ESTIMATING LENGTHS AND ORDERS OF DELAYS IN SYSTEM DYNAMICS MODELS

by

Margaret S. Hamilton
Office of Resource Analysis
U.S. Geological Survey, Reston, Va. 22092

ABSTRACT

Delays are a ubiquitous feature of dynamic systems; they are present at every stage of an action. An understanding of delays is necessary if policy makers are to foresee the consequences of their actions. It is often not sufficient to rely on "expert" opinion to tell how long it will take for the repercussions of an action to be complete, because even the "experts" can seriously underestimate delay times. It is, therefore, important to have systematic methods of estimating the length of delays in system dynamics models. The time structure of delays is also important. Whether a delay is destabilizing or stabilizing will depend on whether the repercussions are concentrated or dispersed, as well as whether the time lag is long or short. Systematic methods of estimating the order of delays are, therefore, also useful. This paper presents five statistical methods that can be used to estimate lengths and orders of delays in system dynamics models. The presentation contains a discussion of when each method is applicable and what problems may be encountered in using it. Empirical results from applying two of the methods are discussed. The empirical studies respectively involve the problem of estimating the delay between changes in export prices and changes in export market shares and the problem of estimating the delay between capital appropriations and capital expenditures. The paper also offers guidelines for choosing an estimation technique and discusses validation of the estimates obtained.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	595
II.	TYPES OF DELAY	597
	Length and order of delays	597
	Material and information delays	599
III.	GENERAL FORMULATION OF THE PROBLEM	600
IV.	ESTIMATION TECHNIQUES	602
	Direct estimation	602
	Prespecifying the coefficients	602
	Geometric lag distribution	604
	Polynomial lag distribution	608
	Fitting Erlang distributions	610
V.	EMPIRICAL ESTIMATIONS	614
	An empirical study using the direct method	614
	An empirical study using polynomial lags	619
VI.	CHOOSING AN ESTIMATION TECHNIQUE	622
VII.	VALIDATION	625
	BIBLIOGRAPHY	627

I. INTRODUCTION

Delays are a ubiquitous feature of dynamic systems; they are present at every stage of an action. Time is required to recognize a problem, to decide what to do about it, and to implement a decision once made. Policy makers must understand delays if they are to foresee the consequences of their actions. Many decisions turn out to be faulty because people underestimate the length of delays. Wage-price controls are adopted and then abandoned as ineffectual before they have had the desired effect. Monetary policy is changed before results can be seen. It is often not sufficient to ask an "expert" how long it will take for the repercussions of an action to be complete, because even the "experts" can seriously underestimate delay times. For example, in 1971, collective wisdom, based on intuition, put the time that it would take for most of the effects of the Smithsonian currency realignment to work through at 1½ to 2 years. However, empirical research has shown that the time required for such changes is closer to 5 years. (Junz and Rhomberg, 1973). The "experts" were off by a factor of 3. It is, therefore, important to have systematic methods of estimating adjustment times in dynamic systems.

The time structure of delays must also be understood by policy-makers. Whether a delay is destabilizing or stabilizing will depend on whether the repercussions are concentrated or dispersed, as well as whether the time lag is long or short. For example, governments often try to stimulate investment in a counter-cyclical fashion. If there is a long lag between changes in policy instruments and changes in actual expenditure, and if the final results are concentrated in time, policies designed to stimulate investment may be destabilizing instead of stabilizing. Under such circumstances, the forecasting of economic conditions and the timing of policy measures will need to be very precise, and no policy at all may be preferable to one that does not take these considerations into account. It is thus also important to have systematic methods of estimating the time structure of delays.

Not all lengths and orders of delays in a model need to be estimated. It is often the case that system dynamics models are completely insensitive to the order of a delay and occasionally the case that even the length is unimportant. Sensitivity analysis should be employed to determine the sensitive parameters before substantial effort is expended on estimation. Once the sensitive parameters have been determined, one of the statistical techniques presented here can be used for estimation purposes. This paper describes five estimation techniques, offers guidelines for choosing among them, and discusses validation of the estimates obtained.

II. TYPES OF DELAYS

Length and order of delays

When an action occurs, its consequences are felt over a period of time. A delay can be specified by describing the time path of the repercussions. Such a time path can be characterized by its total length (the time required for the effects of an action to be fully worked out), its average length (essentially the half-way point; the time required for one half of the consequences to be felt[1]), and its shape (whether it is dispersed or concentrated, whether effects are felt immediately or only after some time has elapsed.)

Delays in system dynamics models are characterized by two parameters: their length and their order. The length of a delay, or adjustment time AT, is the average length defined above. The order of a delay is an integer greater than or equal to 1. A first-order delay of a variable produces an exponential average of past values of that variable (weighting more recent values more heavily); an nth-order delay is a sequence of n first-order delays. Consequently, low-order delays have an immediate response that is dispersed; high-order delays have a deferred response that is more concentrated[2]. By varying the length and order of a system dynamics delay, one

1) Strictly speaking, the average length of a delay in a system dynamics model is not the abscissa of the vertical line that divides the area under the time path into two equal parts; it is the abscissa of the center of gravity of the area under this curve.

2) For a more complete description of system dynamics delays, see Forrester (1968).

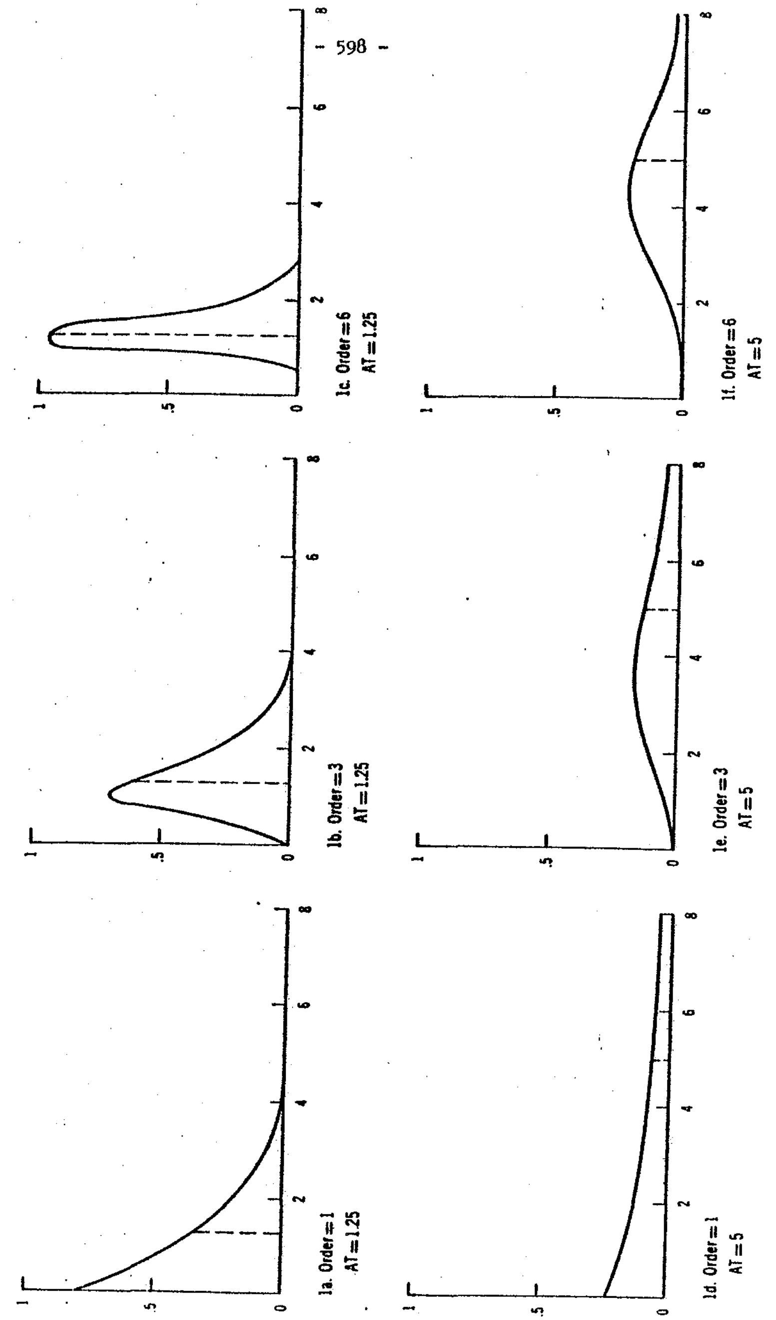


Figure 1 System dynamics delays

can approximate most well-behaved time paths. Figure 1 illustrates this variation. The first three delays have the same length but are different orders. The three delays in the second row are of the same respective order as the delays immediately above them but have different lengths.

Material and information delays

Delays can be subdivided into material and information delays. A material delay modifies a physical flow such as letters in the mail. If one hundred letters are mailed, they are not immediately received. Their arrival will be dispersed over a period of time, the average length of which will depend on postal conditions.

In contrast, information delays are delays in perception or reaction. People may not respond to changed conditions unless they feel that the change is permanent. A temporary increase in the price of a commodity will not cause producers to increase the supply of that commodity. If, however, the increase persists over a period of time, the change will be perceived as permanent and increased production will be initiated.

Most real-world delays combine information and material delays. For example, in making an investment expenditure, time is required to collect information about market conditions and to reach a decision on the basis of this information. This period of time is an information delay. When the order for capital equipment has been placed, time is required for the production, delivery, and installation of the equipment.

This period of time is a material delay. Since the time paths of material and information delays do not differ very much, there is no need to distinguish between the two types in estimating their lengths and orders.

III. GENERAL FORMULATION OF THE PROBLEM

Suppose that you mail X_t letters in time period t and receive, on the average, w_1X_t of those letters in period t+1, w_2X_t in period t+2, and so forth up to w_mX_t in period t+m (where $\Sigma w_i=1$). Then, if the system of reactions is constant over time, X_t^*, the expected number of letters received in any period t, can be written as a linear function of the m previous values of X_t.

1) $$X_t^* = w_1X_{t-1} + w_2X_{t-2} + \dots + w_mX_{t-m}$$
$$w_i \geq 0 \qquad i = 1, \dots, m$$
$$\sum_{i=1}^{m} w_i = 1$$

Equation 1 is quite general. X_t^* could be the expected price of a commodity in period t, determined by a weighted average of prices in previous periods, or it could be the expected income of an individual, determined by a weighted average of past incomes. The numerical values of the weights w_i will depend on the postal system in the first example; they will depend on the way in which expectations are formed in the second two cases.

Expressed in this fashion, a delay can be thought of as a probability distribution; w_i is the probability of a discrete random variable taking on the values 1, 2, ..., m. In the postal example, w_i is the probability that a letter mailed in period t will arrive in period t+i. The adjustment time of the delay, AT, or average length of the delay, is then the mean of the probability distribution.

$$2)\quad AT = \sum_{i=1}^{m} iw_i$$

In other words, AT is the average length of time a letter spends in the mail. The problem is to estimate the weights w_i from data.

More generally, we may wish to estimate Y_t, assumed to be a linear function of X_t^* .

$$3)\quad Y_t = a + bX_t^* + u_t$$

where u_t is a random disturbance. Substituting equation 1 into equation 3 we obtain

$$4)\quad Y_t = a + b(w_1X_{t-1} + w_2X_{t-2} + \ldots + w_mX_{t-m}) + u_t$$

$$5)\quad Y_t = a + b_1X_{t-1} + b_2X_{t-2} + \ldots + b_mX_{t-m} + u_t$$

where $b_i = w_ib \qquad i = 1, \ldots, m.$

The adjustment time AT can then be written

$$6)\quad AT = \sum_{i=1}^{m} iw_i = \sum_{i=1}^{m} (ib_i/\sum_{i=1}^{m} b_i)$$

If we know the coefficients b_i in equation 5, we can calculate the adjustment time from equation 6. Plotting the weights w_i against lagged time gives us the time structure of the delay.

IV. ESTIMATION TECHNIQUES

Direct estimation

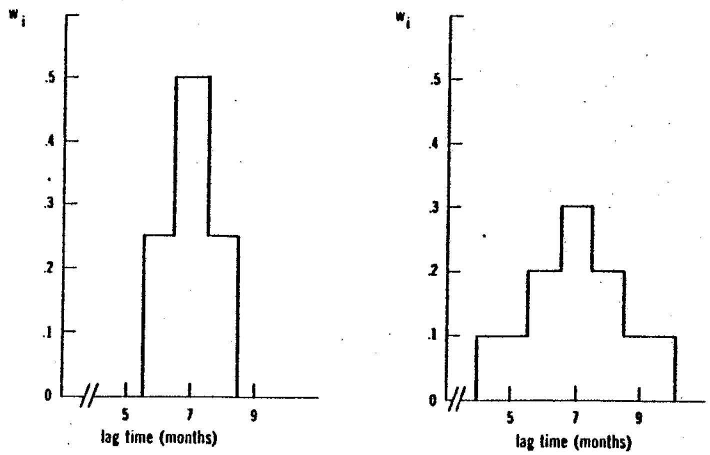
In some cases it is possible to use ordinary least-squares to estimate equation 5 directly. However, problems often arise. If the length of the lag, the number of periods over which the effect of an action is distributed, is long, there will be a large number of parameters to estimate (m+1). There may not be sufficient degrees of freedom to perform the estimation.

Another serious problem is multicollinearity. Frequently the various lagged values of X are highly intercorrelated, leading to imprecise estimates of the lagged coefficients (their variances and covariances will be large.) If the estimates of the coefficients are poor, estimates of the length and order of the delay will be unreliable as well. Multicollinearity is most apt to occur with time-series data when the period of observation is short.

In many estimation problems, detecting multicollinearity can be quite difficult. Fortunately this is not apt to be the case for direct estimates of time structures. Since the regressors are lagged values of the independent variable X, one should not expect collinearity to occur in a very complicated form, and examining the simple correlation coefficients between the lagged variables should be sufficient to test for multicollinearity.

Prespecifying the coefficients

Weymar (1968) tried to use the direct technique to estimate delays in his study of the dynamics of the world cocoa market. While working with a small number of observations of monthly



2a. Narrow distribution

2b. Wide distribution

Figure 2 Weymar's lag structures for monthly cocoa consumption

time-series data, he encountered severe problems with multicollinearit and degrees of freedom. To get around these problems he specified a priori two lag distributions (shapes of time paths), one wide and one narrow, and shifted these distributions along the time axis to test different lag times. This procedure amounts to prespecifying the coefficients w_i instead of estimating them. He chose the distribution and average lag that explained the largest amount of variance in cocoa consumption. Figure 2 shows Weymar's wide and narrow distributions for monthly cocoa consumption. He tested each distribution with several average lag times and arrived at a mild preference for the wide distribution with an average lag time of seven months.

Weymar's procedure is very cumbersome, and, unless the length and order of the delay are fairly well known in advance, involves performing a very large number of regressions. Several other methods of avoiding the problems associated with the direct approach are available. These methods consist of specifying a priori some assumption about the form of the weights, but do not involve specifying numerical values for them.

Geometric lag distribution

If there is reason to believe that a delay is a first order delay, a geometric lag can be used to estimate its length. A geometric lag model assumes that the weights w_i decline exponentially.

7) $w_i = (1 - \lambda)\lambda^{i-1} \qquad 0 < \lambda < 1 \qquad i = 1, 2, \ldots$

Here, the most recent values of X are given the largest weights. The effect of X on Y extends indefinitely into the past ($m \to \infty$), but the weights decline in a fixed proportion so that the effect of distant values of X eventually becomes negligible. Substituting equation 7 into equation 4 we obtain

8) $Y_t = a + b(1 - \lambda)(X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \ldots) + u_t$

Equation 8 can not be estimated as is, since it involves an infinite number of regressors. However, Koyck (1954) showed how to simplify the equation. Lag equation 8 one period, multiply through by λ, and subtract the result from the original equation to obtain

9) $Y_t = a(1 - \lambda) + b(1 - \lambda)X_{t-1} + \lambda Y_{t-1} + v_t$

$v_t = u_t - \lambda u_{t-1}$

Equation 9 contains only three parameters to estimate: a, b, and λ and two regressors: Y_{t-1} and X_{t-1}. There should be no problem with degrees of freedom or multicollinearity.

The adjustment time AT can be calculated from equation 2.

$$AT = \sum_{i=1}^{\infty} i w_i = (1 - \lambda) \sum_{i=0}^{\infty} i\lambda^{i-1} \quad [3]$$

$$= (1 - \lambda) \frac{d}{d\lambda} \{\sum_{i=0}^{\infty} \lambda^i\} = (1 - \lambda) \frac{d}{d\lambda} \{\frac{1}{1 - \lambda}\}$$

$$= \frac{(1 - \lambda)}{(1 - \lambda)^2} = \frac{1}{1 - \lambda}$$

10) $AT = \frac{1}{1 - \lambda}$

This result can also be obtained directly by recognizing equation 7 as a geometric probability distribution with mean $\frac{1}{1 - \lambda}$.

3) Since $i\lambda^{i-1} = 0$ for $i = 0$, the subscript i can go from 0 to ∞.

The DYNAMO[4] rate equation of a first-order delay macro can be expanded to show that it is precisely equivalent to a geometric lag.

	DYNAMO equations	Equations with t subscripts (assuming DT = 1)
R	O·KL = L·K/AT	$O_t = L_t/AT$
L	L·K = L·J + (DT)(I·JK - O·JK)	$L_t = L_{t-1} + I_{t-1} - O_{t-1}$

$$O_t = L_t/AT$$

$$= (L_{t-1} + I_{t-1} - L_{t-1}/AT)/AT$$

$$= I_{t-1}/AT + \frac{(1 - \frac{1}{AT})L_{t-1}}{AT}$$

$$= \frac{I_{t-1}}{AT} + \frac{(1 - \frac{1}{AT})I_{t-2}}{AT} + \frac{(1 - \frac{1}{AT})^2 I_{t-3}}{AT} + \ldots$$

From equation 10, $AT = \frac{1}{1 - \lambda}$ and $\lambda = 1 - \frac{1}{AT}$. Substituting, we obtain

11) $O_t = (1 - \lambda)I_{t-1} + (1 - \lambda)\lambda I_{t-2} + (1 - \lambda)\lambda^2 I_{t-3} + \ldots$

The output O_t is a geometric lag of the input I_t.

There are two drawbacks to using geometric lags to estimate adjustment times. First, it is necessary to assume a priori that the delay is first-order. But it is preferable to let the data determine the order of the delay (shape of the time path.) Second, and more serious, it is not possible to use ordinary least-squares to estimate the coefficients a, b, and λ.

4) DYNAMO is a computer language used in simulating system dynamics models. It is described in Pugh (1970).

The disturbance term v_t in equation 9 is correlated with Y_{t-1} which is one of the explanatory variables. In fact

$$E\{v_t Y_{t-1}\} = E\{(u_t - \lambda u_{t-1})\ (a + b(1 - \lambda)$$
$$(X_{t-2} + \lambda X_{t-3} + \ldots) + u_{t-1})\}$$
$$= -\lambda\sigma^2$$

where σ^2 is the variance of u_t. Ordinary least-squares estimates of the coefficients will be inconsistent (they will be biased, and the bias will persist even as the number of observations approaches infinity.) Some other estimation technique, such as the method of instrumental variables, must be used[5].

In system dynamics models, higher-order delays are obtained by aggregating several sequential first-order delays (cascading several levels.) The probability distribution for an nth-order delay can be obtained by convoluting (summing sequentially) n independent and identically distributed geometric random variables.

---

5) A description of the method of instrumental variables can be found in any standard econometrics text, such as Kmenta (1971, pp. 479-480). For a good discussion of the problems of estimation with serially-correlated disturbances, see Griliches (1967).

The resulting random variable has a negative binomial distribution.

$$12)\quad w_i = \binom{i + n - 1}{i}(1 - \lambda)^n \lambda^i \qquad i = 0, 1, \ldots$$

with mean $n\lambda/(1 - \lambda)$. A distributed lag model with weights defined by equation 12 is known as a Pascal lag model. Such a model can be estimated and the adjustment time determined from the estimated coefficients, but the procedure is not recommended. Pascal models are very difficult to estimate and, in addition, have all the problems associated with the geometric distributed lag. (The order must be specified in advance. The disturbance terms are serially correlated.) Other methods of estimating higher-order delays exist which, while not mathematically equivalent to higher-order system dynamics delays, are easy to use and give similar results.

Polynomial lag distribution

A polynomial lag, suggested by Almon (1965), is a very flexible approach and can be used to estimate both the length and the order of a system dynamics delay. The idea is to approximate the time path with a polynomial. The method is based on Weierstrass's theorem, which states that a function continuous in a closed interval can be approximated with any prespecified accuracy by a polynomial of suitable degree. Since most time paths are reasonably well-behaved, a polynomial of fairly low degree (3 or 4) should suffice.

The degree n of the polynomial and the total length of the lag k must be chosen in advance. These choices can be tested by varying both parameters. The weights w_i will lie along a polynomial of the chosen degree.

13) $w_i = \lambda_0 + \lambda_1 i + \lambda_2 i^2 + \ldots + \lambda_n i^n \qquad i = 1, \ldots, k.$

The number of parameters to be estimated depends only on the degree of the polynomial, not on the length of the lag. Therefore, if the degree is fairly low, there should be no problem with degrees of freedom. Since the explanatory variables involve different powers of i, multicollinearity should not be a problem. The original disturbance term appears as the disturbance term in the transformed equation; so, if there was no serial correlation to begin with, none has been introduced. As a result, ordinary least-squares can be used to perform the estimation, and the resulting estimates should have all the desirable properties. In particular, they will be unbiased (and, therefore, consistent). Standard econometric texts discuss how to transform the original equation into one suitable for estimation, as well as how to reduce the number of parameters that must be estimated (by constraining the polynomial to cross the time axis at particular points and by forcing the weights to sum to 1.) For example, see Kmenta (1971, pp. 492-493).

Once the weights have been estimated, the adjustment time can be calculated from equation 6. The weights w_i can be plotted against lagged time and compared to computer plots of various-order system dynamics delays with the calculated adjustment time. This procedure should give a fair estimate of the order of the delay. One should not try to match the order too accurately, since there is no exact equivalence between an nth-order system dynamics delay and a polynomial lag distribution. It should be sufficient to determine if the delay is of 1st, 3rd, or higher order.

<u>Fitting Erlang distributions</u>

The estimation techniques discussed so far have all been forms of distributed lags. Using these techniques requires time-series data on the dependent and independent variables Y and X (the letter receiving rate and letter mailing rate, for example). It is not necessary to be able to associate particular letters received with particular letters mailed. In some cases, however, additional data may be available. We could perform an experiment: mail 100 letters and count how many arrive on the 1st day, on the second day, and so forth, to obtain a histogram of the mail system. Consider another example. If the system to be modeled is a hospital, the hospital may keep records of the number of patients that stay between 0 and 1 week, between 1 and 2 weeks, and so forth, as well as records of the number of patients being admitted and discharged each week. When this sort of detailed data are available, there is a particularly simple method of calculating the length and order of the delay (the delay being the stay in the hospital.) The method is based on the fact that the continuous probability distribution associated with a kth-order system dynamics delay is an Erlang type k distribution.

The probability distributions considered so far (the geometric and negative binomial distributions corresponding to 1st-and higher-order delays) have been discrete, of necessity, because the data were discrete. A system dynamics model, however, is continuous. The true probability distribution associated with a first-order delay is the exponential distribution.

14) $(1 - \lambda)e^{-(1 - \lambda)t} \qquad t > 0$

This is the limiting case of the geometric distribution as the period of observation (time increment) becomes smaller and smaller. The distribution associated with a kth-order delay can be obtained by convoluting (summing sequentially) k independent and identically distributed exponential random variables. The resulting random variable has an Erlang type k distribution.

15) $\dfrac{\{k(1 - \lambda)\}^k}{(k - 1)!}\, t^{k-1} e^{-k(1 - \lambda)t} \qquad t > 0$

The exponential distribution is the special case of the Erlang obtained by setting k = 1. The Erlang type k distribution has mean $1/(1 - \lambda)$ and variance $1/\{k(1 - \lambda)^2\}$. Figure 3 shows Erlang distributions with various values of k. As $k \to \infty$, the

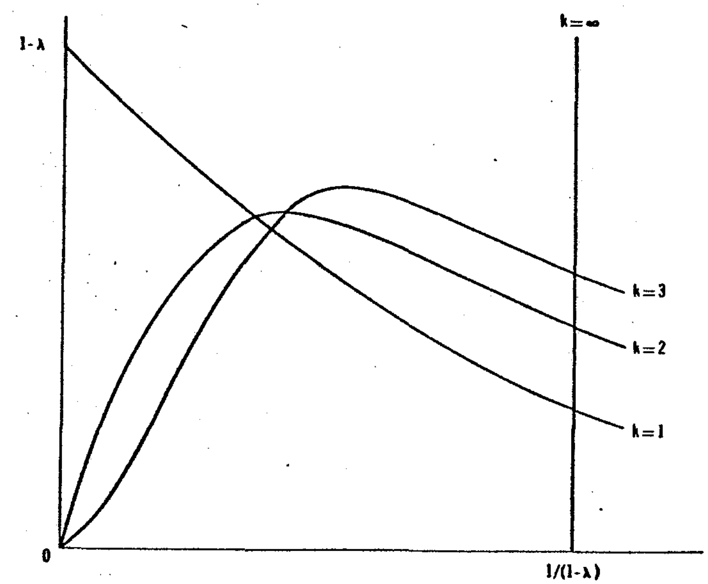


Figure 3 A family of Erlang distributions with mean $1/(1 - \lambda)$

distribution becomes completely deterministic (it becomes concentrated at the mean.) This outcome can also be seen from the formula for the variance; the variance becomes 0 as $k \to \infty$.

The formulas for the mean and variance of an Erlang distribution can be used to calculate the length and order of the delay. Let n_i be the number of patients staying in the hospital from i - 1 to i weeks, i = 1, 2, ... The sample mean is

$$16)\quad \bar{X} = \sum_i in_i / \sum_i n_i \doteq \frac{1}{(1-\lambda)} = AT$$

The adjustment time AT is approximately equal to the sample mean. We can also calculate the sample variance.

$$s^2 = \frac{\sum_i n_i (i - \bar{X})^2}{\sum_i n_i} \doteq \frac{1}{k}\,\frac{1}{(1-\lambda)^2} \doteq \frac{1}{k}\,\bar{X}^2$$

This means that

$$k \doteq \frac{\bar{X}^2}{s^2}$$

The order of the delay is approximately equal to the square of the sample mean divided by the sample variance. Of course k won't necessarily be an integer, but if k = 2.85, for example, one can conclude that the delay is 3rd-order. This calculation is so simple (it involves no regression) that it should be used whenever the appropriate data are available.

V. EMPIRICAL ESTIMATIONS

An empirical study using the direct method

Junz and Rhomberg (1973) used the direct approach to estimate the timing of the effects of relative price changes on export flows. Such changes in relative prices can be brought about either by alterations in exchange rates or by changes in export prices measured in national currencies. Proportional changes in market shares were related to proportional changes in relative export prices, yielding a price elasticity of market shares. Such a study aggregates several information and material delays: the recognition lag required for the changed competitive situation to be perceived, the replacement lag required for inventories to be depleted or equipment to wear out, the production lag required for producers to undertake the expense of shifting from supplying one market to supplying another or of adding capacity in order to supply additional markets, to name only a few. Whether it is advisable to model (and therefore estimate) each delay separately or combine them in one aggregate delay depends on the model being constructed and whether disaggregation changes the dynamic behavior of the system. In what follows, all of the constituent delays have been combined.

The data consist of annual observations for the years 1958-1969 (allowing for 5 lags in price data going back to 1953) for 13 exporting countries and 13 markets in each country. The fitted equation is

$$\widehat{PCMS}_t = .019 - \underset{(2.2)}{.52PCP_t} - \underset{(1.2)}{.29PCP_{t-1}} - \underset{(2.4)}{.58\ PCP_{t-2}}$$

$$- \underset{(4.2)}{.98PCP_{t-3}} - \underset{(1.0)}{.24PCP_{t-4}} - \underset{(1.2)}{.27PCP_{t-5}}$$

where PCMS stands for percent change in market share and PCP for percent change in relative export price. The numbers shown in parentheses are t ratios. Three of the six regression coefficients are statistically significant at the 95% confidence level.

The response can be divided into an immediate response (for items that are produced quickly and are relatively homogeneous across suppliers) and a delayed response. Figure 4a shows the delayed response plotted against lagged time in years. A smooth curve has been drawn to approximate the time structure. It can be seen that in the 5th year there is still quite a lot of adjustment taking place. Figure 4b shows the cumulative delayed response (the sum of the responses up to any point in time) plotted against time. The adjustment time can be calculated from equation 6.

$$AT = \sum_{i=1}^{5} ib_i / \sum_{i=1}^{5} b_i = 2.84 \text{ years}$$

The peak response occurs in the third year.

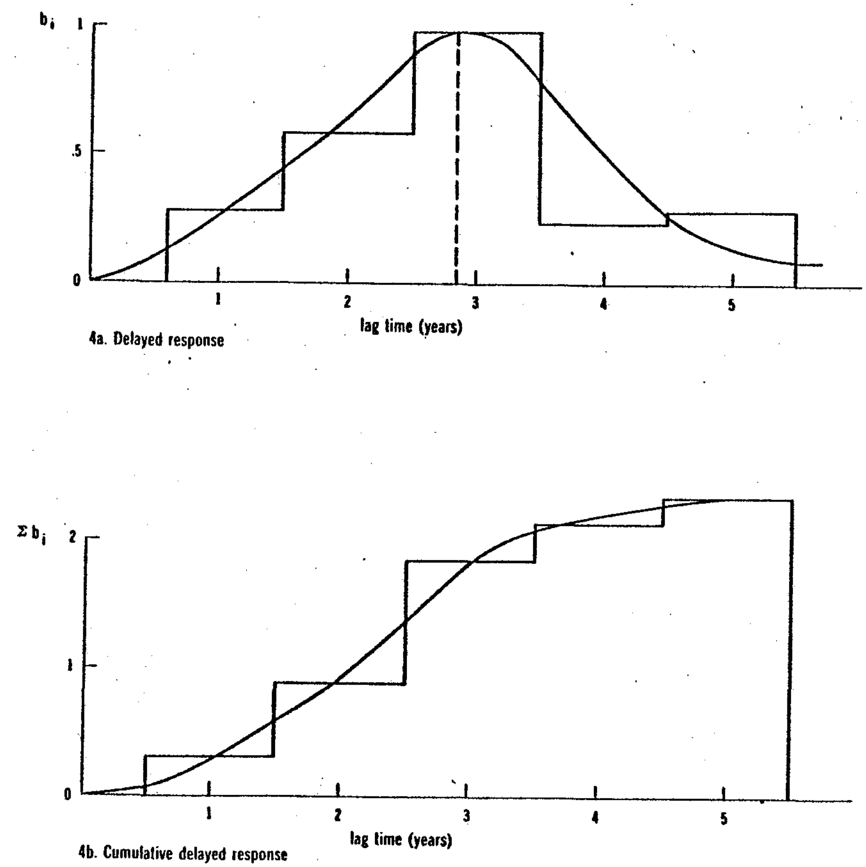


Figure 4 Junz and Rhomberg estimated time structures

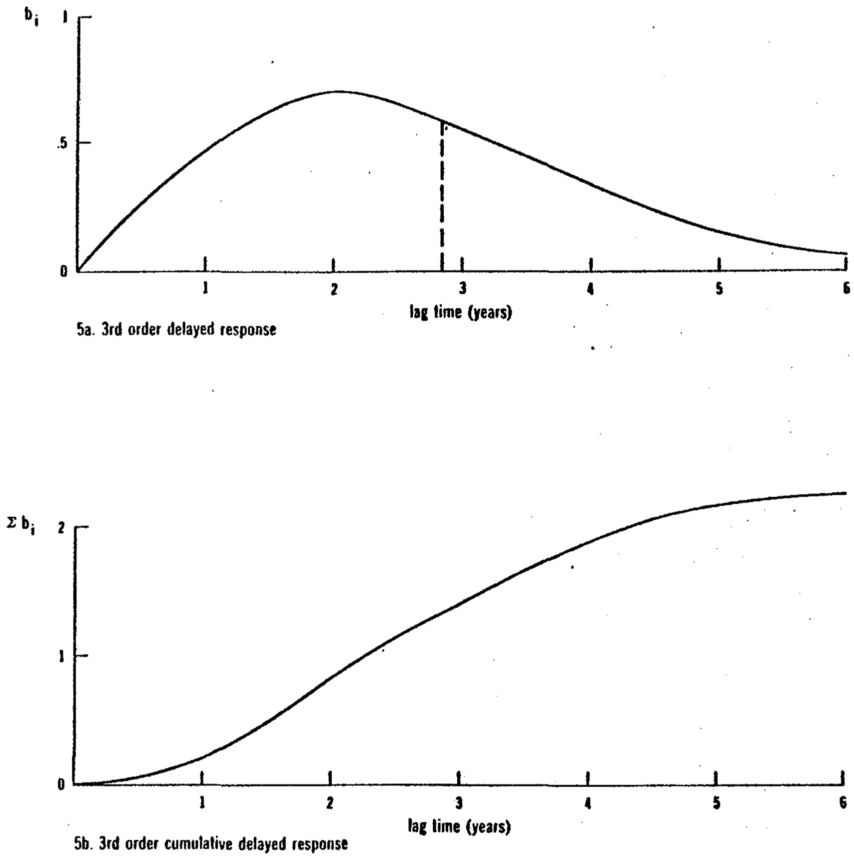


Figure 5 System dynamics 3rd order delays with adjustment times of 2.84 years

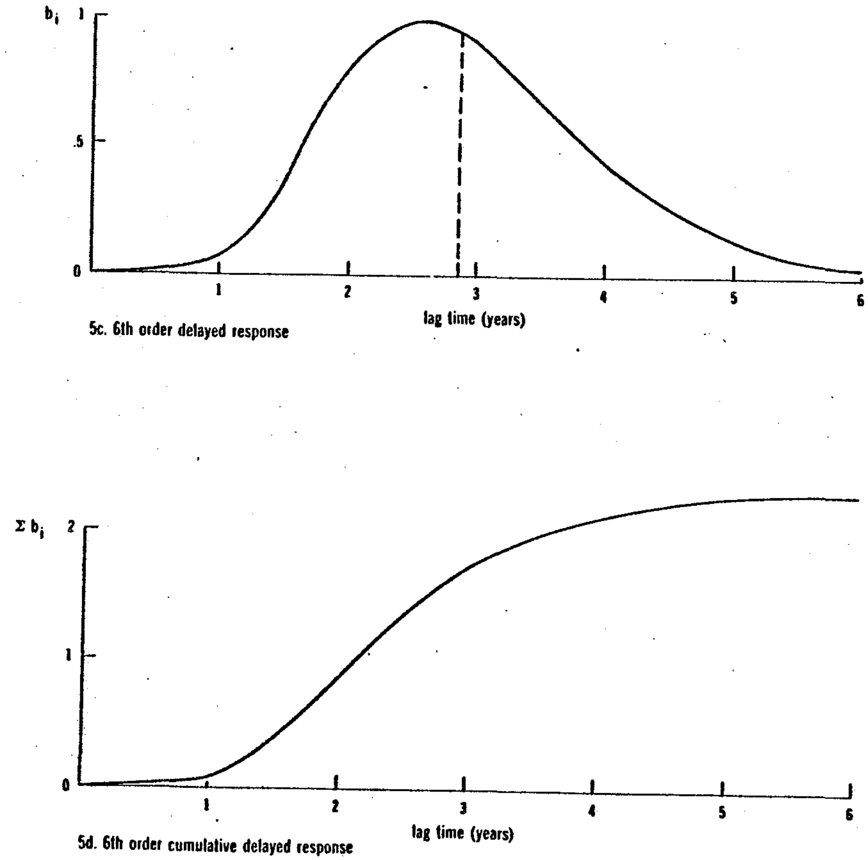


Figure 5 System dynamics 6th order delays with adjustment times of 2.84 years

Figures 5a and 5b show pulse and cumulative responses of 3rd-order system dynamics delays with adjustment times of 2.84 years. Figures 5c and 5d give similar plots for 6th order delays. The 6th-order delay gives a closer approximation, but not too much weight should be given to the exact order of the delay. What is important is that the delay is of higher-order.

<u>An empirical study using polynomial lags</u>

Almon (1965) employed polynomial lags to predict quarterly capital expenditures in manufacturing industries from present and past capital appropriations. The quarterly data on appropriations and expenditures came from the survey conducted by the National Industrial Conference Board among the thousand largest manufacturing companies in the U.S. Estimates were based on the nine years 1953-1961. Dummy variables were added to the equation to remove the effects of seasonal variation in expenditures. The length of the lag was assumed to be seven years. The resulting equation for all manufacturing industries (neglecting the seasonal variables and the constant term) is

$$\begin{aligned}\widehat{KE}_t = {} & \underset{(2.09)}{.048KA_t} + \underset{(6.19)}{.099KA_{t-1}} + \underset{(10.85)}{.141KA_{t-2}} + \underset{(7.17)}{.165KA_{t-3}} \\ & + \underset{(7.26)}{.167KA_{t-4}} + \underset{(11.23)}{.146KA_{t-5}} + \underset{(6.56)}{.105KA_{t-6}} + \underset{(2.21)}{.053KA_{t-7}}\end{aligned}$$

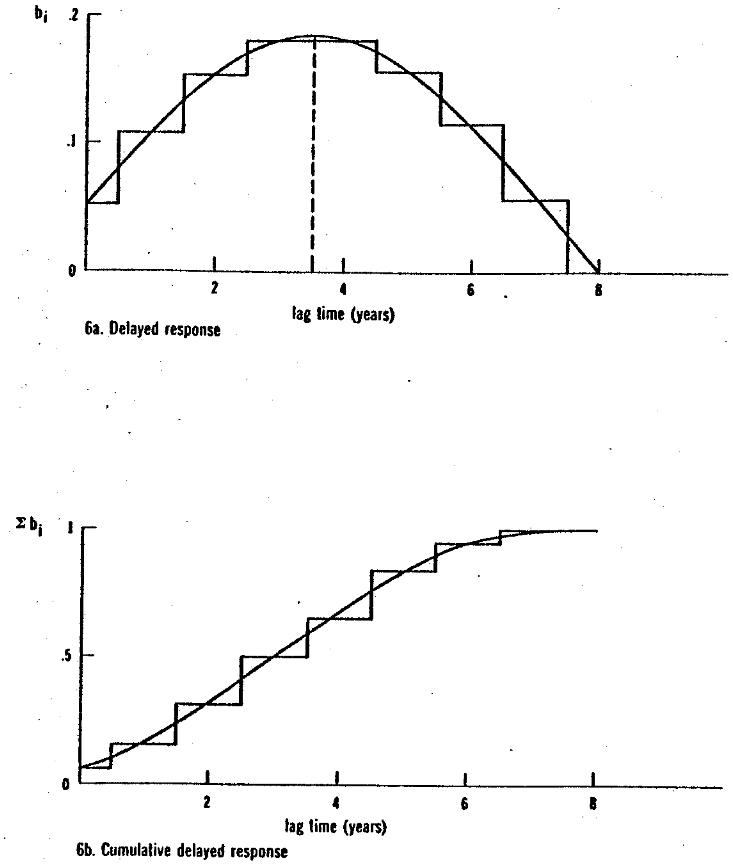
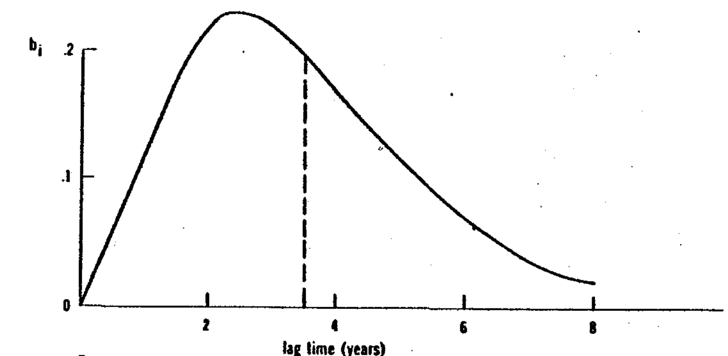
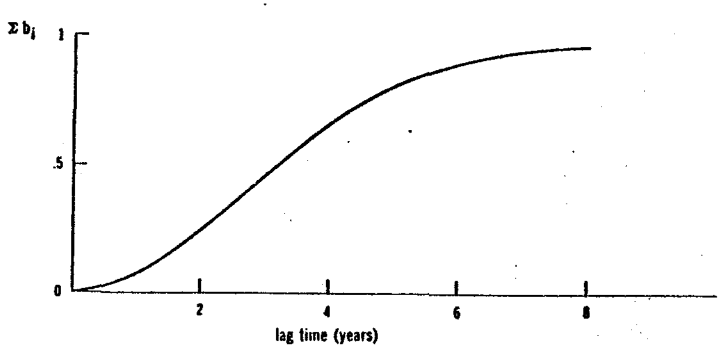


Figure 6 Almon estimated time structures



7a. Delayed response



7b. Cumulative delayed response

Figure 7 System dynamics 3rd order delays with adjustment times of 3.55 years

where KE and KA stand for capital expenditures and appropriations, respectively. The numbers shown in parentheses are t ratios. All of the lagged variables are very significant. The weights, which were not constrained to sum to 1, add up to .924. The difference between .924 and 1 can be very nearly accounted for by cancellations. The adjustment time can be calculated as

$$AT = \sum_{i=0}^{7} ib_i / \sum_{i=0}^{7} b_i = 3.55 \qquad 6)$$

The peak response comes in the 4th year. Figures 6a and 6b show the pulse and cumulative responses of the estimated time paths; figures 7a and 7b show similar responses for 3rd-order system dynamics delays with the same adjustment time.

VI. CHOOSING AN ESTIMATION TECHNIQUE

This paper has presented five techniques for estimating lengths and orders of delays in system dynamics models and has discussed the advantages and disadvantages of each. Table 1 summarizes this information.

The modeler, having decided through sensitivity analysis that a delay should be estimated and faced with a choice among

---

6) The subscript i runs from 0 to 7 in this example because Almon chose to include the present value of X in the lag structure.

Table 1

Method	Advantages	Disadvantages
Direct	Uses observations directly. No need to specify the order of the delay.	There may be too many parameters to estimate (too few degrees of freedom.) The explanatory variables may be highly intercorrelated (multi-collinearity.)
Prespecified coefficients	Uses observations directly.	Length of delay must be specified. Shape of delay must be specified. Many regressions may have to be performed.
Geometric lag	Uses observations directly. Simple functional form. Exact equivalence with 1st order delay.	Useful only for 1st order delays. Serial correlation of residuals (ordinary least squares can not be used).
Polynomial lag	No need to specify order of the delay.	Variables must be transformed before estimation can take place.
Fitting an Erlang distribution	Very simple calculation. Exact equivalence with system dynamics delay. No need to specify the order of the delay.	Requires additional data

five techniques, should consider a few guidelines. [7)]

1) If the data required to fit an Erlang distribution are available, this method should be used; it is by far the simplest approach and is exactly equivalent to a system dynamics delay.
2) If there are sufficient observations, and multicollinearity does not present a problem, then the direct technique should be used.
3) If there is reason to believe that a delay is 1st-order, or if the order of the delay is not important, and if the modeler has access to a program that corrects for serial correlation, then a geometric lag should be used.
4) The polynomial lag, though somewhat more cumbersome, provides the greatest flexibility and overcomes most of the special conditions that need to be met before the other methods can be used. This technique will work in nearly all circumstances.

---

7) The techniques are listed in order of preference. That is, the modeler should consider 1) then 2) and so forth.

VII. VALIDATION

Once estimates have been obtained, how does the modeler develop confidence in them? There are several things that can be done to test the robustness of the estimates.

1) The length of the lag can be varied for either the direct or the polynomial method; the degree of the polynomial can be varied for the latter.
2) The data can be disaggregated and new estimates made. For example, in the estimation of market shares, single markets can be considered; in the estimation of capital expenditures, manufacturing can be separated into durables and non-durables or even further disaggregated.
3) As new data become available they can be added to the old and new estimates made.

If the resulting time structures are similar to the original ones, confidence in the estimates will be improved.

The two applications described in this paper differ in one fundamental respect. Whereas nearly all capital expenditures can be accounted for by appropriations minus cancellations, variation of relative prices is only one of the determinants of export market shares. In the latter case, variables not included in the model are thrown into the residual category. It would therefore be of interest

to know just how sensitive the estimates are to noise. Sensitivity to noise can be tested by adding a stochastic variable to the data which might be correlated with one of the explanatory variables or with the calculated residuals. If the estimates still do not change very much, confidence will be further increased.

B I B L I O G R A P H Y

1) Almon, Shirley, Jan., 1965, "The Distributed Lag Between Capital Appropriations and Expenditures", Econometrica, Vol. 33, No. 1, pp. 178-196.

2) Forrester, Jay W., 1968, Principles of Systems, Wright-Allen Press, Cambridge, Mass.

3) Griliches, Zvi, Jan., 1965, "Distributed Lags, A Survey", Econometrica, Vol. 35, No. 1, pp. 16-48.

4) Junz, Helen B., and Rhomberg, Rudolf R., May, 1973, "Price Competitiveness in Export Trade Among Industrial Countries," American Economic Review, pp. 412-418.

5) Kmenta, Jan., 1971, Elements of Econometrics, the Macmillan Co., New York.

6) Koyck, L.M., 1954, Distributed Lags and Investment Analysis, North Holland Publishing Co., Amsterdam.

7) Pugh, Alexander L., 1970, DYNAMO II User's Manual, M.I.T. Press, Cambridge, Mass.

8) Weymar, F. Helmut, 1968, The Dynamics of the World Cocoa Market, M.I.T. Press, Cambridge, Mass.