3-1-1967

# Distributed Lags and Unobserved Components in Economic Time Series

Marc Nerlove

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 221



# DISTRIBUTED LAGS AND UNOBSERVED COMPONENTS IN ECONOMIC TIME SERIES

Marc Nerlove

March 13, 1967

# DISTRIBUTED LAGS AND UNOBSERVED COMPONENTS IN ECONOMIC TIME SERIES*

by

Marc Nerlove

## I. Introduction

Irving Fisher [9] was the first, to my knowledge, to use and discuss the concept of a distributed lag. Theoretically, such a lag arises when any economic cause (for example, a price or an income change) produces its effect (for example, on the quantity demanded or on the quantity supplied) only after some lag in time, so that this effect is not felt all at once at a single point in time but is distributed over a period of time. In a later paper [10, p. 323], Fisher described the central problem in applying the theory of distributed lags as primarily a statistical one. It was, he said, "... to find the 'best' distribution of lag, by which is meant the distribution such that ... the total combined effect [of the lagged values of the variables taken with a distributed lag has]... the highest possible correlation with the actual statistical series with which we wish to compare it." We have learned, however, in the years since Fisher wrote, an important principle of which he himself was

---

* This paper is an outgrowth of research on methods of analyzing economic time series conducted jointly by David Grether and the author under National Science Foundation Grant NSF-GS-818 to Yale University. I am indebted to Mr. Grether not only for his important general contribution but also for a large number of detailed comments and suggestions on the present paper. T. C. Koopmans has also made a number of helpful suggestions which are gratefully acknowledged.

certainly not unaware: In econometrics we can rarely solve our problems by statistical means alone; many assumptions are necessary before such data as we have can begin to yield new knowledge; these assumptions are more fruitful the less arbitrary they are and the better founded in the general corpus of economic theory and empirical knowledge of economic behavior.

Early users of distributed lags such as Fisher himself, Alt [1], Tinbergen [34], and later Koyck [19], recognized the need for assumptions on the form of the lag distribution. But the assumptions they made were largely based on considerations of statistical convenience and had little or no underpinning in theoretical models of economic behavior. Following the lead of Cagan [5] and Friedman [12], Nerlove [23, 24, 25, and 26] formulated models of expectation formation and partial adjustment towards equilibrium. In simple cases, these models resulted in a form of distributed lag originally proposed by Koyck for reasons of statistical convenience. For systems, however, of related behavioral equations (e.g., related demand functions or supply functions for industries consisting of multiproduct firms) the expectational model and the partial adjustment model were found to yield implications with important differences. Both the expectational model and the partial adjustment model were given a certain amount of theoretical underpinning, but this was only of a most general character and at no time could the explicit difference equation formulations of the models be regarded as more than a relatively crude approximation for statistical purposes. There is clearly a need, as Griliches [14] has recently reemphasized, for more rigorous derivation of the precise forms of

distributions of lag from theoretical models of behavior.

In the case of partial adjustment models of consumer behavior, for example, a natural approach would be to assume that a consumer has only local knowledge of his utility function so that when prices or his income change, he cannot proceed immediately to a new equilibrium position but must iterate by some convergent gradient procedure to the new maximizing position. Such a procedure might be chosen arbitrarily or, for example, deduced from a model assuming maximization of discounted utility over time. In the case, of models of firm behavior, distributed lags of a partial adjustment sort may be derived on the basis of models assuming additional costs of rapid, as opposed to slower, changes in input levels. See, for example, Nerlove [27], Nerlove and Arrow [29], and Eisner and Strotz [8, pp. 61-116]. The distributions of lag implied by such models are more complex than the simple distribution suggested by the original partial adjustment model which assumed adjustment in proportion to the difference between current and equilibrium positions.

The simple adaptive expectations model has recently been given a foundation in terms of optimal forecasting procedures; see Muth [22], Nerlove and Wage [30], and Couts, Grether, and Nerlove [6]. In this approach, the distribution of lag resulting from simple adaptive expectations or somewhat more complicated adaptation schemes, is derived as a minimum mean-square-error forecast on the basis of certain assumed stochastic structures for the causative variable in the analysis. Thus, the justification for the lag distribution involves a number of distinct hypotheses: First, we assume that the economic agent's behavior is in response not to the actual value of a causative variable but rather to a forecast of a future value. For example, it

is plausible that farmers plant not in response to what prices actually are but in response to what they think they will be at harvest time. Often, the precise date of the forecast is not specified, but rather some "normal" or long-term average is sought [25, pp. 51-2]. It is a remarkable fact, however, that for certain nonstationary time series the optimal forecast, in a sense to be described below, is the same for all future periods. Second, in the application of this approach we assume that economic agents base their forecasts on the past values of the variable in question and that they optimize their forecasts given knowledge of some stochastic specification of the mechanism generating the time series of the causative variable. At this point, one may adopt either of two hypotheses, each of which has quite different implications for estimation. One might argue, for example, that the economic agents have a clear conception of what the stochastic mechanism really is; then analysis of their behavior might plausibly begin by estimation of the stochastic structure of the time series to which they react, then determine optimal predictors, and, finally, use such directly as variables in a subsequent statistical analysis of the behavioral relation or indirectly to transform the equation to be estimated. Alternatively, we might assume some rather general stochastic structure, derive the form of the optimal predictors, then determine the actual values by means of an analysis of the relation between the predicate variable and past values of the causative variables the form of which will have been suggested by our earlier assumptions.

It is, of course, clear that precisely to what forecasts our economic agents ought to react, and indeed the sense in which such forecasts ought to be optimal, should be determined by some sort of theory of optimization over time. To do so, however, greatly complicates the analysis of behavior, and it thus appears fruitful to explore approaches similar to those described in the preceeding paragraph while recognizing the desirability of deeper levels of analysis. At such levels, considerations underlying attempts to derive lags due to partial adjustment, merge with those which have implications for what forecasts should be considered and how they ought to reflect past values of the variables in question.

In the present paper, I shall examine an approach to the derivation of lag distributions which is very much in the spirit of derivations as optimal forecasts. In contrast, however, to the forecasting justification, I shall assume that the causative variables are divided by economic agents into two or more unobserved components having definite stochastic properties, and that these agents react not to the observed variables but rather to estimates of the current values of the unobserved components. We will derive such estimates in a manner which minimizes the expected value of the squared difference between the estimated and the actual value; that is to say, our theory will be that the values of the variables to which our economic agents react are the minimum mean-square-error "extractions" of the unobserved components. Such extractions are related to, but not, except in special cases, identical to predictions of the observed variables. The sort of justification for distributed lags offered by this approach is thus distinct from one based

on the notion that certain lag distributions arise because economic agents react to forecasts rather than actual values of certain variables.

The idea that one may divide an economic time series into several unobserved, but separately meaningful components goes back at least to the work of Buys-Ballot [4]. The class of models embodying this idea was once fashionable in economics and includes the well-known errors-in-variables models. In the simplest case of such a model, we suppose that the two observed variables, $x_t$ and $y_t$, are divided into two unobserved components, the "true" values, $\xi_t$ and $\eta_t$, respectively, and the "errors," $u_t$ and $v_t$, respectively:

$$(1.1) \qquad \begin{cases} x_t = \xi_t + u_t \\ y_t = \eta_t + v_t \end{cases}$$

It is also generally supposed that the "true" values of $x_t$ and $y_t$ are connected by an exact relation, e. g.,

$$(1.2) \qquad \eta_t = \alpha\xi_t + \beta ,$$

while $u_t$ and $v_t$ are unrelated. The problems of estimating $\alpha$ and $\beta$ have been the subject of extensive investigation; see, for example, Malinvaud [21, pp. 326-363]. The essential point, however, for our purpose is that the observed variables $x_t$ and $y_t$ are divided into two unobserved components, and corresponding components in each series are related differently, i.e., $u_t$ is related to $v_t$ in quite a different manner than $\xi_t$ is related to $\eta_t$.

Although interest in models of errors-in-variables waned considerably following the sweep of the field by the "shock" model in the forties, the errors in variables approach is not entirely without recent application. Indeed, as is well-known, the "permanent income" hypothesis of consumption behavior developed by Friedman [12] can be viewed in exactly this way. Without attempting to do justice to Friedman's ingenious and complex theory, it may be described in brief as follows: Both income and consumption as observed may be divided into two unobserved components called, respectively, the permanent and the transitory component. Friedman supposes that there exists an exactly proportional relationship between the two permanent components, but no relationship between the two transitory components or between these, on the one hand, and permanent component of either income or consumption. Friedman argues that "... if a consumer unit knows that its receipts in any one year are unusually high and expects lower receipts subsequently, it will surely tend to adjust its consumption to its 'normal' receipts rather than to its current receipts." [12, p. 10]. The permanent component of income, according to Friedman, "... is to be interpreted as reflecting the effect of those factors that determine the consumer unit's capital value or wealth ...," while the transitory component "... is to be interpreted as reflecting all 'other' factors, factors that are likely to be treated by the unit affected as 'accidental' or 'chance' occurrences, though they may, from another point of view, be the **predictable** effect of specifiable forces, for example, cyclical fluctuations in economic activity." [12, pp. 21-22]. When Friedman comes to the analysis of aggregate consumption and income over time, he specifies a

relation between past income and what people consider to be the permanent component of income which amounts to taking consumption as a function of income with an exponential distribution of lag.[1] Indeed, Friedman's justification for this form is the continuous analogue of the adaptive expectations hypothesis which leads to distributions of lag based on geometrically declining weights.

The early economic statisticians also practiced the decomposition of observed time series into unobserved components. The work of Persons [32], Frickey [11], and especially of Kuznets [20] developed the idea in a way which still forms the basis of present-day methods of seasonal adjustment. In one version of this approach the observed time series is divided into three unobserved components: trend-cycle, seasonal, and irregular. These components may be assumed additive or multiplicative, independent or related. Although it is usually not very clear or explicitly indicated just what else is assumed about the nature of these components, it appears evident that the components of different sorts in different series are related differently. The division itself would have little purpose were this not the case. More important, the attempt to decompose the observed series into its unobserved components would make little sense except on the supposition that the various components were affected differently by various economic policies or by movements of corresponding or non-corresponding components in other series. From an analytical point of view, it is plausible, for example, that a manufacturer deciding on inventory levels will react somewhat differently to a change in sales he regards as being purely seasonal in character than he will to one he

regards as more permanent or longer lasting or one he regards as exceptionally ephemeral.

As before, at a deeper level of analysis, the nature of the decomposition into unobserved components would be derived from a model of optimization over time. However, in this paper, I shall take for granted that an economic time series may be divided into economically meaningful unobserved components with certain general stochastic characteristics and that these components, to the extent they are known or estimated, affect behavior differently. I shall assume that these components are estimated by minimum mean-square-error methods given certain assumptions about the underlying stochastic characteristics of the time series in question. Together, these assumptions suffice to determine distributed lag relationships. In particular, I shall exhibit a model which leads to a slight generalization of the geometric lag distribution for income in a consumption function which may be interpreted as a variant of the permanent income approach. A more complicated lag distribution is also derived which is related to, but not identical with, Solow's Pascal lag distribution [33] and which includes the geometric lag distribution as a special case. Finally, more complicated models involving seasonality will be introduced and a general discussion of the type of distributed lag relation given in which reactions to seasonal movements are different from those to movements of other types. Some tentative implications for the estimation of distributed lag relationships are drawn in a final section.

Since much of the underlying theory is unfamiliar we begin with some preliminary remarks on:

## II. The Elementary Theory of Optimal Extraction and Prediction.[2]

It will be convenient for present purposes to deal with processes which are stationary, at least to second order,[3] and defined at discrete points in time. The restriction to stationary processes will be relaxed occasionally in a very limited fashion. Let such a process be denoted by $\{y_t\}$ where the index $t$ ranges over the positive and negative integers and 0. In 1938, Wold [38] showed that every discrete stationary process $\{y_t\}$ could be decomposed into the sum of two mutually independent processes $\{\xi_t\}$ and $\{\eta_t\}$ such that

$$y_t = \xi_t + \eta_t \,, \tag{2.1}$$

where $\{\xi_t\}$ is the so-called linearly deterministic process which may be predicted with zero mean-square error from all past observations and where $\eta_t$ is a stationary, possibly infinite moving average process:

$$\eta_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} \,, \qquad b_o = 1 \,, \tag{2.2}$$

where

$$\Sigma |b_j|^2 < \infty$$

and

$$E\, \epsilon_t \epsilon_{t'} = \begin{cases} \sigma^2 \,, & t = t' \\ 0 \,, & \text{otherwise} \,. \end{cases}$$ [4]

For a variety of reasons, including realism and economic relevance as well as convenience, our further discussion is restricted to purely

nondeterministic processes of the sort representable by (2.2), i.e., representable in general by a one-sided infinite moving average of what is often called a white noise input.

To facilitate our discussion we introduce the backwards shift operator $U$ defined by

$$U^k y_t = y_{t-k} . \tag{2.3}$$

Then we may write (2.2) as

$$y_t = B(U)\epsilon_t \tag{2.4}$$

where

$$B(U) = \sum_{j=0}^{\infty} b_j U^j .$$

The generating transform or z- transform of a sequence $\{ \ldots a_{-1} , a_0 , a_1 , a_2 , \ldots \}$ is defined as

$$A(z) = \sum_{k=-\infty}^{\infty} a_k z^k \tag{2.5}$$

when the summation on the right converges. (When it does in some region it represents the Laurent expansion of the function $A(z)$ there.) Note that $z$ is complex. Clearly,

$$B(z) = \sum_{j=0}^{\infty} b_j z^j$$

converges in a region encompassing the unit circle and

$$B(z^{-1}) = \sum_{j=0}^{\infty} b_j z^{-j}$$

converges outside a circle inside the unit circle, so that the function $B(z)B(z^{-1})$ is defined and analytic in an annulus about the unit circle. As we shall see this function evaluated on the unit circle is proportional to the spectral density of the time series $y_t$.

The autocovariance function of a stationary time series is defined as

$$c(k) = E\, y_t y_{t-k} \tag{2.6}$$

and is a function only of the lag $k$. For processes which are stationary and contain no purely linearly deterministic component, the autocovariance generating transform exists and is given by

$$g_{yy}(z) = \sum_{k=-\infty}^{\infty} c(k)z^k = \sigma^2 B(z)B(z^{-1}) \tag{2.7}$$

as can be readily deduced from (2.2), (2.4), and (2.6). On the unit circle, i.e., for $z = e^{-i\theta}$, $-\pi \leq \theta \leq \pi$, we have

$$\begin{aligned} g_{yy}(e^{-i\theta}) = 2\pi f_{yy}(\theta) &= \sum_{k=-\infty}^{\infty} c(k)e^{-ik\theta} \\ &= c(0) + 2\sum_{k=1}^{\infty} c(k)\cos k\theta , \end{aligned} \tag{2.8}$$

so we see that on the unit circle the autocovariance generating transform is proportional to the spectral density function. Furthermore,

$$f_{yy}(\theta) = \sigma^2 \left| B(e^{i\theta}) \right|^2 \tag{2.9}$$

so that because $B(z)B(z^{-1})$ is analytic in an annulus containing the unit circle we see that spectral density functions for processes of the type considered are absolutely continuous functions of $\theta$.

Equation (2.9) shows why the representation of $g_{yy}(z)$ as $\sigma^2 B(z)B(z^{-1})$ is often called the canonical factorization of the spectral density function. This factorization must evidentally exist for all processes of the type considered but it may not be unique unless one sets conditions on the zeros of $g_{yy}(z)$ and may not be readily obtainable either unless certain other conditions are met; see Hannan [16, pp. 13-16].

Although all stationary time series with no linearly deterministic component have a one-sided moving average representation, not all have an autoregressive representation. A process defined by a sequence $\{x_t\}$ satisfying

$$A(U)x_t = \epsilon_t \tag{2.10}$$

where $A(U)$ is a polynomial in $U$ and $\epsilon_t$ is a white noise input is called an autoregressive process and may or may not be stationary. In stationary cases it is not necessary to restrict the degree of $A(U)$ to be finite. When $g_{yy}(z)$ is the autocovariance generating transform of a stationary process which has no zeros and is analytic both on the unit circle and in an annulus about the unit circle, then the process $\{y_t\}$ has both a moving average and an autoregressive representation. For, under these circumstances $\log g_{yy}(z)$ is analytic in an annulus about the unit circle and therefore has a Laurent expansion there:

$$\log g_{yy}(z) = \sum_{j=-\infty}^{\infty} c_j z^j = c_o + \sum_{j=1}^{\infty} c_j z^j + \sum_{j=1}^{\infty} c_{-j} z^{-j} . \tag{2.11}$$

Clearly we can take

$$\begin{cases} B(z) = e^{\sum_{j=1}^{\infty} c_j z^j} \\ \sigma^2 = e^{c_o} . \end{cases} \tag{2.12}$$

$B(z)$ is analytic inside a circle with radius greater than 1 and therefore has a Taylor's series expansion there of the form

$$B(z) = \sum_{j=0}^{\infty} b_j z^j \tag{2.13}$$

which yields immediately the moving average representation of the process by equating powers of $z$ with length of lag in (2.2). On the assumption that $\log g_{yy}(z)$ is analytic, we have also that

$$[B(z)]^{-1} = e^{-\sum_{j=1}^{\infty} c_j z^j} = A(z) \tag{2.14}$$

is analytic inside the same circle and thus has a Taylor's series expansion

$$A(z) = \sum_{j=0}^{\infty} a_j z^j \, . \tag{2.15}$$

$A(z)$ is the generating transform of the autoregressive representation. Clearly a necessary condition that $A(z)$ exist for a stationary process whose spectral density in canonical form is $\sigma^2 B(z)B(z^{-1})$ is that $B(z)$ shall have no zeros on the unit circle. Indeed to make the factorization unique we observe that $g_{yy}(z)$ being symmetric has a zero outside the unit circle corresponding to every one inside the unit circle so that we can separate these zeros by appropriate choice of the factors $B(z)$ and $B(z^{-1})$. If this is done so that $B(z)$ has zeros only outside the unit circle the factorization will be unique and $A(z)$ will be given as in (2.14).

Although perfectly acceptable stationary processes such as $y_t = \epsilon_t - \epsilon_{t-1}$ do not possess autoregressive representations, we will

generally suppose throughout the remainder of this paper that the processes with which we deal have both moving average and autoregressive representations.

Of substantial practical importance is the case of a time series with a rational spectral density function. In this case, by definition, the autocovariance generating function may be written as the ratio of two polynominals:

$$g_{yy}(z) = \frac{P(z)}{Q(z)} \,. \tag{2.16}$$

If $Q(z)$ has roots on the unit circle, then $\{y_t\}$ cannot be regarded as stationary for it has no moving average representation. On the other hand, if $Q(z)$ has no such roots, we know that because $g_{yy}(z)$ is symmetric in $z$ and $z^{-1}$, both $P(z)$ and $Q(z)$ must be as well, and, hence, can be factored as

$$g_{yy}(z) = \frac{\sigma^2 \prod_{k=1}^{m} (1 - \beta_k z)(1 - \beta_k z^{-1})}{\prod_{k=1}^{n} (1 - \alpha_k z)(1 - \alpha_k z^{-1})} \,. \tag{2.17}$$

Note $\sigma^2$ has been chosen so that the leading coefficients of $P(z)$ and $Q(z)$ are both one.[5/] In line with the convention mentioned earlier to ensure a unique factorization, we suppose $|\beta_k| \leq 1$ and $|\alpha_k| < 1$. If the strict inequality holds in the first instance the process has an autoregressive as well as a moving average prepresentation; the latter has generating transform

$$(2.18) \qquad B(z) = \frac{\prod_{k=1}^{m} (1 - \beta_k z)}{\prod_{k=1}^{n} (1 - \alpha_k z)}$$

and the former $A(z) = [B(z)]^{-1}$ . Processes with rational spectral density are typically represented as an autoregression equal to a noise input which is not white, i.e.,

$$(2.19) \qquad \prod_{k=1}^{n} (1 - \alpha_k U)\, y_t = \prod_{k=1}^{m} (1 - \beta_k U)\, \epsilon_t$$

Consider now two jointly stationary nondeterministic processes $\{y_t\}$ and $\{x_t\}$ . The $k^{th}$ lag covariance of $y_t$ and $x_t$ , in that order, is given by

$$(2.20) \qquad c_{yx}(k) = E\, y_t x_{t-k} \qquad k = 0 , \pm 1 , \pm 2, \ldots .$$

Note this is different from $E\, y_{t-k} x_t = c_{xy}(k) = c_{yx}(-k)$ . The generating transform of $c_{yx}(k)$ is

$$(2.21) \qquad g_{yx}(z) = \sum_{k=-\infty}^{\infty} c_{yx}(k) z^k$$

and may be termed the cross-covariance generating function inasmuch as its value on the unit circle is proportional to the cross-spectral density function of the series $\{y_t\}$ and $\{x_t\}$ .

Consider the problem of estimating $y_t$ for fixed $t$ given the entire past of the series $\{x_t\}$ up to and including that time. Let us consider only predictors which can be expressed as linear combinations of past $x$'s ,

$$(2.22) \qquad \hat{y}_t = \sum_{j=0}^{\infty} \gamma_j x_{t-j} = \gamma(U) x_t ,$$

and consider optimal that choice of $\gamma_j$ for which

$$E(\hat{y}_t - y_t) \text{ is minimized.}$$

Because $\{x_t\}$ is stationary we can write

$$x_t = B(U)\epsilon_t \tag{2.23}$$

where $\{\epsilon_t\}$ is a white noise sequence with variance $\sigma^2$ and $b_o = 1$. If $\varphi(z)$ represents the generating transform

$$\varphi(z) = \gamma(z)B(z) \tag{2.24}$$

we can express $\hat{y}_t$ in terms of the past of $\{\epsilon_t\}$ to $t$:

$$\hat{y}_t = \varphi(U)\epsilon_t = \sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j} \tag{2.25}$$

It is, in fact, more convenient to find $\varphi(z)$ or $\hat{y}_t$ in the form (2.25), then determine $\gamma(z)$ from (2.24), provided $[B(z)]^{-1}$ exists, and so express $\hat{y}_t$ in the form (2.22).

Under the minimum mean-square-error criterion, we seek to minimize

$$\begin{aligned} E(\hat{y}_t - y_t)^2 &= E[\sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j} - y_t]^2 \\ &= \text{var}(y) + \text{var}(\sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}) - 2\,\text{cov}(\sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}, y_t) \\ &= \text{var}(y) + \sigma^2 \sum_{j=0}^{\infty} \varphi_j^2 - 2\sum_{j=0}^{\infty} c_j \varphi_j \end{aligned} \tag{2.26}$$

where $c_j = E(y_t - Ey_t)\epsilon_{t-j}$. Completing the square, we obtain

$$(2.27)\qquad E(\hat{y}_t - y_t)^2 = \mathrm{var}(y) + \sigma^2 \sum_{j=0}^{\infty} \left[\varphi_j - \left(\frac{c_j}{\sigma^2}\right)\right]^2 - \frac{1}{\sigma^2}\sum_{j=0}^{\infty} c_j^2$$

$$\geq \mathrm{var}(y) - \frac{1}{\sigma^2}\sum_{j=0}^{\infty} c_j^2$$

with equality only for $\varphi_j = c_j/\sigma^2$. Whence

$$(2.28)\qquad \min E(\hat{y}_t - y_t)^2 = \mathrm{var}(y) - \sigma^2 \sum_{j=0}^{\infty} \varphi_j^2 .$$

The following notation will be used extensively throughout the remainder of this paper: If $\{\ldots h_{-1}, h_0, h_1, \ldots\}$ is a sequence with generating transform

$$H(z) = \sum_{j=-\infty}^{\infty} h_j z^j$$

we denote by $[H(z)]_+$ that part of $H(z)$ having only nonnegative powers of $z$, i.e.,

$$[H(z)]_+ = \sum_{j=0}^{\infty} h_j z^j$$

and that part having only negative powers by

$$[H(z)]_- = \sum_{j=-\infty}^{-1} h_j z^j .$$

Using this notation we see that (2.27) implies

$$(2.29)\qquad \varphi(z) = \frac{1}{\sigma^2}\sum_{j=0}^{\infty} c_j z^j = \frac{1}{\sigma^2}\left[g_{y\epsilon}(z)\right]_+ .$$

We may assume $\{y_t\}$ has zero mean without loss of generality. Then

$$(2.30) \qquad g_{yx}(z) = \sum_{k=-\infty}^{\infty} (E\, y_t x_{t-k}) z^k$$

$$= \sum_{k=-\infty}^{\infty} z^k E\, y_t \sum_{j=0}^{\infty} b_j \epsilon_{t-j-k}$$

$$= \sum_{k=-\infty}^{\infty} z^k \sum_{j=0}^{\infty} b_j E\, y_t \epsilon_{t-j-k}$$

$$= \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} b_j z^{-j} z^{j+k} c_{j+k}$$

$$= g_{y\epsilon}(z)\, B(z^{-1})$$

Hence,

$$(2.31) \qquad \varphi(z) = \frac{1}{\sigma^2} \left[ \frac{g_{yx}(z)}{B(z^{-1})} \right]_+$$

or, using (2.24), and assuming the process has an autoregressive representation,

$$(2.32) \qquad \gamma(z) = \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{yx}(z)}{B(z^{-1})} \right]_+ .$$

Equation (2.32) is the fundamental formula for optimal signal extraction and prediction. To obtain the result for prediction we set $y_t = x_{t+\nu}$, $\nu > 0$. Then

$$(2.33) \qquad g_{yx}(z) = \sum_{k=-\infty}^{\infty} z^k E\, x_{t+\nu} x_{t-k}$$

$$= z^{-\nu} \sum_{k=-\infty}^{\infty} z^{k+\nu} c(k+\nu)$$

$$= \frac{\sigma^2 B(z) B(z^{-1})}{z^{\nu}} ,$$

whence

$$(2.34) \qquad \gamma(z) = \frac{1}{\sigma^2 B(z)} \left[ \frac{\sigma^2 B(z)B(z^{-1})}{B(z^{-1})z^{\nu}} \right]_+$$

$$= \frac{1}{B(z)} \left[ \frac{B(z)}{z^{\nu}} \right]_+ ,$$

where $\gamma(z)$ is now the generating transform for the $\nu$-step predictor. There is no need to restrict $\{y_t\}$ to observable series, it is only that we have sufficient information about its stochastic properties to be able to specify $g_{yx}(z)$. Equation (2.32) then expresses the generating transform for the estimate $\hat{y}_t$. If $y_t$ is an unobserved component of the series $x_t$, we say $\gamma(z)$ is the generating transform of the optimal extraction.

Clearly, if $\{y_t\}$ is an unobserved component of an observed time series to an estimate of which an economic agent reacts, we can determine a distributed lag model by means of (2.32). Thus, suppose the behavioral relation is

$$(2.35) \qquad h_t = f(\hat{y}_t)$$

where $f$ may include a stochastic disturbance. If we assume $\hat{y}_t$ is determined by optimal extraction from the past of the observed series $\{x_t\}$, we have

$$(2.36) \qquad h_t = f\left( \sum_{j=0}^{\infty} \gamma_j x_{t-j} \right)$$

where $\gamma_j$ is determined by (2.32). Equation (2.36) is the distributed lag relation sought.

## III. Derivation of Some Distributed Lag Models by Means of Optimal Extraction

In this section, we derive various distributed lag models along the lines suggested above for a variety of stochastic models relating the observed time series to its unobserved components over time. We shall examine a simple two-component model in considerable detail, showing that in one special case we get a geometrical lag distribution. When this model is allowed to become nonstationary in a specified way adaptive expectations result and the $\nu$-step predictor of the observed series becomes identical with the $\nu$-step extractor of the unobserved component. A generalization of this simple model leads to a lag distribution related to Solow's Pascal lag distribution. Finally, a three-component model and related behavioral equation are considered in a fair degree of generality. This model is capable of allowing differing reactions to what might be termed "seasonal" and "long-term" movements of the causative variable.

Consider the following model:

$$(3.1) \qquad \begin{cases} x_t = y_t + u_t \\ y_t = \alpha y_{t-1} + v_t \end{cases}, \qquad |\alpha| < 1 ,$$

where $\{u_t\}$ and $\{v_t\}$ are independent zero-mean white noise inputs with variance ratio

$$(3.2) \qquad \lambda = \frac{E\, v_t^2}{E\, u_t^2} = \frac{\sigma_v^2}{\sigma_u^2}$$

To apply (2.32), we must determine $B(z)$, $\sigma^2$, and $g_{yx}(z)$.

First, let us determine $B(z)$ and $\sigma^2$. From (3.1), we see

$$x_t - \alpha x_{t-1} = y_t - \alpha y_{t-1} + u_t - \alpha u_{t-1} \tag{3.3}$$

or

$$(1 - \alpha U)x_t = v_t + (1 - \alpha U)\, u_t \,. \tag{3.3'}$$

Because of the assumed independence of $v_t$ and $u_t$ and their assumed lack of serial dependence, it follows straightforwardly that

$$(1 - \alpha z)(1 - \alpha z^{-1})g_{xx}(z) = \sigma_v^2 + (1 - \alpha z)(1 - \alpha z^{-1})\, \sigma_u^2 \,, \tag{3.4}$$

so that

$$g_{xx}(z) = \sigma_u^2 \left[ \frac{\lambda + (1 - \alpha z)(1 - \alpha z^{-1})}{(1 - \alpha z)(1 - \alpha z^{-1})} \right] . \tag{3.5}$$

In order to express this in the form $\sigma^2 B(z)B(z^{-1})$, it is necessary to factor the polynomial $\lambda + (1 - \alpha z)(1 - \alpha z^{-1})$ which appears in the numerator on the right.

When $\alpha > 0$ the two roots of this polynomial are

$$\begin{cases} \beta = \dfrac{(1 + \lambda + \alpha^2) - \sqrt{(1 + \lambda + \alpha^2)^2 - 4\alpha^2}}{2\alpha} < 1 \\ \dfrac{1}{\beta} = \dfrac{(1 + \lambda + \alpha^2) + \sqrt{(1 + \lambda + \alpha^2)^2 - 4\alpha^2}}{2\alpha} > 1 \,. \end{cases} \tag{3.6}$$

When $\alpha < 0$ the inequalities are reversed, bearing this in mind, however, there is no loss in generality in taking $\alpha > 0$. Thus $\beta$ in (3.6)

is the root lying inside the unit circle. Thus we take

$$B(z) = \frac{1 - \beta z}{1 - \alpha z} \tag{3.7}$$

To determine the constant $\sigma^2$, observe that, since $\beta$ and $\frac{1}{\beta}$ are roots of $\lambda + (1 - \alpha z)(1 - \alpha z^{-1})$,

$$\lambda = -(1 - \alpha\beta)(1 - \alpha/\beta) \tag{3.8}$$

Inserting this in the polynomial and equating like powers of $z$ in

$$(1 - \alpha z)(1 - \alpha z^{-1}) - (1 - \alpha\beta)(1 - \alpha/\beta) = c(1 - \beta z)(1 - \beta z^{-1}),$$

we easily deduce that $c = \alpha/\beta$, so that

$$\sigma^2 = \frac{\alpha}{\beta}\ \sigma_u^2 . \tag{3.9}$$

Next, let us determine $g_{yx}(z)$. In so doing, we use the independence of $u_t$ and $v_{t'}$ for all $t, t'$, which in turn implies the independence of $y_t$ and $u_{t'}$ for all $t, t'$. Lagging the first of equations (3.1) by $k$ periods multiplying by $y_t$ and taking expected values we deduce

$$\begin{aligned} g_{yx}(z) &= g_{yy}(z) + 0 \\ &= \frac{\sigma_v^2}{(1 - \alpha z)(1 - \alpha z^{-1})} \end{aligned} \tag{3.10}$$

where the last equality follows from the final equation of (3.1). $\sigma_v^2 = \lambda\sigma_u^2$ so substituting for $B(z)$, $\sigma^2$, and $g_{yx}(z)$ in (2.32) we obtain

$$(3.11)\qquad \gamma(z) = \frac{1}{\sigma^2 B(z)}\left[\frac{g_{yx}(z)}{B(z^{-1})}\right]_+$$

$$= \frac{\beta\lambda(1-\alpha z)}{\alpha(1-\beta z)}\left[\frac{1}{(1-\alpha z)(1-\beta z^{-1})}\right]_+$$

$$= -\frac{\beta(1-\alpha\beta)(1-\alpha/\beta)(1-\alpha z)}{\alpha(1-\beta z)(1-\alpha\beta)}\left[\frac{1}{1-\alpha z} + \frac{\beta z^{-1}}{1-\beta z^{-1}}\right]_+ .$$

Now the second term in the $[\ ]_+$ sign contains only negative powers of $z$, whereas the first contains only positive powers of $z$, so that the $[\ ]_+$ operator annihilates the second term and does not affect the first. Consequently, we find at last that

$$(3.12)\qquad \gamma(z) = \frac{\alpha-\beta}{\alpha}\cdot\frac{1}{1-\beta z}$$

$$= \frac{\alpha-\beta}{\alpha}\sum_{j=0}^{\infty}\beta^j z^j$$

since $\sum_{j=0}^{\infty}\beta^j z^j$ converges in a circle containing the unit circle by virtue of the choice of $\beta$ such that $|\beta| < 1$.[6/]

It follows from equation (3.12) that the optimal extraction $\hat{y}_t$ is expressible as a geometrically weighted average of past observed values of $x_t$:

$$(3.13)\qquad \hat{y}_t = \frac{\alpha-\beta}{\alpha}\sum_{j=0}^{\infty}\beta^j x_{t-j} .$$

It is interesting to note that by differentiating (3.8) with respect to $\lambda$ holding $\alpha$ constant we obtain

$$(3.14)\qquad \frac{\partial\beta}{\partial\lambda} = \frac{1}{\alpha(1-\frac{1}{\beta^2})} < 0$$

since we assume $\alpha > 0$ and so $\beta < 1$. The expression (3.14) shows that as the ratio $\lambda = \sigma_v^2 / \sigma_u^2$ increases less weight is given to observations in the distant past and more to recent observations. If we interpret, as must now seem plausible, $y_t$ as the "permanent" component of, say, income, and $u_t$ as the "transitory" component, (3.13) and (3.14) together assert that the greater the proportion of variance of measured income contributed by the transitory component, the smaller $\lambda$, and the less weight will be given to recent values of measured income in the calculation of each estimate of permanent income.

By subtracting $\alpha\hat{y}_{t-1}$ from both sides of (3.13) we arrive at an equation relating values of $\hat{y}_t$ over time which bears a remarkable resemblance to the adaptive expectations model earlier proposed for the generation of distributed lags

$$\hat{y}_t - \alpha\hat{y}_{t-1} = \frac{\alpha - \beta}{} \{x_t - \alpha\hat{y}_{t-1}\} . \tag{3.15}$$

Indeed, if we let $\alpha$ tend to one from below [7] (3.15) reduces to the familiar form of the adaptive expectations model with coefficient of expectations $(1-\beta)$ :

$$\hat{y}_t - \hat{y}_{t-1} = (1 - \beta) \{x_t - \hat{y}_{t-1}\} . \tag{3.16}$$

Of course, when $\alpha = 1$ the model (3.1) yields a nonstationary time series, but this may be reasonable for economic applications.[8] Furthermore, one can show that all results derived for stationary models can be extended to nonstationary models of this limiting variety in exactly the fashion here proposed.

When we consider the nonstationary model, (3.1) with $\alpha = 1$, another remarkable conclusion emerges. Consider the minimum mean-square-error forecast of $x_{t+\nu}$. Let the generating transform of the weights applied to $x_t$, $x_{t-1}$, .... , be

$$(3.17) \qquad \theta_\nu(z) = \frac{1}{B(z)} \left[ \frac{B(z)}{z^\nu} \right]_+$$

$$= \left( \frac{1 - \alpha z}{1 - \beta z} \right) \left[ \frac{1 - \beta z}{(1 - \alpha z) z^\nu} \right]_+$$

$$= \left( \frac{1 - \alpha z}{1 - \beta z} \right) \left[ z^{-\nu} \left( \sum_{j=0}^{\infty} \alpha^j z^j - \beta \sum_{j=0}^{\infty} \alpha^j z^{j+1} \right) \right]_+$$

$$= \frac{1 - \alpha z}{1 - \beta z} \cdot \frac{\alpha^\nu (1 - \beta/\alpha)}{1 - \alpha z}$$

$$= \left( \frac{\alpha - \beta}{\alpha} \right) \alpha^\nu \sum_{j=0}^{\infty} \beta^j z^j .$$

When $\alpha = 1$ this reduces to $(1 - \beta) \sum_{j=0}^{\infty} \beta^j z^j$ which is exactly the formula for the optimal extraction $\hat{y}_t$ when $\alpha = 1$. Thus, for this simple nonstationary model, we see that <u>the same weighted moving average of past observed values $x_t$ gives both the optimal prediction for all future periods and the optimal extraction for the current period as well as all future periods.</u> 2/

In [25, pp. 51-59], I proposed that farmers base planned output not on prices currently prevailing at time of planting but rather on "expected normal price" which I interpreted essentially as

an average of future prices expected to prevail. Furthermore, in the course of analyzing the results it was found that the coefficient of expectation ($1 - \beta$ in the above formulation) varied inversely with the variance of actual prices for the three crops considered [25, p. 221]. On the assumption that increasing observed variance is associated with a greater than proportional increase in the variance of the "transitory component" in prices, this result is implied by (3.14). Thus (3.1) with $\alpha = 1$ provides an extraordinarily strong justification of the "expected normal price" model. Since, however, we would not expect farmers to base decisions on a stochastic model of price behavior very far out of line with reality, the present discussion also suggests a further test of the hypothesis advanced in [25]: namely, to explore the degree to which (3.1) with $\alpha = 1$ actually fits annual data on agricultural prices for the period prior to 1933.

Although (3.1) leads to the most common form of distributed lag employed in current studies, it represents an extremely simplified scheme as far as most economic time series are concerned. The results of Orcutt [31] and Ames and Reiter [3] suggest that autoregressive schemes of higher than first order may provide a better description of the slower moving components of many important economic time series. In what follows, we derive a distributed lag scheme from an optimal extraction of a component which obeys an autoregressive scheme of arbitrarily high, but finite order. A specialization of this result leads to a distributed lag which is similar in some respects to the Solow Pascal distribution.

Consider the following model:

$$(3.18)\qquad \begin{cases} x_t = y_t + u_t \\ Q(U)y_t = v_t \end{cases}$$

where $Q(U)$ is a polynomial of degree $n$ in the backward shift operator $U$ with generating transform

$$(3.19)\qquad Q(z) = \prod_{j=1}^{n} (1 - \alpha_j z)$$

such that $|\alpha_j| < 1$ for all $j$. Note that $Q(z)$ is normalized so that the coefficient of the current value of $y$ in $Q(U)y_t$ is one. The restriction $|\alpha_j| < 1$ for all $j$ ensures that the roots of $Q(z)$ all lie <u>outside</u> the unit circle; hence, if we take $\{u_t\}$ and $\{v_t\}$ as independent, zero-mean white noise inputs with variance ratio

$$\lambda = \frac{E\, v_t^2}{E\, u_t^2} = \frac{\sigma_v^2}{\sigma_u^2} \quad ,$$

the series $\{y_t\}$ will have a one-sided moving average representation and $\{x_t\}$ will be a stationary process. Should some roots of $Q(z)$ lie <u>on</u> the unit circle the process $\{x_t\}$ will be nonstationary but our results will hold if we let some of the $\alpha_j$'s tend to values for which $|\alpha_j| = 1$.

To determine the distributed lag implied by (3.18), we determine the minimum mean-square-error extraction of $y_t$ by the formula given above, equation (2.32). First, note that

$$(3.20) \qquad Q(U)x_t = Q(U)y_t + Q(U)u_t$$
$$= v_t + Q(U)u_t$$

so that, by methods used above, we deduce

$$(3.21) \qquad g_{xx}(z) = \sigma_u^2 \left\{ \frac{\lambda + Q(z)Q(z^{-1})}{Q(z)Q(z^{-1})} \right\} .$$

Consider the polynomial

$$(3.22) \qquad P(z) = \lambda + Q(z)Q(z^{-1}) = \lambda + \prod_{j=1}^{n} (1 - \alpha_j z)(1 - \alpha_j z^{-1}) .$$

Clearly if $z_o$ is a root, so is $1/z_o$. There are thus $2n$ roots which come in reciprocal pairs: one member of each pair lies <u>inside</u> the the unit circle, the other <u>outside</u>. Let the roots be $\beta_j$ and $1/\beta_j$, $j = 1, \ldots, n$, and let $\beta_j$ be the member of each pair lying <u>inside</u> the unit circle. Thus factor $P(z)$ as

$$(3.23) \qquad P(z) = c \prod_{j=1}^{n} (1 - \beta_j z)(1 - \beta_j z^{-1}) ,$$

where $c$ is a constant of proportionality so chosen as to make coefficients for corresponding powers of $z$ in (3.22) and (3.23) equal. Since $\beta_k$, $k = 1, \ldots, n$ is a root of $P(z)$, we have

$$(3.24) \qquad \lambda = - \prod_{j=1}^{n} (1 - \alpha_j \beta_k)(1 - \alpha_j/\beta_k) , \quad k = 1, \ldots, n .$$

It may thus be readily verified that $c$ should be chosen so that

$$(3.25) \qquad c = \frac{\lambda + \prod_{j=1}^{n}(1+\alpha_j^2)}{\prod_{j=1}^{n}(1+\beta_j^2)} = \frac{\prod_{j=1}^{n}(1+\alpha_j^2) - \prod_{j=1}^{n}(1-\alpha_j\beta_k)(1-\alpha_j/\beta_k)}{\prod_{j=1}^{n}(1+\beta_j^2)}$$

where $k$ may take on any value from 1 to $n$.[10/] Following the convention, described above, that we take both numerators and denominators in the canonical factorization of a rational spectral density to have roots <u>outside</u> the unit circle, we choose

$$(3.26) \qquad B(z) = \prod_{j=1}^{n}\left(\frac{1-\beta_j z}{1-\alpha_j z}\right), \quad |\beta_j| < 1, \ |\alpha_j| < 1,$$

and

$$(3.27) \qquad \sigma^2 = \sigma_u^2 c = \sigma_u^2 \left\{ \frac{\prod_{j=1}^{n}(1+\alpha_j^2) - \prod_{j=1}^{n}(1-\alpha_j\beta)(1-\alpha_j/\beta)}{\prod_{j=1}^{n}(1+\beta_j^2)} \right\}$$

where $\beta$ is any root of $P(z)$.

To find $g_{yx}(z)$ from (3.18), we again make use of the fact that $u_t$ and $v_{t'}$ are independent for all $t$ and $t'$, so that $y_t$ and $u_{t'}$ are independent for all $t$ and $t'$. Thus,

$$(3.28) \qquad g_{yx}(z) = g_{yy}(z) = \frac{\sigma_v^2}{Q(z)Q(z^{-1})} = \frac{\sigma_v^2}{\prod_{j=1}^{n}(1-\alpha_j z)(1-\alpha_j z^{-1})}.$$

Applying (2.32), we immediately have

$$(3.29)\qquad \gamma(z) = \left(\frac{\lambda}{c}\right) \cdot \prod_{j=1}^{n} \left(\frac{1 - \alpha_j z}{1 - \beta_j z}\right) \left[ \frac{\prod_{j=1}^{n} (1 - \alpha_j z^{-1})}{\prod_{j=1}^{n} (1 - \alpha_j z)(1 - \alpha_j z^{-1})(1 - \beta_j z^{-1})} \right]_+$$

$$= \left(\frac{\lambda}{c}\right) \prod_{j=1}^{n} \left(\frac{1 - \alpha_j z}{1 - \beta_j z}\right) \left[ \frac{1}{\prod_{j=1}^{n} (1 - \alpha_j z)(1 - \beta_j z^{-1})} \right]_+ ,$$

where $\lambda$ and $c$ are given by (3.24) and (3.25) respectively. Unfortunately, the expression inside and including the operator $[\ ]_+$ is not so very easy to evaluate explicitly. If $\alpha_j$, $j = 1, \ldots, n$, and $\lambda$ were known, $\beta_j$, $j = 1, \ldots, n$, could be found numerically; then the expression inside the operator $[\ ]_+$ could be expanded in partial fractions, the coefficients of which might easily be determined by numerical methods. Having separated terms involving $\frac{1}{1 - \alpha_j z}$ and $\frac{1}{1 - \beta_j z^{-1}}$ it would be relatively easy to obtain that portion of the expression involving only nonnegative powers of $z$. Analytically, however, it is difficult to proceed beyond (3.29) at this level of generality. Suffice it to say, $\gamma(z)$ in (3.29) is the generating transform of a rather general lag distribution which includes the geometric lag distribution as a special case. It also includes an interesting special case to be derived next.

Suppose that $Q(z)$ in (3.19) can be written

$$(3.30) \qquad Q(z) = (1 - \alpha z)^n , \quad |\alpha| < 1 .$$

Thus $Q(U)$ is what I have called elsewhere [28, p. 257] an nth-order quasi-difference with parameter $\alpha$. It has been found that, apart from seasonality, quasi-differences of order 1, 2, or 3 with parameter $\alpha = 0.75$ reduce many economic time series to a good approximation to white noise.[11] It is thus not implausible to assume that the long-term or slowly moving components of many economic time series can be well approximated by a relatively low-order quasi-difference scheme of the form:

$$(1 - \alpha U)^n y_t = v_t ,$$

where $v_t$ is white-noise. In this case the polynomial $P(z)$ in (3.22) which must be factored in order to obtain $B(z)$ takes the form

$$(3.31) \qquad P(z) = \lambda + [(1 - \alpha z)(1 - \alpha z^{-1})]^n .$$

Although $Q(z)Q(z^{-1}) = 0$ has only two distinct roots, it is not in general true that $P(z) = 0$ will have only this limited number. The $2n$ roots of $P(z)$ will, however, come in pairs $\beta_j$ and $1/\beta_j$, one member of which lies inside the unit circle and the other outside. As before let $\beta_j$ be the root lying inside and factor $P(z)$ as

$$(3.32) \qquad P(z) = c \prod_{j=1}^{n} (1 - \beta_j z)(1 - \beta_j z^{-1}) ,$$

where now

$$(3.33) \qquad c = \frac{\lambda + (1 + \alpha^2)^n}{\prod_{j=1}^{n} (1 + \beta_j^2)}$$

and

$$(3.34) \qquad \lambda = -\prod_{j=1}^{n}(1 - \alpha/\beta_k)(1 - \alpha\beta_k) \qquad k = 1, \ldots, n$$

$$= -[(1 - \alpha/\beta)(1 - \alpha\beta)]^n,$$

where $\beta$ is any root of $P(z)$, we now have

$$(3.35) \qquad B(z) = \frac{\prod_{j=1}^{n}(1 - \beta_j z)}{(1 - \alpha z)^n}$$

and

$$(3.36) \qquad \sigma^2 = \sigma_u^2 c$$

with $c$ given by (3.33), and

$$(3.37) \qquad g_{yx}(z) = \frac{\sigma_v^2}{[(1 - \alpha z)(1 - \alpha z^{-1})]^n} \cdot$$

Applying (3.32), we find the transform for the optimal extraction to be

$$(3.38) \qquad \gamma(z) = \left(\frac{\lambda}{c}\right) \frac{(1 - \alpha z)^n}{\prod_{j=1}^{n}(1 - \beta_j z)} \left[ \frac{1}{(1 - \alpha z)^n \prod_{j=1}^{n}(1 - \beta_j z^{-1})} \right]_+$$

The expression inside and including the operator $[\ ]_+$ may be evaluated by means of a highly useful theorem due to Whittle (Theorem 1, [37, p. 93]):

THEOREM. Let $R(z)$ be analytic in an annulus about the unit circle and let $\theta$ be a parameter such that $|\theta| < 1$. Then

$$\Phi(z) = (1 - \theta z)^p \left[ \frac{R(z)}{(1 - \theta z)^p} \right]_+$$

$$= \Pi_p(z) + [R(z)]_+ ,$$

where $\Pi_p(z)$ is a polynomial of degree $p-1$ in $z$ such that

$$\Pi_p(z) = \sum_{j=0}^{p-1} \pi_j (z - \theta^{-1})^j .$$

The coefficients $\pi_j$ are determined by

$$\pi_j = \frac{1}{j!} \left. \frac{d^j([R(z)]_-)}{dz^j} \right|_{z = \theta^{-1}} ,$$

where the operator $[\ ]_-$ means we consider a function defined by taking only the negative powers in the Laurent expansion of $R(z)$ in the annulus about the unit circle.

To apply this theorem we set up the following correspondences:

$$\text{(3.39)} \qquad \begin{cases} p = n \\ \theta = \alpha \\ R(z) = \dfrac{1}{\prod_{j=1}^{n} (1 - \beta_j z^{-1})} = \sum_{k=0}^{\infty} r_k z^{-k} . \end{cases}$$

The power series on the right hand side of (3.39) is obtained by expanding each term

$$\frac{1}{1 - \beta_j z^{-1}} = \sum_{k=0}^{\infty} \beta_j^k z^{-k}$$

which can surely be done in an annulus about the unit circle since $| \beta_j | < 1$ .[12] It follows that

$$(3.40) \qquad \left[ R(z) \right]_+ = \left[ \sum_{k=0}^{\infty} r_k z^{-k} \right]_+ = 1 ,$$

and that

$$(3.41) \qquad \Pi_p(z) = \sum_{j=0}^{n-1} \pi_j (z - \alpha^{-1})^j$$

with

$$(3.42) \qquad \pi_j = \frac{1}{j!} \left. \frac{d^j \left\{ \sum_{k=1}^{\infty} r_k z^{-k} \right\}}{dz^j} \right|_{z = \alpha^{-1}}$$

$$= (-1)^j \sum_{k=1}^{\infty} \binom{k + j - 1}{j} r_k \alpha^{k+j}$$

If we let

$$g_j = (-1)^j \alpha^{-j} \pi_j$$

we can then write $\gamma(z)$ in (3.38) as

$$(3.43) \qquad \gamma(z) = \frac{(\lambda/c)}{\prod_{j=1}^{n}(1-\beta_j z)} \left\{ 1 + \sum_{j=0}^{n-1} g_j (1-\alpha z)^j \right\}$$

If $P(z)$ had had only two distinct roots, the first term in this expression would be the generating transform of the weights in a Pascal distribution; Solow [33, p. 396]. In general, however, there will be more distinct roots. The first term will still represent a "cascaded" series of geometric lags, but one more general than the Pascal case. The presence of the second term, adds a further degree of complexity, for this term implies that the first $n$ weights (counting the zero order lag as the first), will differ still more from those of a Pascal distribution. It does not appear to be true that, as $\alpha \to 1$, the distribution of lag (3.43) tends to a Pascal distribution, but I have been unable to find a convenient form for the limit.[13/] In any event, the case $\alpha = 1$ is surely not highly plausible from an economic standpoint, and for this reason the more complicated form (3.43) may be preferable. Furthermore, the estimation problems raised by a distribution of lag defined by (3.43) are very similar to those encountered in estimating a Pascal distributed lag. In the next section, we will see that the advantages for estimation of the more complicated form arising from its interpretation as an optimal extraction outweigh the advantages of using the simpler Pascal form.

Economic time series consisting of observations taken at intervals of less than a year, say at quarterly or monthly intervals, exhibit a much more complicated sort of behavior than is easily

describable in terms of the simple two-component model discussed above. In particular, if $L$ observations per year are taken at equal intervals ($L=4$ for quarterly data, $L=12$ for monthly), the spectral density will nearly always exhibit peaks of a variety of widths at the so-called seasonal frequencies: $\frac{2\pi k}{L}$, $k = 1, 2, \ldots, L/2$ if $L$ is even, or $k = 1, 2, \ldots, (L-1)/2$ if $L$ is odd. While such behavior can be generated by relatively high-order autoregressive schemes, it has been found easier and more useful to explain such behavior in terms of three-component models having separate components for predominantly low frequency movements, often termed "trend-cycle," and seasonal movements characterized by a spectral density with peaks of more-or-less equal height at seasonal frequencies. The paper by Couts, Grether, and Nerlove [6] analyzes a nonstationary model of this type and applies it to U. S. unemployment data; Grether [13] analyzes a related stationary model.

The possibility of introducing a third component in the derivation of distributed lags also introduces an interesting solution to an old problem in the application of the standard results to monthly or quarterly data. It has always been something of a puzzle whether or not the simple types of distributed lag models could be directly applied to data containing an important seasonal component. The problem has typically been handled either by using seasonally adjusted data or by introducing dummy variables into the behavioral relation to be estimated; see Wallis [36, pp. 17-19 and pp. 48-51]. Neither of these solutions appears to be completely satisfactory. While it is, in fact, possible to investigate rather complicated lag relations using seasonally

unadjusted data by means of spectral techniques, as Wallis does, a more promising approach is suggested in the first section of this paper: Once we divide a time series into several unobserved components and derive a distributed lag relationship by assuming that economic agents react not to the observed series but to an estimate of one of the unobserved components, it is a short step to the hypothesis that economic agents react in distinct ways to several estimates, each of a different unobserved component. Thus, for example, consider the following model:

$$(3.44) \qquad \begin{cases} x_t = y_t + s_t + u_t \\ h_t = a\hat{y}_t + b\hat{s}_t + e_t \,, \end{cases}$$

where $x_t$ is observed and $y_t$, $s_t$, and $u_t$ are unobserved components which have specified stochastic properties making it plausible, say, to interpret them as "trend-cycle," "seasonal," and "irregular." The variables $\hat{y}_t$ and $\hat{s}_t$ are estimates of $y_t$ and $s_t$, respectively, based upon the assumed stochastic structure. The second equation of (3.44) simply states that the variable $h_t$, which might be inventory investment for example, is differently affected by firms' estimates of "trend-cycle" and their estimates of the "seasonal." $e_t$ is a disturbance in this relationship.

Following the same procedure adopted earlier in this paper, we will derive a distributed lag relationship between $h_t$ and current and past values of $x_t$ by replacing $\hat{y}_t$ and $\hat{s}_t$ by their optimal extractions based on an assumed stochastic structure. The main difference between

the final result in this more complicated case and our earlier results based upon a two-component model, is that the former now involves the parameters of the behavioral relationship in an essential way. Indeed, as will be brought out in the next section, one might plausibly estimate such a relationship by creating two distinct time series, $\hat{y}_t$ and $\hat{s}_t$, from the single observed series, $x_t$, and using both as independent variables in a statistical analysis. Alternatively, knowledge of the optimal extractors in parametric form may lead to transformations of the equation to be estimated.

Suppose then that we augment the first equation of (3.44) by two more and certain other assumptions which suffice to determine completely the stochastic structure of the observed series $x_t$ :

$$(3.45) \qquad \begin{cases} x_t = y_t + s_t + u_t \\ Q(U)yt = v_t \\ S(U)s_t = w_t \ , \end{cases}$$

where $\{u_t\}$, $\{v_t\}$, and $\{w_t\}$ are independent, zero-mean white noise inputs characterized by variance ratios

$$\lambda = \frac{Ev_t^2}{Eu_t^2} = \frac{\sigma_v^2}{\sigma_u^2} \ ,$$

and

$$\mu = \frac{Ew_t^2}{Eu_t^2} = \frac{\sigma_w^2}{\sigma_u^2}$$

and where $Q(U)$ and $S(U)$ are polynomials in the backward shift operator $U$ chosen so as to make the interpretation of $y_t$ and $s_t$ as "trend-cycle" and "seasonal" plausible. At the present time, the details of just how such a choice should be made are the subject of intensive investigation. However, in general, it seems clear that $Q(U)$ should be a relatively low-order polynomial in $U$ such that the spectral density of $y_t$ will have relatively high power near the origin falling off not too sharply with increasing frequency, and $S(U)$ should be roughly the same sort of polynomial, not in $U$ itself, but rather in $U^L$, where $L$ is the number of times per year we observe $x_t$. Grether [13], for example, has analyzed models of the form (3.45) with

$$(3.46) \qquad \begin{cases} Q(U) = (1 - \alpha U)^n \\ S(U) = (1 - \theta U^L)^m \end{cases}$$

with $m = n = 2$.[14/]

In the three-component model there are three possible extractions, $\hat{y}_t$, $\hat{s}_t$, and $\hat{u}_t$, and, hence, there are three generating transforms relating these estimates to the current and past values of the observed variable $x_t$:

$$(3.47) \qquad \gamma(z) = \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{yx}(z)}{B(z^{-1})} \right]_+ ,$$

the generating transform for the weights in

$$\hat{y}_t = \sum_{j=0}^{\infty} \gamma_j x_{t-j} ,$$

$$(3.48) \qquad \varphi(z) = \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{sx}(z)}{B(z^{-1})} \right]_+ ,$$

the generating transform for the weights in

$$\hat{s}_t = \sum_{j=0}^{\infty} \varphi_j x_{t-j} ,$$

and

$$\psi(z) = \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{ux}(z)}{B(z^{-1})} \right]_+ , \tag{3.49}$$

the generating transform for the weights in

$$\hat{u}_t = \sum_{j=0}^{\infty} \psi_j x_{t-j} .$$

As before $B(z)$ and $\sigma^2$ are the appropriate factors in the canonical factorization of $g_{xx}(z)$. Of course, only $\gamma(z)$ and $\varphi(z)$ are relevant directly since only $\hat{y}_t$ and $\hat{s}_t$ enter the behavioral relationship under consideration. Nonetheless, $\psi(z)$ is of interest because it can be shown that

$$\gamma(z) + \varphi(z) + \psi(z) = 1 . \tag{3.50}$$

This follows because

$$g_{xx}(z) = g_{yx}(z) + g_{sx}(z) + g_{ux}(z) , \tag{3.51}$$

on account of the assumed independence of $\{u_t\}$, $\{v_t\}$, and $\{w_t\}$, so that

$$\frac{1}{\sigma^2 B(z)} \left[ \frac{g_{xx}(z)}{z^{\nu} B(z^{-1})} \right]_+ = \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{yx}(z)}{z^{\nu} B(z^{-1})} \right]_+ + \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{sx}(z)}{z^{\nu} B(z^{-1})} \right]_+ + \frac{1}{\sigma^2 B(z)} \left[ \frac{g_{ux}(z)}{z^{\nu} B(z^{-1})} \right]_+ . \tag{3.52}$$

Now the expression on the left is the generating transform of the optimal $\nu$-step predictor for $x_t$, while the three expressions on the right are the generating transforms for the optimal $\nu$-step predictors of the components $y_t$, $s_t$, and $u_t$. For $\nu = 0$, of course, these are nothing more than the optimal extractions of $\hat{y}_t$, $\hat{s}_t$, and $\hat{u}_t$. For $\nu = 0$, however, the optimal predictor of $x_t$ must be $x_t$ itself; clearly

$$\frac{1}{\sigma^2 B(z)} \left[ \frac{\sigma^2 B(z) B(z^{-1})}{B(z^{-1})} \right]_+ = 1 .$$

Thus, (3.52) implies (3.50) when $\nu = 0$.

As before, the independence of $\{u_t\}$, $\{v_t\}$, and now $\{w_t\}$ implies

$$(3.53) \qquad \begin{cases} g_{yx}(z) = g_{yy}(z) = \sigma_v^2 / Q(z)Q(z^{-1}) \\ g_{sx}(z) = g_{ss}(z) = \sigma_w^2 / S(z)S(z^{-1}) \\ g_{ux}(z) = g_{uu}(z) = \sigma_u^2 . \end{cases}$$

Since $\left[ \frac{\sigma_u^2}{B(z^{-1})} \right]_+ = \sigma_u^2$ on account of the normalization $b_o = 1$,

$$(3.54) \qquad \psi(z) = \frac{\sigma_u^2}{\sigma^2 B(z)} ,$$

so that

$$(3.55) \qquad \gamma(z) + \varphi(z) = 1 - \frac{\sigma_u^2}{\sigma^2 B(z)} .$$

We need only determine one of the two generating transforms and the expression $\sigma_u^2/\sigma^2 B(z)$ ; the remaining generating transform is then deduced from (3.55). This may be of considerable practical importance since one of the transforms, usually $\varphi(z)$ , may be quite difficult to obtain directly. For the general model (3.45) we readily obtain

$$(3.56)\quad \begin{cases} \gamma(z) = \dfrac{\sigma_u^2 \lambda}{\sigma^2 B(z)} \left[ \dfrac{1}{Q(z)Q(z^{-1})B(z^{-1})} \right]_+ \\ \varphi(z) = \dfrac{\sigma_u^2 \mu}{\sigma^2 B(z)} \left[ \dfrac{1}{S(z)S(z^{-1})B(z^{-1})} \right]_+ \end{cases}$$

$$= 1 - \frac{\sigma_u^2}{\sigma^2 B(z)} \left[ 1 - \frac{\lambda}{Q(z)Q(z^{-1})B(z^{-1})} \right]_+$$

The function $B(z)$ is obtained as

$$(3.57)\qquad B(z) = \frac{\prod_{j=1}^{N} (1 - \beta_j z)}{Q(z)S(z)} ,$$

where $\beta_j$ , $j = 1, \ldots, N$ are the $N$ roots of the polynomial

$$(3.58)\quad P(z) = \lambda S(z)S(z^{-1}) + \mu Q(z)Q(z^{-1}) + S(z)S(z^{-1})Q(z)Q(z^{-1}) ,$$

which lie <u>outside</u> the unit circle. If we assume $Q(z)$ is of degree $n$ and $S(z)$ is of degree $mL$ , then $P(z)$ will be of degree $2N$ where $N = n + mL$ . Since $P(z)$ is symmetric in $z$ and $z^{-1}$ , the $2N$ roots will come in pairs; one member of each pair lies outside the unit circle, the other, inside. The variance $\sigma^2$ is determined as $c\sigma_u^2$ where $c$ is

chosen so as to make

$$(3.59) \qquad P(z) = c \prod_{j=1}^{N} (1 - \beta_j z)(1 - \beta_j z^{-1})$$

equal to the right-hand side of (3.58).[16]

Having obtained $\gamma(z)$ and $\varphi(z)$, we readily obtain the second equation of (3.44) in distributed lag form

$$(3.60) \qquad h_t = \sum_{j=0}^{\infty} (a\gamma_j + b\varphi_j)x_{t-j} + e_t .$$

Note that the behavioral parameters a and b now enter into the lag structure in a more intimate way than in the distributed lag relation resulting from a two-component model; in the latter case the behavioral parameter may be factored out of the lag distribution.

As a specific illustration of the type of model implied by (3.44) and the stochastic specification (3.45), we refer the reader to the model analyzed by Grether [13, pp. 16-19], namely (3.46) with $n = m = 2$ and $L$ unspecified. The distributed lag relation implied by Grether's results is extremely complicated and does not resemble any lag distribution so far proposed, although, like nearly everything, it may be well approximated by one of Jorgenson's rational distributed lag functions, [17].[17]

## IV. Some Implications for Estimation

The general problems of estimating relationships involving even simple distributed lags are well-known. Put simply they are, first, the strongly nonlinear way in which the parameters of the lag distribution enter the relationship, and, second, the difficult problems raised by the possibility of serial correlation in the disturbances of the relation to be estimated. This last problem is especially severe when the relationship to be estimated is transformed into an autoregressive one in order to circumvent certain of the difficulties caused by the nonlinear way in which the parameters of the lag distribution enter.

To illustrate these problems consider the lag distribution given by (3.13) and the relationship

$$h_t = a\hat{y}_t + e_t \tag{4.1}$$

where the constant term has been dropped for convenience. This simple case contains all the elements of the more complicated models discussed above. By virtue of (3.12), we may write

$$(1 - \beta U)h_t = a\left(\frac{\alpha-\beta}{\alpha}\right)x_t + (1 - \beta U)e_t \ , \tag{4.2}$$

or

$$h_t = a\left(\frac{\alpha-\beta}{\alpha}\right)x_t + \beta h_{t-1} + e_t - \beta e_{t-1} \ . \tag{4.2'}$$

Clearly, if $\{e_t\}$ was white noise to start the disturbance in (4.2') will not be white. In ordinary regression, the effect of such serial correlation would be loss of efficiency in the estimation of the

regression slopes and bias in the estimation of the residual variance and standard errors. In the estimation of autoregressive relationships the problem, however, is more serious. The estimates of $a(\frac{\alpha-\beta}{\alpha})$ and $\beta$ will typically be inconsistent as well as inefficient.

Except for the parameter $\alpha$ (which makes (4.2) unidentified without further assumptions), the relationship poses no really serious estimation problems other than the nonlinear way in which the parameters enter provided we assume $\{e_t\}$ is white noise. Let $\alpha = 1$, therefore, and

$$\epsilon_t = e_t - \beta e_{t-1} . \tag{4.3}$$

The generating transform of the autocovariances of the sequence $\{\epsilon_t\}$ is $\sigma^2(1 - \beta z)(1 - \beta z^{-1})$ where $\sigma^2$ is the variance of $e_t$. Thus, the variance-covariance matrix of $\epsilon = (\epsilon_1, \ldots, \epsilon_T)'$ is

$$\Omega = E\epsilon\epsilon' = \sigma^2 \begin{bmatrix} 1+\beta^2 & -\beta & 0 & \cdots & 0 \\ -\beta & 1+\beta^2 & -\beta & \cdots & 0 \\ 0 & -\beta & 1+\beta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1+\beta^2 \end{bmatrix} . \tag{4.4}$$

Thus, if $\epsilon_t$ were assumed to be distributed normally with zero mean and variance $\sigma^2$, we could set up the logarithmic likelihood function in a straightforward fashion. As Klein [18] shows, maximization in the usual way leads to highly nonlinear equations. Various numerical methods are available for the solution of this type of problem; see Traub [35]. Alternatively one may consider this a problem in the

minimization of a weighted sum of squared residuals, i.e.,

$$(4.5) \qquad \min_{a,\beta} \left\{ (h - ax - \beta h_{-1})' \, \Omega^{-1} (h - ax - \beta h_{-1}) \right\}$$

where $h = (h_1, \ldots, h_T)'$ , $h_{-1} = (h_0, \ldots, h_{T-1})'$, and $x = (x_1, \ldots, x_T)'$ . This is equivalent to maximizing the logarithmic likelihood function if the term involving $\det \Omega$ is neglected. Asymptotically, the procedure leads to the same estimates.[18]

When $\alpha$ is not assumed known as above, serious difficulties arise. Let $\hat{a}_o$ be an estimate of the coefficient of $x_t$ in (4.2') and let $\hat{\beta}$ be an estimate of the coefficient of $h_{t-1}$ ; then any pair $\hat{\alpha}$ and $\hat{a}$ satisfying the equation

$$(4.6) \qquad \hat{\alpha} = \frac{\hat{\beta}}{\hat{a}_o - \hat{a}}$$

will maximize the likelihood function if $\hat{a}_o$ and $\hat{\beta}$ do. This identification problem clearly arises in all of the models discussed above in which the distributed lags are assumed to be a result of optimal extractions of components of stationary time series. However, as we shall see shortly, the very assumptions giving rise to the difficulty provide a natural solution.

If nothing is assumed about the stochastic structure of $\{e_t\}$ , except perhaps of the most general sort, the problem of estimating (4.2) becomes more than one of merely great computational difficulty. This is not, it should be emphasized, intended to minimize severity of the computational problems arising in even very simple distributed lag

models, but to stress the entirely new dimension given to the problem when the covariance structure of $\{e_t\}$ is known only in a very general way. Provided one were willing to make sufficiently strong parametric assumptions about the covariance structure of $\{e_t\}$ it is again true that maximum-likelihood methods may be employed. Let $\{e_t\}$, for example, be generated by a first-order autoregressive scheme:

$$e_t - \rho e_{t-1} = \epsilon_t \tag{4.7}$$

where $\{\epsilon_t\}$ is white noise. Then, assuming $\alpha = 1$, (4.2) becomes

$$(1 - \beta U)h_t = a(1 - \beta)x_t + (1 - \beta U)(1 - \rho U)^{-1}\epsilon_t \tag{4.8}$$

or

$$(1 - \rho U)(1 - \beta U)h_t = a(1 - \beta)(1 - \rho U)x_t + (1 - \beta U)\epsilon_t \,. \tag{4.8'}$$

Equation (4.8') shows that maximum-likelihood estimates are possible but computationally complicated.[19/]

When dealing with monthly or quarterly data, it will seldom be reasonable to assume anything so simple as first-order serial correlation of the disturbances. In general, the presence of seasonality induces much higher order serial correlation. If we are willing to provide a parametric stochastic structure, we can again achieve maximum likelihood estimates, but our ability to specify the nature of the stochastic structure in such fine detail might well be limited. Under these circumstances, methods of the sort suggested by Hannan [16a] and further elaborated by Dhrymes [7] should prove useful. If the spectrum of the disturbances were known, as of course it would be if $\Omega$ were fully specified, a

form of generalized least squares would be appropriate for estimating distributed lag relationships in the form (4.2). The trouble, if $\{e_t\}$ is assumed to be white noise, is that $\Omega$ depends on one of the parameters we are trying to estimate. When $\Omega$, however, is more complicated, it may well pay to ignore this information. Hannan's procedure, then amounts to replacing $\Omega$ by a consistent estimate and proceeding, in the frequency domain, to obtain generalized least-squares estimates of the regression parameters. The advantage of working in the frequency domain is that relatively complicated stochastic structures of the disturbances in (4.2) may be consistently estimated without tight parametric specifications. The price one pays is that a great deal of data is required before one can do this in a reasonably refined fashion. The consistent estimate of $\Omega$, or its equivalent in the frequency domain, must be obtained in a "first round" in which consistent estimates of $a(1 - \beta)$ and $\beta$ are obtained. The natural estimates in this case would be found by using $x_t$ and $x_{t-1}$ as instruments. The estimated residuals are then used to compute estimates of the spectral densities function of $\{e_t - \beta e_{t-1}\}$. Even apart from the dependence of this spectrum on one of the parameters to be estimated, the "second-round" estimates are known not to be fully efficient when the relationship to be estimated contains an autoregressive component.[20/]

The general conclusion one may draw from consideration of even this simplest of cases, is that the problem of inference about distributed lag relationships is exceptionally complex. Nonetheless, the interpretation of certain distributions of lag as optimal extractions may make inference

about at least these lags easier. The rationale is a simple one: If we believe that the lags are due to the fact that the economic agents under consideration react not to the observed values of certain variables but to unobserved components thereof, and if, further, we are willing to specify the stochastic structure of the time series involved in order to obtain parametrically specified distributions of lag, it is not implausible to assume that the perception of this structure by the economic agents in question is reasonably accurate. The significance of such an additional assumption is that we can now introduce a prior stage in our analysis of behavior: We can try to obtain optimal extractions of those components which we believe to enter the behavioral relations directly from the observed time series. Such estimates may then be used in place of the complicated distributed lag structures deduced above. Such an approach has the further advantage that the identification problems noted above do not arise in the second stage of the analysis. If, for example, prior estimates of $\alpha$ and $\beta$ are used to obtain $\hat{y}_t$ in (4.1), and then this series is used to estimate $\underline{a}$, the specific identification problem occurring in connection with (4.2) does not arise.

The approach suggested does not, of course, solve all difficulties; indeed, it creates several new ones.

First, the estimation of the optimal extractors used in estimating the behavioral relations requires estimates of parameters such as $\alpha$ and $\beta$ appearing in the lag distributions derived earlier in this paper. Estimates of such parameters with desirable properties are not trivial to obtain. Crude estimates can be found easily, as, for example, in Couts, Grether, and Nerlove [6, pp. 18-19], by matching theoretical spectra to the observed spectrum of the time series to be decomposed. But the properties of such estimates are not well-known, although it is

fairly obvious that they are not likely to be very good. Both for this reason, and because we do not have infinite past series of observations, the components entering our behavioral relations must be considered to be measured with error. In a multi-component model, the errors attaching to different component estimates cannot assuredly be assumed independent. It may, however, be reasonable to assume the errors are independent of the disturbance in the behavioral relation to be estimated. As a first approximation, at least, we may be prepared to overlook such difficulties.

Second, it is apparent that the behavior of the series $\{h_t\}$, which we take as dependent in the behavioral relation to be estimated, must cast some light on the stochastic structure of the series $\{x_t\}$ under the assumptions made. It is important in view of the difficulties of estimating this structure to try to take all information into account. By breaking the problem down into two stages, we have, in effect, separated ourselves from this additional information. Such a view, however, is perhaps too pessimistic, for knowledge of the behavioral parameters enables us to use information on the dependent variable to make further inferences on the stochastic structure of the series $\{x_t\}$. Furthermore, if perceptions of this structure on the part of the economic agents are assumed to be correct, the same optimal extraction will appear in several relationships (e.g., the same "permanent income" appears in consumption functions for categories of total consumption). This means that more than one dependent variable bears on the nature of the series $\{x_t\}$.

Finally, the problem of serial dependence of the disturbances in the behavioral relation remains. However, neglecting the difficult matter of errors of measurement in the extracted components, it is apparent that the approach suggested reduces the problem to a different order of complexity, for now we do not have to take into account the dependence of the stochastic structure of the disturbances upon some of the very same parameters we wish to estimate. Furthermore, we are not now trying to estimate an autoregressive relationship. This is true whether we use estimates of $\alpha$ and $\beta$ to obtain direct estimates of $\hat{y}_t$ above and estimate $\underline{a}$ in (4.1) or whether we use these estimates to define new variables

$$(4.9) \qquad \begin{cases} h_t^* = h_t - \beta h_{t-1} \\ x_t^* = \left(\frac{\alpha-\beta}{\alpha}\right) x_t \end{cases}$$

and estimate

$$(4.10) \qquad h_t^* = a x_t^* + \epsilon_t$$

where $\epsilon_t$ is given by (4.3) and has variance-covariance matrix depending on $\beta$ as well as the stochastic properties of $\{e_t\}$.

The importance of deducing the explicit lag structure from explicit assumptions about the structure of the observed time series is two-fold. First, we obviously need such a lag structure to estimate the components entering the behavioral relation. More important, however, is that knowledge of the explicit form of the lag structure will often

permit a greatly simplifying transformation to be carried out. One such is given in (4.9) but the power of the method may be better appreciated by examination of the case in which $\hat{y}_t$ is determined by past values of $x_t$ using the weights generated by (3.43). In this case, prior estimates of $\alpha$ and $\beta_1, \ldots, \beta_n$ permit us to make the transformation

$$(4.11) \qquad \begin{cases} h_t^* = \prod_{j=1}^{n} (1 - \beta_j U) h_t \\ x_t^* = \left(\frac{\lambda}{c}\right)\left\{1 + \sum_{j=0}^{n-1} g_j (1 - \alpha U)^j\right\} x_t \ , \end{cases}$$

which are both finite moving averages. A relationship such as (4.10) may then be estimated. Estimates of $\underline{a}$ and of the spectrum of $\{\epsilon_t\}$ then permit us to draw inferences from a cross-spectral analysis of $h_t$ and $x_t$ about the parameters $\alpha$ and $\beta_1, \ldots, \beta_n$ .

## FOOTNOTES

1/ Friedman [12, pp. 142-152]. Friedman also allows for trend in his estimates, but this need not concern us here.

2/ This section is based on Whittle [37], Chapters 1-3, 6, and 8, and on Grether [13].

3/ I.e., have means which do not depend on the time index $t$ and autocovariances which depend only on the lag involved and not on the absolute value of the time index.

4/ Note the normalization $b_0 = 1$ has been imposed. We could equally well have imposed the alternative normalization $\sigma^2 = 1$.

5/ Because of the nature of $g_{yy}(z)$ these coefficients cannot be zero.

6/ It follows from the fact that $\beta$ is the root of $\lambda + (1 - \alpha z)(1 - \alpha z^{-1})$ which is less than one that $\beta < \alpha$ if $\alpha > 0$. Thus, the expression $\frac{\alpha - \beta}{\alpha}$ is positive. If $\alpha < 0$, we choose the other root (now less than one), and then this value can be shown to be greater than $\alpha$; in this case, therefore, $\frac{\alpha - \beta}{\alpha} > 0$ too.

7/ We cannot from above, since $\alpha$ must be less than one for convergence of most of our expressions in the preceding pages.

8/ $\beta$ now depends only on the ratio of variance $\lambda$, but it is still less than 1.

9/ It is easy to show as well that the same formula suffices for <u>predictions</u> of $y_{t+\nu}$. To do so, observe that when we replace $y_t$ in the calculation of $g_{yx}(z)$ by $y_{t+\nu}$, we obtain not $g_{yy}(z)$ but $g_{yy}(z)/z^{\nu}$.

10/ All values so obtained must obviously be equal. This result is obtained most simply by equating the coefficients of $\zeta = z + z^{-1}$ in the two representations of $P(z)$.

11/ The value 0.75, curiously enough, although an arbitrary choice, seems to have been almost inspired. Even the tides in Sidney harbor seem to follow a quasi-difference scheme with parameter 0.75! See Hamon and Hannan [15, p. 6034].

12/ Thus $r_o = 1$

$$r_1 = \sum_{j \neq 1} \beta_j + \sum_{j \neq 2} \beta_j + \ldots + \sum_{j \neq n} \beta_j$$

and so on.

13/ It would be a useful exercise for the reader to show that (3.43) reduces to (3.12) when $n = 1$. This is a great deal simpler than deducing the limit as $\alpha \to 1$.

14/ When $\theta = 1$, this is the nonstationary model analyzed in Couts, Grether, and Nerlove [6]. $L$, of course, is typically 12 or 4, since monthly or quarterly data are usually the subject of analysis.

15/ Of course, we will certainly wish to assume $Q(z)$ and $S(z)$ only have roots outside the unit circle. Again the nonstationary cases arising when they have roots on the unit circle may be handled easily in the analysis which follows by our earlier device in which we let certain roots outside the circle tend to limits on the circle.

16/ The algebra necessary to determine $c$ explicitly in terms of the roots of $Q(z)$ and $S(z)$ is so similar to that carried out above that we shall not present it here.

17/ For it to be an exact special case, both $\gamma(z)$ and $\varphi(z)$ must be expressible as polynomials of finite degree divided by similar polynomials. This is clearly the case for the lag distribution associated with the generating function in (3.43), but it is not obvious that Grether's results permit of a similar characterization. Jorgenson [17, pp. 139-142] shows, however, that an arbitrary distributed lag function may be approximated to any desired degree of accuracy by a rational distributed lag function.

18/ Solow [33] suggests this procedure for the even more general Pascal lag distribution and shows (pp. 400-401), that one can express the resulting problem as one in concave programming. See also Jorgenson [17, pp. 145-148]. Use of "weighted regression" amounts to treating (4.2') as an errors-in-variables model with one variable measured without error; viz.,

$$(h_t - e_t) = \beta(h_{t-1} - e_{t-1}) + a\left(\frac{\alpha-\beta}{\alpha}\right)x_t .$$

19/ One can see that $\hat{\rho}$ and $\hat{\beta}$ are separately identifiable as follows: Let $a_1$ be the coefficient of $x_t$ in (4.8'), let $a_2$ be the coefficient of $x_{t-1}$, $a_3$ of $h_{t-1}$ and $a_4$ of $h_{t-2}$. Then either of the following relations determine $\rho$

(Footnote 19 continued on next page)

19/ Footnote 19 continued from Page 55.

$$\text{(i)} \qquad -a_3 = \hat{\rho}\left(1 + \frac{a_1}{a_2}\right) + 1$$

or

$$\text{(ii)} \qquad a_4 = \hat{\rho}\left\{\hat{\rho}\left(\frac{a_1}{a_2}\right) + 1\right\} .$$

Because two relations may be used, $\hat{\beta}$ and $\hat{\rho}$ are in fact over identified. One might also consider estimating $\hat{\beta}$ and $\hat{\rho}$ from

$$\text{(iii)} \qquad \begin{cases} \hat{\beta} + \hat{\rho} = -a_3 \\ \hat{\beta}\hat{\rho} = a_4 , \end{cases}$$

see Malinvaud [21, p. 469]. However, a little reflection shows that the two quadratics obtained from (iii), one in $\hat{\beta}$, the other in $\hat{\rho}$, have identical roots. Although there are two, there is no way to identify one as $\hat{\beta}$ and the other as $\hat{\rho}$ except by making use of the other coefficients, $a_1$ and $a_2$.

20/ See Amemiya and Fuller [2].

REFERENCES

[1] Alt, F., "Distributed Lags," Econometrica, 10: 113-28 (1942).

[2] Amemiya, T., and W. Fuller, "A Comparative Study of Alternative Estimators in a Distributed-Lag Model," Tech. Rept. 12 (NSF-GS-142), Stanford University (July 23, 1965).

[3] Ames, E., and S. Reiter, "Distribution of Correlation Coefficients in Economic Time Series," Jour. Amer. Stat. Assoc., 56: 637-56 (1961).

[4] Buys-Ballot, C. H. D., Les Changements Périodiques de Température (Utrecht: 1847).

[5] Cagan, P., "The Monetary Dynamics of Hyper-Inflation," pp. 25-117 in M. Friedman, ed., Studies in the Quantity Theory of Money (Chicago: University of Chicago Press, 1956).

[6] Couts, D., D. Grether, and M. Nerlove, "Forecasting Non-Stationary Economic Time Series," Management Science, 13: 1-21 (1966).

[7] Dhrymes, P. J., "Estimation of the General Rational Lag Structure by Spectral Techniques," Discussion Paper No. 35, Department of Economics, University of Pennsylvania (1967).

[8] Eisner, R., and R. Strotz, "Determinants of Business Investment," pp. 59-337 in Impacts of Monetary Policy, Studies Prepared for the Commission on Money and Credit (Englewood Cliffs: Prentice-Hall, 1963).

[9] Fisher, I., "Our Unstable Dollar and the So-Called Business Cycle," Jour. Amer. Stat. Assoc., 20: 179-202 (1925).

[10] _____, "Note on a Short Cut Method for Calculating Distributed Lags," Bul. Inst. Inter. Stat., 29: 323-27 (1937).

[11] Frickey, E., Economic Fluctuations in the United States (Cambridge: Harvard University Press, 1942).

[12] Friedman, M., A Theory of the Consumption Function (Princeton: Princeton University Press, 1957).

[13] Grether, D., "Application of Signal Extraction Techniques in the Study of Economic Time Series," Cowles Foundation Discussion Paper No. 202 (February 25, 1966).

[14] Griliches, Z., "Distributed Lags: A Survey," Econometrica, forthcoming.

[15] Hamon, B. V., and E. J. Hannan, "Estimating Relations Between Time Series," Jour. Geophys. Res., 68: 6033-6041 (1963).

[16] Hannan, E. J., Time Series Analysis (London: Methuen, 1960).

[16a] _____, "The Estimation of Relationships Involving Distributed Lags," Econometrica, 33: 206-24 (1965).

[17] Jorgenson, D., "Rational Distributed Lag Functions," Econometrica, 32: 135-49 (1966).

[18] Klein, L. R., "The Estimation of Distributed Lags," Econometrica, 26: 553-65 (1958).

[19] Koyck, L. M., Distributed Lags and Investment Analysis (Amsterdam: North-Holland Publishing Co., 1954).

[20] Kuznets, S., Seasonal Variations in Industry and Trade (New York: National Bureau of Economic Research, 1933).

[21] Malinvaud, E., Statistical Methods of Econometrics (Chicago: Rand-McNally and Co., 1966).

[22] Muth, J. F., "Optimal Properties of Exponentially Weighted Forecasts," Jour. Amer. Stat. Assoc., 55: 299-306 (1960).

[23] Nerlove, M., "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," Jour. Farm Econ., 38: 496-509 (1956).

[24] _____, "The Implications of Friedman's Permanent Income Hypothesis for Demand Analysis," Agricultural Economics Research, 10: 1-14 (January 1958).

[25] _____, The Dynamics of Supply: Estimation of Farmers' Response to Price (Baltimore: The Johns Hopkins Press, 1958).

[26] _____, Distributed Lags and Demand Analysis, Agricultural Handbook No. 141 (Washington: U.S. Government Printing Office, 1958).

[27] _____, "Notes on the Estimation and Identification of Cobb-Douglas Production Functions," hectographed (1959), section 6. Later published as Estimation and Identification of Cobb-Douglas Production Functions (Chicago: Rand-McNally, 1965), Chapter 7, pp. 131-56.