array}{lllllll} \hat Y_t = 25{,}845.0 & & +\ 1.1150X_0 + & 0.6836X_{t-1} + & 0.1321X_{t-2} - & 0.5394X_{t-3} & \\ \text{se} = & (6596.99) & (0.5381) & (0.4672) & (0.4656) & (0.5656) & \textbf{(17.13.16)} \\ t = & (3.9177) & (2.0718) & (1.4630) & (0.2837) & (-0.9537) & \end{array}$$

Geometrically, the estimated β_i is as shown in Figure 17.8.

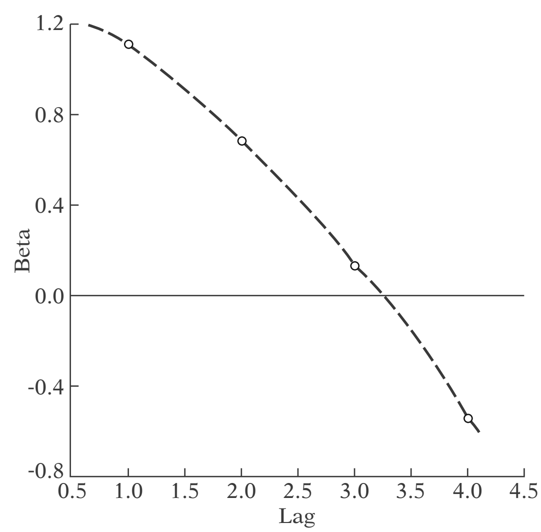

FIGURE 17.8 Lag structure of the illustrative example.

Our illustrative example may be used to point out a few additional features of the Almon lag procedure:

1. The standard errors of the *a* coefficients are directly obtainable from the OLS regression (17.13.14), but the standard errors of some of the $\hat\beta$

coefficients, the objective of primary interest, cannot be so obtained. But they can be obtained from the standard errors of the estimated *a* coefficients by using a well-known formula from statistics, which is given in exercise 17.18. Of course, there is no need to do this manually, for most statistical packages can do this routinely. The standard errors given in (17.13.15) were obtained from Eviews 4.

2. The $\hat{\beta}$'s obtained in (17.13.16) are called *unrestricted estimates* in the sense that no a priori restrictions are placed on them. In some situations, however, one may want to impose the so-called **endpoint restrictions** on the β's by assuming that β_0 and β_k (the current and kth lagged coefficient) are zero. Because of psychological, institutional, or technical reasons, the value of the explanatory variable in the current period may not have any impact on the current value of the regressand, thereby justifying the zero value for β_0. By the same token, beyond a certain time the kth lagged coefficient may not have any impact on the regressand, thus supporting the assumption that β_k is zero. In our inventory example, the coefficient of X_{t-3} had a negative sign, which may not make economic sense. Hence, one may want to constrain that coefficient to zero.[51] Of course, you do not have to constrain both ends; you could put restriction only on the first coefficient, called near-end restriction or on the last coefficient, called far-end restriction. For our inventory example, this is illustrated in exercise 17.28. Sometimes the β's are estimated with the restriction that their sum is one. But one should not put such restrictions mindlessly because such restrictions also affect the values of the other (unconstrained) lagged coefficients.

3. Since the choice of the number of lagged coefficients as well as the degree of the polynomial is at the discretion of the modeler, some trial and error is inevitable, the charge of data mining notwithstanding. Here is where the **Akaike and Schwarz information criteria** discussed in Chapter 13 may come in handy.

4. Since we estimated (17.13.16) using three lags and the second-degree polynomial, it is a *restricted least-squares* model. Suppose, we decide to use three lags but do not use the Almon polynomial approach. That is, we estimate (17.13.11) by OLS. What then? Let us first see the results:

$$\begin{aligned}
\hat{Y}_t &= 26{,}008.60 + 0.9771X_t + 1.0139X_{t-1} - 0.2022\,X_{t-2} - 0.3935X_{t-3} \\
\text{se} &= (6691.12) \quad (0.6820) \quad (1.0920) \quad (1.1021) \quad (0.7186) \\
t &= (3.8870) \quad (1.4327) \quad (0.9284) \quad (-0.1835) \quad (-0.5476) \\
& R^2 = 0.9755 \quad d = 0.1571 \quad F = 379.51
\end{aligned} \tag{17.13.17}$$

If you compare these results with those given in (17.13.16), you will see that the overall R^2 is practically the same, although the lagged pattern in (17.13.17) shows more of a humped shape than that exhibited by (17.13.16).

[51]For a concrete application, see D. B. Batten and Daniel Thornton, "Polynomial Distributed Lags and the Estimation of the St. Louis Equation," *Review,* Federal Reserve Bank of St. Louis, April 1983, pp. 13–25.

As this example illustrates, one has to be careful in using the Almon distributed lag technique, as the results might be sensitive to the choice of the degree of the polynomial and/or the number of lagged coefficients.

17.14 CAUSALITY IN ECONOMICS: THE GRANGER CAUSALITY TEST[52]

Back in Section 1.4 we noted that, although regression analysis deals with the dependence of one variable on other variables, it does not necessarily imply causation. In other words, the existence of a relationship between variables does not prove causality or the direction of influence. But in regressions involving time series data, the situation may be somewhat different because, as one author puts it,

> . . . time does not run backward. That is, if event *A* happens before event *B*, then it is *possible* that *A* is causing *B*. However, it is not *possible* that *B* is causing *A*. In other words, events in the past can cause events to happen today. Future events cannot.[53] (Emphasis added.)

This is roughly the idea behind the so-called Granger causality test.[54] But it should be noted clearly that the question of causality is deeply philosophical with all kinds of controversies. At one extreme are people who believe that "everything causes everything," and at the other extreme are people who deny the existence of causation whatsoever.[55] The econometrician Edward Leamer prefers the term **precedence** over causality. Francis Diebold prefers the term **predictive causality.** As he writes:

> . . . the statement "y_i causes y_j" is just shorthand for the more precise, but long-winded, statement, "y_i contains useful information for predicting y_j (in the linear least squares sense), over and above the past histories of the other variables in the system." To save space, we simply say that y_i causes y_j.[56]

The Granger Test

To explain the Granger test, we will consider the often asked question in macroeconomics: Is it GDP that "causes" the money supply *M* (GDP → *M*)

[52]There is another test of causality that is sometimes used, the so-called **Sims test of causality.** We discuss it by way of an exercise.

[53]Gary Koop, *Analysis of Economic Data,* John Wiley & Sons, New York, 2000, p. 175.

[54]C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica,* July 1969, pp. 424–438. Although popularly known as the Granger causality test, it is appropriate to call it the **Wiener–Granger causality test,** for it was earlier suggested by Wiener. See N. Wiener, "The Theory of Prediction," in E. F. Beckenback, ed., *Modern Mathematics for Engineers,* McGraw-Hill, New York, 1956, pp. 165–190.

[55]For an excellent discussion of this topic, see Arnold Zellner, "Causality and Econometrics," *Carnegie-Rochester Conference Series,* 10, K. Brunner and A. H. Meltzer, eds., North Holland Publishing Company, Amsterdam, 1979, pp. 9–50.

[56]Francis X. Diebold, *Elements of Forecasting,* South Western Publishing, 2d ed., 2001, p. 254.