DISTRIBUTED LAGS: A SURVEY

by

Phoebus J. Dhrymes

Discussion Paper Number 24
October 1972

Distributed Lags: A Survey

by

Phoebus J. Dhrymes[†]

1. Introduction and Preliminary Concepts

The problem of distributed lags in economics has its origin in the following considerations. Suppose we are interested in an economic phenomenon, y, which has as its (systematic or non-stochastic) determinant some economic variable, x. The influence of the latter on the former, however, is not exerted instantaneously but is spread or distributed over many (perhaps infinitely many) time periods. This leads to the formal model

$$y_t = \sum_{i=0}^{\infty} w_i x_{t-i} + u_t \tag{1}$$

where y_t represents the economic phenomenon of interest (at time t), x_t is the determining variable, w_i, $i = 0,1,2,\dots$, are fixed, but unknown, constants and u_t is a suitably specified random variable. It is then readily seen that a change in x does not translate instantaneously and fully into a change in the conditional mean of y. For example, consider the sequence

$$\begin{aligned} x_t &= 0, \qquad t = 0, -1, -2, \dots \\ &= 1, \qquad t = 1,2,3,\dots \end{aligned}$$

[†] This is an invited survey of certain aspects of the literature on distributed lags and is to appear in the Soviet journal <u>Economics and Mathematical Methods</u>.

The conditional mean of y is given by

$$(2) \qquad E(y_t \mid x_{t-i}, i = 0,1,2,\ldots) = 0 \qquad t = 0, -1, -2, \ldots$$
$$= \sum_{i=0}^{t-1} w_i \qquad t = 1,2,\ldots$$

Thus, even though a permanent change has occured in the determining variable the impact on the dependent variable is distributed over an infinite period, the instantaneous impact being w_0, the cumulative impact after two periods being $w_0 + w_1$, after three $w_0 + w_1 + w_2$ and so on. It is useful at this stage to introduce

Definition 1. The model in (1) is termed the general infinite distributed lag model. The sequence $\{w_i : i = 0,1,2,\ldots\}$ is termed the general infinite lag structure. The elements, w_i of the sequence above are said to be the lag coefficients.

Various special cases of the model in (1) have been employed in empirical work. Perhaps the earliest such application is to be formed in Irving Fisher [10]. Fisher's application is a rather crude one in that he proposed the scheme

$$w_i = 0 \qquad i > n$$
$$w_i = \alpha(n + 1 - i) \qquad i = 0,1,2,\ldots,n \; .$$

An independent development along similar lines is that of Almon [1] who proposed the scheme

$$w_\tau = 0 \qquad \tau > n$$

$$w_\tau = P^*(t) \qquad \tau = 0,1,2,\ldots,n$$

where

(3) $$P^*(t) = \sum_{j=0}^{k} b_j s_j(t)$$

the $s_j(\cdot)$ being suitably defined Lagrange interpolation polynomials. Almon's application was to the problem of determining the pattern by which appropriations are translated into expenditures. Thus, firms or governments may decide on a certain level of budgetary appropriations for a certain type of expenditure. Typically, the authority embodied in the budgetary appropriation lapses after a certain number of time periods (years). In this case there is a good institutional reason why one may take the lag coefficients to be zero for $\tau > n$. It may be shown that the introduction of Lagrange polynomials above is quite superfluous and we may best define this class of distributed lag models by the model in (1) subject to the conditions

(4) $$w_\tau = 0,\ \tau > n,\ w_\tau = P(\tau) \qquad P(t) = \sum_{i=0}^{k} \beta_i t^i$$

We have, therefore,

Definition 2. The model in (1) together with the restrictions in (4) is said to be the polynomial lag model of order n and degree k.

This model has been employed quite extensively in recent empirical work; particularly so in the context of economy wide economic models such as the Wharton Model [26] and the MPS Model[1] [29] - both of which are models of the U. S. economy. Similar applications are contemplated with respect to project LINK which is essentially a global model of trade flows.

By far the most extensive empirical application of distributed lag models, however, originated with the formulation first suggested in the work of the Dutch econometician L. M. Koyck [22]. His problem was that of "explaining" fixed investment and his suggestion was that after some period, say, τ_0 one should specify the following behavior for lag coefficients

$$w_\tau = \alpha \lambda^\tau \qquad \tau = \tau_0, \tau_0 + 1, \dots \quad (5)$$

If for simplicity of exposition we take $\tau_0 = 0$ then the model in (1) may be rewritten as

$$y_t = \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t \quad (6)$$

This leads to

Definition 3. The model in (6) is said to be the geometric distributed lag model.

---

[1] Formerly known as the MIT-FRB Econometric Model

Since Koyck's first formulation a number of very ingenious rationalizations of the geometric lag have been proposed, notably those due to Cagan [5] and Nerlove [28].

The discussion of the relationships among these diverse developments and the general exposition of such models will be much facilitated if we introduce the notion of the <u>lag operator</u>. The discussion of these aspects here will be somewhat elliptical. Greater detail may be found in Dhrymes [8, Ch. 2] or Griliches [13].

<u>Definition 4</u>. Let X be the set of all functions $\{x : N \rightarrow R\}$ where N is the set $\{0, \pm 1, \pm 2, \ldots\}$ and R is the one dimensional Euclidean space. The <u>lag operator</u>, L, is defined by

$$Lx(t) = x(t-1) \qquad \forall x \in X .$$

Powers of the lag operator are defined by iteration, i.e.,

$$L^2 x(t) = L[Lx(t)] = Lx(t-1) = x(t-2) . \tag{7}$$

In general

$$L^k x(t) = x(t-k) \qquad k = 1,2,\ldots . \tag{8}$$

and by convention

$$L^0 \equiv I, \qquad Ix(t) = x(t) \ \forall x \in X . \tag{9}$$

The operator I is said to be the <u>identity operator</u>. It is easily established, in view of (7) and (8), that

$$L^{k+s} = L^{s+k}, \quad (L^k)^s = L^{ks} \tag{10}$$

Moreover, it may be shown, Dhrymes [8 , Ch. 2] that the set of all finite linear combinations of elements of the set $\{I, L, L^2 \ldots\}$ over the field of complex numbers constitutes an <u>algebra</u> which is <u>isomorphic to the algebra of polynomials</u> in the complex ideterminate.

The import of the preceding is that in dealing with polynomials (or power series) in L we can treat the latter in the same way as we do the complex indeterminate. The convenience of this framework will be evident in the discussion to follow.

In the notation just introduced, we may rewrite the general infinite distributed lag model in (1) as

$$y_t = W(L)x_t + u_t, \quad W(L) = \sum_{i=0}^{\infty} w_i L^i \tag{11}$$

<u>Definition 5</u>. The quantity $W(L)$ of (11) is said to be the <u>general infinite lag operator</u>.

The specification introduced by Koyck and given in (6) may be written as

$$y_t = \frac{\alpha I}{I - \lambda L} x_t + u_t \tag{12}$$

provided $|\lambda| < 1$. The latter is a condition implicit in the model given in (6).

Cagan's contribution, alluded to above, relates to the adaptive expectations rationale for the model in (12). The, suppose that economic agents act according to

$$y_t = \alpha p_t^* + u_t \tag{13}$$

Above, y is some observable economic variable and p^* is an expectational quantity. The adaptive expectations model hypothesizes that economic agents revise their expectations linearly according to the most recently observed deviation of expectations from realizations. In particular, if p is the observed value of the variable the hypothesis states

$$p_t^* - p_{t-1}^* = \beta(p_{t-1} - p_{t-1}^*) \qquad \beta \in [0,1) . \tag{14}$$

Solving (14) by the use of lag operator methods we have

$$p_t^* = \frac{(1-\gamma)I}{I-\gamma L} \, p_{t-1} \qquad \gamma = 1-\beta \tag{15}$$

Inserting in (13) we find

$$y_t = \frac{\alpha(1-\gamma)I}{I-\gamma L} \, p_{t-1} + u_t \tag{16}$$

which is the geometric distributed lag model.

Nerlove's formulation is equally ingenious. In his scheme one supposes that economic agents, given the information conveyed by their environment, determine the optimal quantity of a certain economic variable, say y_t^*. At the same time they hold a certain magnitude of this variable say y_{t-1}. The adjustment they make to their holdings is then asserted to be proportional to the gap between actual and desired holdings, i.e., one hypthesizes that

$$(17) \qquad y_t - y_{t-1} = \alpha(y_t^* - y_{t-1}) + u_t \qquad \alpha \in [0,1)$$

it being understood that the relationship only holds when $y_t^* - y_{t-1} \geq 0$. For example if y_t^* is desired capital stock and if $y_t^* - y_{t-1} < 0$ then decumulation can occur only at the rate of depreciation, it being assumed that second hand markets for fixed capital do not exist. Actually, this assumption may not be very realistic for individual firms although it is clearly a very reasonable one for the economy as a whole. At any rate solving (17) we find

$$(18) \qquad y_t = \frac{\delta I}{I - \delta L} y_t^* + \frac{I}{I - \delta L} u_t \qquad \delta = 1 - \alpha .$$

But y_t^* is not an observable quantity; we may, however, postulate that optimization occurs according to

$$(19) \qquad y_t^* = ax_t$$

where x_t is an observable economic quantity. Consequently,

$$(20) \qquad y_t = \frac{\alpha a I}{I - \delta L} x_t + \frac{I}{I - \delta L} u_t .$$

Except for the appearence of the random component this is the geometric lag model exhibited in (12). The ingenuity of the Nerlove formulation will become evident when we consider the estimation problem posed by such models.

A somewhat unsatisfactory feature of geometric lag structures is that the sequence of lag coefficients declines monotonically, so that the largest impact of a change in the determining variable is registered instantaneously; thereafter the lag coefficients decline in magnitude. In many economic problems, however, we have reason to believe that lag coefficients first increase, reach a peak and then decline. Such a behavior cannot be described by the geometric lag structure. This prompted Solow [32] to suggest the Pascal lag structure which postulates that

$$w_i = (1-\lambda)^r \binom{r+i-1}{i} \lambda^i, \qquad i = 0,1,2,\ldots \tag{21}$$

$$\lambda \in [0,1) \qquad r \in (0,\infty) .$$

The operator suggested by this specification is

$$W(L) = \frac{(1-\lambda)^r I}{(I-\lambda L)^r} . \tag{22}$$

In empirical applications - of which there have been only a few - one usually takes r to be an integer. It may be easily demonstrated that the operator above results from the successive application of r geometric lag operators, all having the same parameter λ. More appropriately the lag structure exhibited in (21) is termed the negative binomial lag structure.

A final development in the evolution of distributed lag specification in the literature of econometrics is the work of Jorgenson [19]. Jorgenson suggested the rational lag operator

$$W(L) = \frac{A(L)}{B(L)} \tag{23}$$

where

$$A(L) = \sum_{i=0}^{n} a_i L^i , \qquad B(L) = \sum_{j=0}^{n} b_j L^j , \qquad m \leq n , \tag{24}$$

it being understood that the polynomials $A(\cdot)$, $B(\cdot)$ contain no common factors. Although Jorgenson sought to justify the rational lag specification as an approximation to the general infinite lag structure exhibited in (1) by appealing to the Weierstrass approximation theorem, this justification is rather unsatisfactory. For a discussion of this aspect see Dhrymes [8, Ch. 3] and Sims [31]. The specification in (23) and (24) is best looked upon as a hypothesis about the structure of the economic phenomenon under study rather than as an approximation to a perfectly general underlying lag structure.

In closing this section it ought to be noted that an uncanny parallel development has occurred in the engineering literature; in particular, the rational lag distribution has received extensive attention. See, for example, Steiglitz and McBride [33], and the recent survey paper by Astrom and Eykhoff [3].

2. Characterization of Lag Structures

Often, it may be useful to give a summary characterization of the lag distribution. In particular, we may be interested in whether the impact of the determining on the dependent variable is "concentrated" or "diffuse", or in the average lag by which such impact is registered. For certain type of lag structures such summary characterizations are readily available. To this effect note that from (21) we obtain

$$(25) \qquad \lim_{t \to \infty} E(y_t | x_{t-i}, \ i = 0,1,2,\dots) = \sum_{i=0}^{\infty} w_i = \omega$$

It is clear that the economics of the problem could typically require the condition $|\omega| < \infty$; frequently, it would also be required that the lag coefficients be of the same sign. When these two conditions hold we may define

$$(26) \qquad w_i^* = \frac{w_i}{\omega} \qquad i = 0,1,2,\dots .$$

We then observe that

$$(27) \qquad w_i^* \in [0,1], \quad \sum_{i=0}^{\infty} w_i^* = 1 .$$

We have

<u>Definition 6</u>. If the lag structure $\{w_i : i = 0,1,2,\dots\}$ has the properties

$$w_i w_j \geq 0 \quad \text{for all} \quad i,\ j$$

$$\sum_{i=0}^{\infty} w_i = \omega < \infty$$

then it is said to be a <u>normalizable lag structure</u>.

<u>Remark 1</u>. Normalizable lag structures are isomorphic to probability mass functions. Indeed, some of the early work on distributed lags relied heavily on this isomorphism. Thus, recall that if y is a geometric (discrete) random variable, then the mass function of $x = y - 1$ is given by

$$\Pr\{x = i\} = (1 - \lambda)\lambda^i \qquad i = 0,1,2,\ldots$$

whose ordinates give the geometric lag structures, introduced by Koyck [22].

Let $\{x_i : i = 1,2,\ldots,r\}$ be mutually independent identically distributed (i.i.d.) random variables having the mass function above; then

$$x = \sum_{i=1}^{r} x_i$$

has the mass function

$$\Pr\{x = i\} = (1 - \lambda)^r \binom{r + i - 1}{i} \lambda^i \qquad i = 0,1,2,\ldots\ .$$

But the ordinates of this mass function give the Pascal lag structure introduced by Solow [32].

Making use of the isomorphism noted above we may define various neasures of central tendency or dispension in exactly the same manner as we do with random variables.

Definition 7. Let $\{w_i : i = 0,1,2,\ldots\}$ be a lag structure and t be a real indeterminate. Then

$$W(t) = \sum_{i=0}^{\infty} w_i t^i$$

is said to be the lag generating function (of the lag structure).

If the lag is normalizable, then by suitably redefining the lag coefficients, if necessary, we observe that

$$W'(1) = \sum_{i=0}^{\infty} i\, w_i = m(w) \tag{28}$$

Definition 8. If $\{w_i : i = 0,1,2,\ldots\}$ is a normalized lag structure, i.e., if $w_i \in [0,1]$, $\Sigma_{i=0}^{\infty} w_i = 1$, then the quantity in (28) is said to be the mean lag; the quantity

$$W''(1) - [W'(1)]^2 = V(w) \tag{29}$$

is said to be the lag variability.

The quantities of Definition 8 are analogous to the mean and variance of discrete random variables and have exactly the same interpretation. Note, however, that such quantities (although they may certainly still be defined) are meaningless when the lag structure is not normalizable.

Example 1. For the geometric lag structure we have

$$(30) \qquad W(t) = (1-\lambda) \sum_{i=0}^{\infty} (\lambda t)^i = \frac{1-\lambda}{1-\lambda t}, \quad m(w) = \frac{\lambda}{1-\lambda}, \quad V(w) = \frac{\lambda^2}{(1-\lambda)^2}$$

We observe that both mean lag - $m(w)$ - and lag variability - $V(w)$ - are increasing functions of λ.

Example 2. For the negative binomial (Pascal) lag distribution we have

$$(31) \qquad W(t) = \frac{(1-\lambda)^r}{(1-\lambda t)^r}, \quad m(w) = \frac{\lambda r}{1-\lambda}, \quad V(w) = \frac{\lambda^2 r}{(1-\lambda)^2}$$

Here mean lag and lag variability are increasing functions of both λ and r.

Example 3. Suppose that the general rational lag structure is normalizable and has, in fact, been normalized. Then,

$$(32) \qquad \begin{aligned} W(t) &= \frac{A(t)}{B(t)}, \quad m(w) = \frac{A'(1)}{A(1)} - \frac{B'(1)}{B(1)}, \\ V(w) &= \frac{A''(1)}{A(1)} - \left(\frac{A'(1)}{A(1)}\right)^2 - \frac{B''(1)}{B(1)} + \left(\frac{B'(1)}{B(1)}\right)^2 . \end{aligned}$$

3. Estimation of Distributed Lags.

a. Polynomial Lags

We begin by considering the class of polynomial lag structures since the estimation problems in this instance are rather simple.

The formulation is as follows: Given the model

$$(33) \qquad y_t = \sum_{\tau=0}^{n} w_\tau x_{t-\tau} + u_t , \quad t = 1,2,\ldots,T$$

with <u>known</u> n and subject to the conditions

$$(34) \qquad w_\tau = P(\tau) , \quad P(t) = \sum_{i=0}^{k} \beta_i t^i , \quad k < n$$

find efficient estimators of the lag coefficients.

Clearly, the solution to this problem is known under the usual assumptions on the error sequence $\{u_t : t = 1,2,\ldots\}$. Thus, if the latter is a sequence of independent identically distributed (i.i.d.) randam variables with mean zero and variance σ^2, the solution to the problem is as follows, provided that the u-sequence is also independent of the x-sequence.

Using (34) and (33) we observe that

$$(35) \quad y_t = \sum_{i=0}^{k} \beta_i z_{ti} + u_t, \quad z_{ti} = \sum_{i=0}^{n} \tau^i x_{t-i} \quad t = n+1, n+2,\ldots,T .$$

The Gauss-Markov theorem then yields that

$$(36) \quad \hat{\beta} = (Z'Z)^{-1} Z'y , \; Z = (z_{ti}), \; t = n+1, n+2,\ldots,T, \; i = 0,1,2,\ldots,k$$
$$y = (y_{n+1}, y_{n+2},\ldots,y_T)' \quad \beta = (\beta_0,\beta_1,\beta_2,\ldots,\beta_k)'$$

is the best linear unbiased estimate (BLUE) of the vector coefficient β. Deducing best linear unbiased estimators for the lag coefficients is comparatively easy since we have the relation

(37) $w = Q\beta \quad Q = (\tau^i) \quad \tau = 0,1,2,\ldots,n;\ i = 0,1,2,\ldots,k$.

and by convention $\tau^i = 1$ for $\tau = 0,\ i = 0$. It is, then, apparent from (37) that

(38) $$\hat{w} = Q\hat{\beta}$$

is the BLUE of w.

The distribution theory of polynomial lag estimators is rather straightforward; it represents only a slightly higher degree of complexity than that encountered in the usual regression model.

In empirical applications two issues are paramount: First, what is the length of the lag (i.e., what is n) and what is the order of the polynomial (i.e., what is k)? Second, is the specification in (34) compatible with the data?

For the first issue the discussion above proceeded from the premise that both n and k are known; subject to this a priori information one has an efficient procedure for estimating the unknown parameters. Typically, however, in applications neither n nor k are known but must be estimated simultaneously with the lag coefficients. Little systematic study has been directed toward the solution of such problems. The standard procedure is to select a rather narrow range for possible values of n and k estimate parameters under all admissible alternatives and choose the set of estimates that minimizes, say, the residual sum of squares. It may be shown that if such a priori restrictions are true, the procedure above yields the minimum chi sequare estimator. The second issue is discussed in Dhrymes [8, Ch. 8].

b. Geometric Lags

Here we shall consider in some detail the estimation issues raised by the geometric lag structure; the discussion of general rational lags will be considerably less complete.

The model we shall discuss first is

$$y_t = \frac{\alpha I}{I - \lambda L} x_t + u_t \qquad t = 1,2,\ldots,T \tag{39}$$

where $\{u_t : t = 1,2,\ldots\}$ is a sequence of i.i.d. random variables with mean zero and finite (absolute) moments of at least order eight. The exogenous variable is assumed, for the moment, to constitute a (uniformly) bounded sequence of fixed constants, having certain properties to be specified below.

It is evident from (39) that the basic problem here - as indeed in nearly all distributed lag estimation theory - is one of <u>nonlinearities</u>, and the genre of problems encompassed by the model

$$y_t = g_t(\theta) + u_t \tag{40}$$

contains all problems to be dealt with in this section as special cases. In (40) θ is a vector of unknown parameters and the subscript on the function symbol simply "dates" the observation. Referring to the model in (39) we have

$$\theta = (\alpha,\lambda)' \quad g_t(\theta) = \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i} . \tag{41}$$

Estimators of the unknown parameters of the model in (40) may be obtained as

$$\sup_{\theta \in \Theta} \sum_{t=1}^{T} [y_t - g_t(\theta)]^2$$

where Θ is the admissible parameter space. General conditions under which estimators so obtained are consistent are given in Jennrich [18] and Malinvaud [24]. Jennrich also obtains the conditions under which their asymptotic distribution is normal. An earlier discussion of similar problems may be found in Wald [34] and Wolfowitz [35].

Computational algorithms for obtaining such estimators have been examined by Steiglitz and McBride [33] and Dhrymes [6], [8] and a Monte Carlo comparison of their (small sample) properties is given in Morrison [27].

The minimum chi square (MCS) or maximum likelihood (ML) estimator - in the case where the u-process is assumed normal - is obtained as follows: Define

$$(42) \qquad a_0 = \sum_{i=0}^{\infty} \lambda^i x_{-i}\,, \quad x_t^* = \sum_{i=0}^{t-1} \lambda^i x_{t-i}$$

and observe that we can write

$$(43) \qquad y_t = a_0 \lambda^t + \alpha x_t^* + u_t\ .$$

For given λ one can compute

$$(44) \qquad \begin{pmatrix} \tilde{a}_0 \\ \tilde{\alpha} \end{pmatrix} = \begin{bmatrix} \lambda^2(1-\lambda^{2T})/(1-\lambda^2) & \Sigma\,\lambda^t x_t^* \\ \Sigma\,\lambda^t x_t^* & \Sigma\, x_t^{*2} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\,\lambda^t y_t \\ \Sigma\, x_t^* y_t \end{bmatrix}$$

treating a_0 as a parameter to be estimated; a_0, termed the truncation remainder, and its treatment as a parameter to be estimated were first introduced by Klein [21]. An any rate, frequently the economics of the problem would impose the constraint $\lambda \in [0,1)$ and it involves little loss of generality to assert $\lambda \in [0,1-\delta]$, $\delta > 0$; but then we can divide the interval by the points λ_i, $i = 0,1,2,\ldots,r$ such that $\lambda_0 = 0$, $\lambda_r = 1-\delta$ and for each λ_i compute the quantities in (44); the triplet $\tilde{\lambda}$, $\tilde{\alpha}$, $\tilde{a}_0$ which yields the smallest sum of residuals squared is the MCS or ML estimator, as the case may be. It should be remarked in this context that the estimator of a_0 is not a consistent one; indeed, provided $\lim_{T\to\infty}(\Sigma\, x_t^{*2}/T)$ is a positive quantity and $\lim_{T\to\infty}(\Sigma\, x_t^*/T)$ is bounded the matrix to be inverted in (44) is asymptotically singular. Thus, taking into account the truncation remainder is a practice suitable for small samples only; Morrison's results [27] indicate that, for small samples, taking a_0 into account yields appreciable improvement of the estimators for α and λ.

The Steiglitz-McBride, (SM) algorithm proceeds as follows: the first order conditions for the MCS estimator are

$$(45) \qquad \alpha\,\Sigma\,(y_t - \alpha x_t^*)\,\frac{\partial x_t^*}{\partial \lambda} = 0\,, \quad \Sigma\,(y_t - \alpha x_t^*)x_t^* = 0$$

where, now,

$$(46) \qquad x_t^* = \frac{I}{I-\lambda L} x_t = \sum_{i=0}^{\infty} \lambda^i x_{t-i}, \qquad \frac{\partial x_t^*}{\partial \lambda} = \frac{I}{I-\lambda L} x_{t-1}^* = x_{t-1}^{**} .$$

But from (46) we obtain the recursive relations

$$(47) \qquad x_t^* = \lambda x_{t-1}^* + x_t, \quad x_t^{**} = \lambda x_{t-1}^{**} + x_t^* .$$

Then, given λ, the quantities x_t^*, x_{t-1}^{**} can be easily computed from the data, provided initial conditions, x_0^*, x_0^{**} are available. The SM algorithm takes such initial conditions to be zero. This is a convenient assumption and may be shown to leave unaffected the asymptotic properties of the estimator. In fact, using this pair of initial conditions is equivalent to using x_t^* in the manner defined in (42) and neglecting the truncation remainder a_0.

The essential aspect of the SM algorithm is to put

$$y_t = \frac{I-\lambda L}{I-\lambda L} \lambda_t = y_t^* - \lambda y_{t-1}^*, \quad y_t^* = \frac{I}{I-\lambda L} y_t$$

and to write the equations in (45) as

$$(48) \qquad \begin{aligned} &\alpha \Sigma x_t^{*2} + \lambda \Sigma y_{t-1}^* x_t^* = \Sigma y_t^* x_t^* \\ &\alpha \Sigma x_t^* x_{t-1}^{**} + \lambda \Sigma y_{t-1}^* x_{t-1}^{**} = \Sigma y_t^* x_{t-1}^{**} \end{aligned}$$

Given an initial estimate of λ the starred quantities may be computed and thus the first iterate $\begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_1$ may be obtained; recomputing the starred quantities - using $\begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_1$ - we may obtain the second iterate and so on until convergence is attained. Unfortunately, no general theorem exists defining the conditions on the x-process that will produce convergence. It is clear, of course, that if convergence is attained then the converging iterate corresponds to a solution of the equations in (45).

Two facts are to be noted. First, the estimation procedure does not necessarily guarantee that the restrictions on λ will be respected; second, there is no guarantee that the solutions so arrived at will represent the minimum minimorum of the function to be minimized. All that is found by the iteration is a stationary point.

The (asymptotic) properties of the ML (or MCS) estimators - as the case may be - have been determined. Under suitable conditions on the x-sequence it may be shown that, asymptotically, such estimators obey

(49) $$\sqrt{T}\left[\begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix}\right] \sim N(0, \Phi)$$

where

(50) $$\Phi = \sigma^2 \left\{ \lim_{T \to \infty} \frac{1}{T} \begin{bmatrix} \Sigma x_t^{*2} & \alpha \Sigma x_t^{*} x_{t-1}^{**} \\ \alpha \Sigma x_{t-1}^{**} x_t^{*} & \alpha^2 \Sigma x_{t-1}^{**2} \end{bmatrix} \right\}^{-1}$$

and $\begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix}$ is the converging iterate.

Thus, provided we are prepared to rely on asymptotic theory the inference problem is completely solved. There are many variations of this basic model. The most obvious extension is to include additional explanatory variables none of which is subject to a distributed lag scheme entailing parameter nonlinearity. In this case the model is of the form

$$(51) \qquad y_t = \frac{\alpha_o I}{I - \lambda L} x_{to} + \sum_{i=1}^{n} \alpha_i x_{ti} + u_t$$

The estimation problems entailed by this model are of exactly the same type as those dealt with immediately above. A somewhat more complicated set of problems is presented by the model

$$(52) \qquad y_t = \frac{\alpha L}{I - \lambda L} x_t + \frac{I}{I - \rho L} \varepsilon_t$$

where $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \dots\}$ is a sequence of i.i.d. random variables with mean zero and variance σ^2.

We obtain MCS estimators by minimizing

$$(53) \qquad S = (y - \alpha x^*)' V^{-1}(y - \alpha x^*)$$

the symbols having the same meaning as before, where

$$(54) \qquad \text{Cov}(u) = \sigma^2 V, \; u = (u_1, u_2, \dots, u_t)', \; u_t = \frac{I}{I - \rho L} \varepsilon_t$$

$$V = (v_{ts}), \; v_{ts} = \frac{\rho^{|t-s|}}{1 - \rho^2} \quad t, s = 1, 2, \dots, T .$$

Evidently, estimators here may be obtained by a double search over the admissible parameter space for λ __and__ ρ. Thus, if we put $\theta = (\alpha, \lambda, \rho)'$ and $S_T(\theta)$ for the minimand in (53) divided by T, we obtain estimators by the operation

$$\sup_{\theta \in \Theta} [S_T(\theta)] .$$

Here the ML estimator differs slightly from the MCS one in that the former maximizes

$$L_T(\theta, y, x) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{T}{2\sigma^2} S_T(\theta) + \frac{1}{2} \ln(1 - \rho^2) . \tag{55}$$

Upon partial maximization with respect to σ^2 and substitution in (55) we find

$$L_T(\theta, y, x) = -\frac{T}{2} [\ln(2\pi) + 1] - \frac{T}{2} \ln\left[\frac{S_T(\theta)}{(1 - \rho^2)^{1/T}}\right] \tag{56}$$

and see that the ML estimator is obtained through the operation

$$\sup_{\theta \in \Theta} \frac{[S_T(\theta)]}{(1 - \rho^2)^{1/T}}$$

Typically, in both cases it involves little loss of empirical relevance and considerably simplifies the nature of the mathematical arguments, if we assume that $\lambda \in [0, 1 - \delta_1]$, $\rho \in [-1 + \delta_2, 1 - \delta_2]$, $\delta_1, \delta_2 > 0$. Bearing this in mind we see that for any admissible θ $\lim_{T \to \infty} (1 - \rho^2)^{1/T} = 1$. Consequently, we would expect the two estimators to be asymptotically equivalent - which is, in fact, the case.

The asymptotic properties of such estimators have been established. We shall not give, here, the details of the arguments involved; we shall, however, sketch a somewhat novel proof given in Dhrymes [7], [8, Ch. 6] which indicates that such estimators as discussed above converge to the corresponding parameters <u>not only in probability, but with probability one as well</u>. Thus, consider the MCS estimator defined by

$$S_T(\hat{\theta}) \leq S_T(\theta), \quad \forall \theta \in \Theta$$

It is clear that α can be eliminated by partial minimization, thus yielding

$$(57) \qquad S_T(\omega) = \frac{1}{T} y' \left[V^{-1} - \frac{V^{-1}x^*x^{*\prime}V^{-1}}{x^{*\prime}V^{-1}x^*} \right] y = \frac{1}{T} y'V^{-1}y \left[1 - \frac{(y'V^{-1}x^*)^2}{(x^{*\prime}V^{-1}x^*)(y'V^{-1}y)} \right]$$

where $\omega = (\lambda, \rho)'$.

Consequently, provided $x_t^* \neq 0$ for all t, we see that $S_T(\omega)$ is bounded away from zero and is bounded above by $\frac{1}{T} y' V^{-1} y$; thus any asymptotic properties we determine with respect to the MCS estimator will be applicable to the ML estimator as well.

To show convergence with probability one of the MCS estimator we proceed as follows: First, we show that, for any admissible ω, $S_T(\omega)$ converges to its limit, say $S(\omega)$, with probability one, <u>uniformly in</u> ω. Convergence with probability one means the following. Let $\mathcal{S}$ be the sample space over which the error process (u-process) is defined, and let s be the subset of $\mathcal{S}$ over which convergence does not hold.

Convergence with probability one means that the measure of s is zero. Uniformity in ω means that the limiting argument employed in the preceding does not depend on the (admissible) ω considered. Thus, we may say that "for almost all sequences $\{u_t : t = 1,2,\ldots\}$" $S_T(\omega)$ converges to $S(\omega)$ uniformly in $\omega \in \Omega$, Ω being the set of admissible ω-parameters.

Second, we observe that the sequence of estimators $\{\hat{\omega}_T\}$ such that $S_T(\hat{\omega}_T) \leq S_T(\omega) \;\forall\; \omega \in \Omega$ is a bounded infinite sequence and thus contains at least one limit point, say ω_*; moreover, there exists a subsequence $\{\hat{\omega}_{T_i}\}$ converging to ω_*. Third, we show that $\omega_* = \omega_o$, the latter being the true parameter value. Since the argument is valid for "almost all sequences $\{u_t : t = 1,2,\ldots\}$" and ω_* is any limit point, we conclude that $\hat{\omega}_T$ converges to ω_o with probability one provided a certain identifiability condition holds. Since the estimator of α is defined by

$$\hat{\alpha} = (x^{*\prime} V^{-1} x^*)^{-1} x^{*\prime} V^{-1} y \tag{58}$$

we also conclude that $\hat{\alpha}$ converges to α_0 with probability one. This argument may be readily extended to the case where the error process is an n^{th} order autoregression or the model contains additional explanatory variables none of which are subject to an (infinite) distributed lag.

Now we observe that

$$\frac{x^{*\prime} V^{-1} x^*}{T} = \frac{1}{T}(1-\rho^2)\, x_1^{*2} + \frac{1}{T} \sum_{t=2}^{T} (x_t^* - \rho\, x_{t-1}^*)^2 \tag{59}$$

and, thus, setting

$$(60) \qquad h_t(\omega) = x_t^* - \rho\, x_{t-1}^* = \sum_{i=0}^{\infty} \lambda^i (x_{t-i} - \rho\, x_{t-1-i})$$

we have that, provided the x-sequence is bounded

$$(61) \qquad |h_t(\omega)| \leq \frac{2K}{\delta_1}$$

K being the bound of the x-sequence. Moreover, if $\omega_1, \omega_2 \in \Omega$

$$(62) \qquad |h_t(\omega_2) - h_t(\omega_1)| \leq |\lambda_2 - \lambda_1| \frac{2K}{\delta_1^2} + |\rho_2 - \rho_1| \frac{K}{\delta_1}$$

which shows that $\{h_t\}$ is a sequence of uniformly bounded equicontinuous functions. Since

$$(63) \qquad |h_t^2(\omega_2) - h_t^2(\omega)| \leq |h_t(\omega_2) - h_t(\omega_1)|\ |h_t(\omega_2) + h_t(\omega_1)|$$

it follows that $\{h_t^2\}$ is similarly a sequence of uniformly bounded equicontinuous functions. The same is obviously true of

$$(64) \qquad H_T(\omega) = \frac{1}{T} \sum_{t=2}^{T} h_t^2(\omega)$$

But if the x-sequence is such that

$$(65) \qquad c(i,j) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_{t-i}\, x_{t-j}$$

exist as well defined finite quantities then the sequence $\{H_T(\omega)\}$

converges, pointwise, in ω; thus, by the Arzelà-Ascoli theorem [30] it also converges uniformly (in ω).

Next we observe

$$(65)\qquad \frac{1}{T} u'V^{-1}x^* = \frac{1}{T}\left[\sum_{t=1}^{T} u_t x_t^* - \rho \sum_{t=1}^{T-1} u_t x_{t+1}^* - \rho \sum_{t=1}^{T-1} u_{t+1} x_t^* + \rho^2 \sum_{t=2}^{T-1} u_t x_t^*\right]$$

and, thus, if each of the sums converges uniformly in λ then uniform convergence with respect to ρ is obvious. But it will suffice to show that $\frac{1}{T}\Sigma u_t x_t^*$ converges to its limit with probability one uniformly in λ. The other terms can be handled entirely similarly. But

$$(66)\qquad u_t = \rho_o^t u_o + \sum_{j=0}^{t-1} \rho_o^j \varepsilon_{t-j}, \quad x_t^* = \lambda^t x_o^* + \sum_{i=0}^{t-1} \lambda^i x_{t-i}$$

Consequently,

$$(67)\qquad \frac{1}{T}\sum_{t=1}^{T} x_t^* u_t = \frac{1}{T}\left[u_o x_o^* \sum_{t=1}^{T} (\lambda\rho_o)^t + u_o \sum_{t=1}^{T} \rho_o^t \sum_{i=0}^{t-1} \lambda^i x_{t-i}\right.$$
$$\left. + x_o^* \sum_{t=1}^{T} \lambda^t \sum_{j=0}^{t-1} \rho_o^j \varepsilon_{t-j} + \sum_{t=1}^{T} \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \lambda^i \rho_o^j x_{t-i} \varepsilon_{t-j}\right]$$

and it will be sufficient to show that the last term of right member converges to zero uniformly in λ, the other terms being handled similarly. To this effect, observe that changing the order of summation we have

$$(68) \qquad \frac{1}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\lambda^i\rho_o^j x_{t-i}\,\varepsilon_{t-j} = \sum_{i=0}^{T-1}\sum_{j=0}^{T-1}\lambda^i\rho_o^j W_{T,i,j}$$

where

$$(69) \qquad W_{T,i,j} = \frac{1}{T}\sum_{t=\max(i,j)+1}^{T} x_{t-i}\,\varepsilon_{t-j}$$

and it would obviously be sufficient to show that

$$\lim_{T\to\infty}\ \sup_{i,j\le T} |W_{T,i,j}|$$

converges to zero.

Using Chebyshev's inequality for eighth order moments we have

$$(70) \qquad \Pr\{|W_{T,i,j}| > r\} < \frac{E(W_{T,i,j})^8}{r^8}$$

But

$$(71) \qquad E(W_{T,i,j})^8 \le \frac{K_1}{T^4}$$

where K_1 is some bound independent of T, i or j.

Now, for any $r > 0$

$$(72) \qquad \sum_{T=1}^{\infty}\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}\Pr\{|W_{T,i,j}| > r\} \le K_2\sum_{T=1}^{\infty}\frac{1}{T^2} < \infty$$

and by the Borel-Cantelli lemma [4] we conclude

$$(73) \qquad \Pr\{W_{T,i,j} \mid > r, \text{ i.o.}\} = 0 .$$

This, however, easily implies that $\sup_{i,j \geq T} |W_{T,i,j}|$ converges to zero with probability one. Thus, we have established that $\frac{1}{T} \Sigma x_t^* u_t$ converges to zero with probability one, uniformly in λ <u>and</u> ρ_0.

Finally, we deal with the term

$$(74) \qquad \frac{1}{T} u'V^{-1}u = \frac{1}{T}(1-\rho^2)u_1^2 + \frac{1}{T}\left[\sum_{t=2}^{T} u_t^2 - 2\rho \sum_{t=2}^{T} u_t u_{t-1} + \rho^2 \sum_{t=1}^{T} u_t^2\right] .$$

It is clear that if we can show that $\frac{1}{T} \Sigma u_t^2$, $\frac{1}{T} \Sigma u_t u_{t-1}$, approach their respective limits with probability one, uniformity of convergence with respect to ρ will be obvious from (74). We note that

$$(75) \qquad E(u_t^2) = \frac{\sigma_o^2}{1-\rho_o^2}, \cdot E|u_t u_{t-1}| \leq \frac{\sigma_o^2}{1-\rho_o^2}, \quad \forall\ t$$

and moreover that $\{u_t^2 : t = 1,2,\ldots\}$, $\{u_t u_{t-1} : t = 2,3,\ldots\}$ are strictly stationary process. Consequently, by the Birkhoff-Khintchine theorem [11] the quantities $\frac{1}{T} \Sigma u_t^2$, $\frac{1}{T} \Sigma u_t u_{t-1}$ converge to their limits with probability one. But one easily shows that

$$(76) \qquad \operatorname*{plim}_{T \to \infty} \frac{1}{T} \Sigma u_t^2 = \frac{\sigma_o^2}{1-\rho_o^2}, \quad \operatorname*{plim}_{T \to \infty} \frac{1}{T} \Sigma u_t u_{t-1} = \frac{\sigma_o^2 \rho_o}{1-\rho_o^2}$$

and, consequently, the limits have been identified.

What the preceding argument has shown is that $S_T(\omega)$ converges to its limit, say, $S(\omega)$ with probability one uniformity in ω. Moreover, we have obtained that

$$(77) \qquad S(\omega) = \sigma_o^2 + \frac{\sigma_o^2(\rho - \rho_0^2)^2}{1-\rho_o^2} + \varphi(\omega)$$

where

$$(78) \qquad \varphi(\omega) = \alpha_o^2 \lim_{T\to\infty} \left[\frac{x_0^{*'}V^{-1}x_0^*}{T} - \frac{x_0^{*'}V^{-1}x^*(x^{*'}V^{-1}x_0^*/T)}{x^{*'}V^{-1}x^*} \right]$$

and the vector x_0^* is computed with respect to the true parameter λ_0. It is clear that for all admissible ω, $\varphi(\omega) \geq 0$ and, in particular,

$$(79) \qquad \varphi(\omega_o) = 0 .$$

Thus, we have that

$$(80) \qquad S(\omega_0) \leq S(\omega) , \quad \forall \omega \in \Omega .$$

Now, consider the subsequence $\{\hat{\omega}_{T_i}\}$ of the estimator sequence $\{\hat{\omega}_T\}$ the former converging to ω_* which is a limit point of the latter.

By construction we have

$$(81) \qquad S_{T_i}(\hat{\omega}_{T_i}) \leq S_T(\omega) \quad \forall \quad \omega \in \Omega .$$

Consequently,

$$(82) \qquad S(\omega_*) \leq S(\omega) \quad \forall \quad \omega \in \Omega$$

and in view of (80)

(83) $$S(\omega_*) = S(\omega_0) = \sigma_0^2$$

ω_0 being the true parameter vector $(\lambda_0, \rho_0)'$. But (38) immediately implies

(84) $$\rho_* = \rho_0 \quad \varphi(\omega_*) = 0 \; .$$

For the model to be identified we would require the condition that

(85) $$\varphi(\omega_*) = 0$$

implies

(86) $$\lambda_* = \lambda_0 \; .$$

If (85) and (86) are satisfied then we con lude that $\omega_* = \omega_0$ and, since ω_* is <u>any</u> limit point of $\{\hat{\omega}_T\}$, that $\{\hat{\omega}_T\}$ converges to ω_0 with probability one.

<u>Remark 2</u>. It is important to recognize that the identifiability condition in (85) and (86) is not vacuous. Thus, consider the x-sequence $x_t = (-1)^t$; it obeys all requirements on such sequences discussed earlier. On the other hand, we find

(87) $$x_t^* = \sum_{i=0}^{\infty} \lambda^i x_{t-i} = (-1)^t \sum_{i=0}^{\infty} (-\lambda)^i = (-1)^t \frac{1}{1+\lambda}, \; \lim_{T\to\infty} \frac{1}{T} x^{*\prime} V^{-1} x = \frac{1+\rho}{1+\lambda} \; .$$

Consequently

(88) $$\varphi(\omega) = \alpha_0^2 \left[\frac{(1+\rho)^2}{(1+\lambda_0)^2} - \frac{(1+\rho)^2}{(1+\lambda_0)^2} \right] = 0$$

which shows that $S(\omega)$ does <u>not</u> depend on λ; thus, no information is given by the argument above on the limit of the sequence $\{\hat{\lambda}_T\}$. For this particular x-sequence it is also verified that $\{\hat{\alpha}_T\}$ as defined in (58) does not converge to α_0, unless λ_0 is known and utilized in obtaining $\hat{\alpha}_T$.

If the true parameter ω_0 is an interior point of Ω then the asymptotic distribution of the MCS estimator is established through the expansion

$$\text{(87)} \qquad \frac{\partial S}{\partial \theta}(\hat{\theta}) = \frac{\partial S}{\partial \theta}(\theta_0) + \frac{\partial^2 S}{\partial \theta \partial \theta}(\bar{\theta})\,(\hat{\theta} - \theta_0)$$

where $\hat{\theta}$ is the MCS estimator of θ_0; the latter is the true parameter point and $\bar{\theta}$ obeys $|\bar{\theta} - \theta_0| \leq |\hat{\theta} - \theta_0|$.

The argument involves the use of a central limit theorem for m-dependent variables - see, e.g., Dhrymes [8, Ch. 4] - and the approximating argument given originally by Mann and Wald [25], particularly their Lemma 2. We then determine that, asymptotically,

$$\text{(88)} \qquad \sqrt{T}\,(\hat{\theta} - \theta_0) \sim N(0, C_{MCS})$$

where

$$\text{(89)} \qquad C_{MCS} = \sigma_0^2 \lim_{T \to \infty} \begin{bmatrix} \frac{x^{*\prime} V^{-1} x^*}{T} & \frac{1}{T}\alpha_0 x^{*\prime} V^{-1} \frac{\partial x^*}{\partial \lambda} & 0 \\ \frac{1}{T}\alpha_0 \left(\frac{\partial x^*}{\partial \lambda}\right)' V^{-1} x^* & \alpha_0^2 \left(\frac{\partial x^*}{\partial \lambda}\right)' V^{-1} \left(\frac{\partial x^*}{\partial \lambda}\right) & 0 \\ 0 & 0 & \frac{\sigma_0^2}{1-\rho_0^2} \end{bmatrix}^{-1}$$

it being understood that x^* and $\frac{\partial x^*}{\partial \lambda}$ are computed with respect to the true parameter point λ_0.

Remark 3. An interesting feature of the result just cited is that, asymptotically, the estimator for $(\alpha, \lambda)'$ is independent of that for ρ. This has the consequence that it is, here, possible to define a two step instrumental variables (IV) estimator having the same asymptotic distribution as the MCS (or ML) one. This discovery is due to Amemiya and Fuller [2]. To see how this eventuates observe that we can write the model as

$$(90) \qquad y_t = \lambda y_{t-1} + \alpha x_t + v_t, \qquad v_t = u_t - \lambda u_{t-1}.$$

Let

$$(91) \qquad \mathrm{Cov}(v) = \sigma^2 \Phi, \quad v = (v_2, v_3, \ldots, v_T)'.$$

If an initial consistent estimator is available for α, λ say $\tilde{\alpha}$, $\tilde{\lambda}$, then we can compute

$$(92) \qquad \tilde{u}_t = y_t - \tilde{y}_t, \quad \tilde{y}_t = \tilde{\alpha} \sum_{i=0}^{t-1} \tilde{\lambda}^i x_{t-i}, \quad \tilde{\rho} = \frac{\Sigma \tilde{u}_t \tilde{u}_{t-1}}{\Sigma \tilde{u}_{t-1}^2}$$

and thus obtain a consistent estimator of Φ, since the latter depends only on λ and ρ. It may then be shown that the estimator

$$(92) \qquad \begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_{IIV} = [(x, \tilde{y}_{-1})' \tilde{\Phi}^{-1} (x, y_{-1})]^{-1} (x, \tilde{y}_{-1})' \tilde{\Phi}^{-1} y$$

has the same asymptotic distribution as the marginal (asymptotic) distribution of the first two components of $\hat{\theta}$ as exhibited in (88) and (89).

Initial consistent estimates for α and λ are easily obtained from (90) by use of the instruments x_t, x_{t-1}.

Remark 4. It is very tempting to use the simplified procedure above in the context of the following model

$$y_t = \lambda y_{t-1} + \alpha x_t + \frac{I}{I - \rho L} \varepsilon_t \tag{93}$$

since upon solution, and if $|\lambda| < 1$, we obtain

$$y_t = \frac{\alpha I}{I - \lambda L} x_t + \frac{I}{(I - \lambda L)(I - \rho L)} \varepsilon_t \tag{94}$$

When we do, however, the resulting estimators do not have the same asymptotic distribution as the MCS estimator. The "reason" for this case, first discussed by Amemiya and Fuller [2] as well as Dhrymes [8] is that in the model of (93) λ is both a mean and a variance parameter for the y-process as is evident from (94). In the model considered previously λ was only a mean but not a variance parameter (for the y-process). This explanation becomes transparent if one recalls that in the case of normal distributions (which is the case asymptotically) mean and variance parameter estimators are mutually independent. In this connection, Maddala [23] has also pointed out that, in the case where the error process is normal, the information matrix is, appropriately, block diagonal in the first case but not in the case of the model exhibited in (93).

In closing the discussion of geometric lag structures it is instructive to consider the problems posed by the model in (93) and, in particular, the extent to which the IIV estimator considered earlier and other two step variants differ from the MCS estimator.

The MCS estimator of the parameters α, λ, ρ has the asymptotic distribution

$$\sqrt{T}\,(\hat{\theta} - \theta) \sim N(0, \sigma^2 \Omega^*) \tag{95}$$

where $\theta = (\alpha, \lambda, \rho)'$, $\hat{\theta}$ is the MCS estimator of θ and

$$\Omega^* = \lim_{T \to \infty} \left\{ \frac{1}{T} \begin{bmatrix} x'V^{-1}x & x'V^{-1}\bar{y}_{-1} & 0 \\ \bar{y}'_{-1}V^{-1}x & \bar{y}'_{-1}V^{-1}\bar{y}_{-1} + \dfrac{T\sigma^2}{1-\lambda^2} & \dfrac{T\sigma^2}{1-\lambda^2} \\ 0 & \dfrac{T\sigma^2}{1-\lambda\rho} & \dfrac{T\sigma^2}{1-\rho^2} \end{bmatrix} \right\}^{-1} \tag{96}$$

The important difference between the model in (93) and that in (52) is that, in the former, the estimators of ρ and (α, λ) are, even asymptotically, dependent. Consequently, how well ρ is estimated will have a bearing - even asymptotically - in how well (α, λ) is estimated. This is to be contrasted to the case of the model in (52) where the properties of the estimator of ρ - beyond consistency - are of no consequence in determining the (asymptotic) properties of the MCS estimator of (α,λ). Thus, we might expect that the IIV estimator considered earlier would not coincide asymptotically with the MCS estimator of (α,λ). This result is rather transparent, if we observe that the initial IV estimator of (α,λ) is, say,

$$(97) \qquad \begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix} = \left[\begin{pmatrix} x' \\ x'_{-1} \end{pmatrix} (x, y_{-1}) \right]^{-1} \begin{pmatrix} x'y \\ x'_{-1}y \end{pmatrix}$$

and the iterate is

$$(98) \qquad \begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_{IIV} = \left[\begin{pmatrix} x' \\ \tilde{y}'_{-1} \end{pmatrix} \tilde{V}^{-1} (x, y_{-1}) \right]^{-1} \begin{pmatrix} x'\tilde{V}^{-1}y \\ \tilde{y}'_{-1}\tilde{V}^{-1}y \end{pmatrix}$$

The estimate of V^{-1} is obtained through the estimate of ρ given by

$$(99) \qquad \tilde{\rho} = \frac{\sum_{t=3}^{T} \tilde{u}_t \tilde{u}_{t-1}}{\sum_{t=3}^{T} \tilde{u}_{t-1}^2}, \quad \tilde{u}_t = y_t - (x_t, y_{t-1}) \begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}, \qquad t = 2,3,\ldots,T .$$

The asymptotic distribution of the estimator in (98) is, quite clearly,

$$(100) \qquad \sqrt{T} \left[\begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_{IIV} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right] \sim N(0, \sigma^2 C_{IIV})$$

where

$$(101)\qquad C_{IIV} = \lim_{T\to\infty} \left\{ \frac{1}{T} \begin{bmatrix} x' V^{-1} x & x' V^{-1} \bar{y}_{-1} \\ \bar{y}'_{-1} V^{-1} x & \bar{y}'_{-1} V^{-1} \bar{y}_{-1} \end{bmatrix} \right\}^{-1}$$

and

$$(102)\qquad \bar{y}_{-1} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{T-1}) \quad \bar{y}_t = \frac{\alpha I}{I - \lambda L} x_t \ .$$

It can be shown that the marginal asymptotic distribution of the MCS estimator of (α,λ) is given by

$$(103)\qquad \sqrt{T}\left[\begin{pmatrix} \tilde{\alpha} \\ \tilde{\lambda} \end{pmatrix}_{MCS} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right] \sim N(0, C_{MCS})$$

where

$$(104)\qquad C_{MCS} = \lim_{T\to\infty} \left\{ \frac{1}{T} \begin{bmatrix} x' V^{-1} x, & x' V^{-1} \bar{y}_{-1} \\ \bar{y}'_{-1} V^{-1} x, & \bar{y}'_{-1} V^{-1} \bar{y}_{-1} + \dfrac{T\sigma^2 (\lambda-\rho)^2}{(1-\lambda^2)(1-\lambda\rho)^2} \end{bmatrix} \right\}^{-1}$$

Thus, the MCS estimator of (α,λ,ρ) implies an estimator which is (asymptotically) efficient relative to the IIV estimator of (α,λ) provided $\lambda \neq \rho$. When $\lambda = \rho$ the two estimators are equally efficient. It should be stressed, however, that in obtaining estimators in both instances no use is made of the condition $\lambda = \rho$. A number of other two step procedures for estimating the parameters of the model in (93) have been proposed - see for example Gupta [14] and Dhrymes [6]. In general such procedures are inferior to the MCS estimators in terms of asymptotic efficiency. A more extensive discussion of such aspects may be found in Dhrymes [8] and Grether and Maddala [12].

c. The General Rational Lag

As remarked earlier this formulation generalizes the polyonomial, geometric and negative binomial versions of the distributed lag model. The specification is of the form

$$y_t = \frac{A(L)}{B(L)} x_t + u_t \qquad t = 1,2,\dots,T \tag{105}$$

where

$$A(L) = \sum_{i=0}^{m} a_i L^i , \quad \sum_{j=0}^{n} b_j L^j , \quad b_0 \equiv 1 \qquad m \leq n . \tag{106}$$

An approach to the estimation of parameters in this model was put forth by Jorgenson [20]. The approach is as follows. Reducing the model we find

$$y_t = B^*(L) y_t + A(L) x_t + B(L) u_t \tag{107}$$

where, now,

$$B^*(L) = - \sum_{j=1}^{n} b_j L^j . \tag{108}$$

This is an autoregressive moving average model and Jorgenson's suggestion was to put

$$\varepsilon_t = B(L) u_t \tag{109}$$

assume that $\{\varepsilon_t : t = 1,2,\ldots\}$ is a sequence of i.i.d. random variables with mean zero and variance, say, σ^2 and estimate the unknown parameters by ordinary least squares regression. This approach, however, would be appropriate only if, in the orginial model specification of (105), the error process is of the form

$$u_t = \frac{I}{B(L)} \varepsilon_t \tag{110}$$

We recall from equations (18) and (20) that in Nerlove's formulation of the geometric lag model a similar condition held. However, we cannot credibly invoke (110) as a maintained hypothesis. If, as in (105), we invoke the customary assumptions regarding the error process then the application of ordinary least squares (regression) methods to (107) would yield <u>inconsistent</u> estimators since the right hand (explanatory) variables are no longer independent of (or even uncorrelated with) the error process in (107). Thus, as was the case with the (reduced) geometric lag model discussed earlier a different estimation procedure is required.

We first observe that if the error process in (105) is one of i.i.d. random variables then we can obtain estimators by minimum chi square (here nonlinear least squares) methods.

Thus, we consider the problem of minimizing

$$S(\theta) = \sum_{t=1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right)^2, \quad \theta = (a_0, a_1 \ldots a_m, b_1, b_2, \ldots, b_n)' \tag{111}$$

with respect to the unknown elements in $A(\cdot)$ and $B(\cdot)$. In effect the problem of showing the consistency and asymptotic normality of the estimator, say $\hat{\theta}$, obtained by the operation

$$\sup_{\theta} S(\theta)$$

can be solved by the methods employed in section 3.b. The problem here is that, if m is at all large, say, greater than two, a search technique is impractical. Thus, we are reduced to an iterative procedure, which may be obtained as follows: let

$$(112) \qquad a = (a_0, a_1, a_2, \ldots, a_m)', \quad b = (b_1, b_2, \ldots, b_n)'$$

and differentiate (111) to obtain[2] the "normal" equations

$$(113) \qquad \frac{\partial S}{\partial a} = -2 \sum_{t=1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right) \frac{L^i}{B(L)} x_t = 0 \quad i = 0,1,\ldots,m$$

$$(114) \qquad \frac{\partial S}{\partial b} = 2 \sum_{t=1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right) \frac{A(L)}{[B(L)]^2} L^j x_t = 0 \quad j = 1,2,\ldots,n .$$

Put

$$(115) \qquad y_t^* = \frac{I}{I - B^*(L)} y_t, \quad x_t^* = \frac{I}{I - B^*(L)} x_t, \quad x_t^{**} = \frac{A(L)}{I - B^*(L)} x_t^*$$

and rewrite (113) and (114) as

[2] The lower limit on the summation index is, of course, incorrect; in general, since the observations come in the form of pairs $\{(y_t, x_t) : t = 1,2,\ldots,T\}$ it would perhaps, have been simpler to carry the summation t between the limits $n + 1$ and T.

$$(116)\quad \sum_{t=1}^{T}\left(\sum_{i=0}^{m} a_i x^*_{t-i} - \sum_{j=1}^{n} b_j y^*_{t-j}\right)x^*_{t-i} = \sum_{t=1}^{T} y^*_t x^*_{t-i}, \quad i = 0,1,2,\dots,m$$

$$(117)\quad \sum_{t=1}^{T}\left(\sum_{i=0}^{m} a_i x^*_{t-i} - \sum_{j=1}^{n} b_j y^*_{t-j}\right)x^{**}_{t-j} = \sum_{t=1}^{T} y^*_t x^{**}_{t-j} \quad j = 1,2,\dots,n .$$

The interation procedure suggested by Steiglitz and McBride (SM) [33] is a straightforward extension of the procedure outlined in an earlier section when discussing the geometric lag model. Thus, if an initial <u>consistent estimator</u> of the vectors a, b is available, say $\tilde{a}, \tilde{b}$, then the quantities y^*_t, x^*_t, x^{**}_t can be computed (by recursion) using the relations in (115) and $\tilde{a}, \tilde{b}$. But then it is possible to solve equations (116) and (117) for the first iterate $\tilde{a}^{(1)}, \tilde{b}^{(1)}$. Using these we can recompute y^*_t, x^*_t, x^{**}_t and thus obtain the second iterate $\tilde{a}^{(2)}$, $\tilde{b}^{(2)}$ and so on until convergence is obtained. The converging iterate, say $(\hat{a},\hat{b})$ is a solution of (113) and (114). Since we have commenced the iteration with a consistent estimator it may be shown that every subsequent iterate is consistent and thus the converging iterate, if one exists, is a consistent root of the normal equations. Its asymptotic distribution can be established by the usual device of applying the mean value theorem to the gradient $\frac{\partial S}{\partial \binom{a}{b}}$.

Unfortunately, no general theorems exist concerning the precise conditions under which the iteration above will converge.

4. Concluding Remarks

Much of the preceding discussion was concerned with the formulation, interpretation and particularly the estimation problems posed by distributed lag models.

In effect, we have examined in some detail the rational distributed lag model and two of its important special cases, viz., the geometric and polynomial lag models. When examining such models we assumed the error process to be either one of i.i.d. random variables or a first order autoregression. However, there is no reason to restrict ours-selves in this matter. Indeed, an extension to the case where the error is a general linear (covariance stationary) process was given for the geometric lag by Hannan [15] and for the general rational lag by Dhrymes [8, Ch. 10].

A number of issues relating to tests of stability, goodness of fit and model selection in the context of distributed lag models are taken up in Dhrymes [8, Ch. 11] although no definitive solution is provided there. Indeed in the case of nonnested models no satisfactory procedure exists for choosing one model over another. Thus, e.g., if model one specifies a parameter space S_1 while model two specifies the parameter space S_2 and if $S_1 \not\subset S_2$, $S_2 \not\subset S_1$ (and otherwise the two models share the same specification) we have no entirely satisfactory method for choosing, on the basis of sample evidence, model one over model two and vice versa.

The problem of the small sample distribution of distributed lag estimators is also an inadequately explored topic. The essential problem lies in the complexity of the stimators which is a consequence of the strongly nonlinear character of most formulations .

Finally, the problems induced by misspecification are not fully understood. A good beginning in this direction is made in Sims [31].

Thus, while the literature of distributed lag models is a relatively extensive one it is by no means the case that the problems posed by such models and their empirical application are satisfactory resolved.

REFERENCES

1. Almon, S., "The Distributed Lag between Capital Appropriations and Expenditures", Econometrica, vol. 33 (1965), 178-196.
2. Amemiya, T. and W. Fuller, "A Comparative Study of Alternative Estimators in a Distributed Lag Model", Econometrica, vol. 35 (1967), 509-529.
3. Astrom, K. J. and P. Eykhoff, "Identification and Process Parameter Estimation", presented at the 1970 IFAC Symp. Identification, Prague.
4. Breiman, L., Probability, Reading, Mass., Addison Wesley, 1968.
5. Cagan, P., "The Monetary Dynamis of Hyperinflation", in Studies in the Quantity Theory of Money, M. Friedman (Ed.) University of Chicago Press, Chicago, 1946.
6. Dhrymes, P. J., "Efficient Estimation of Distributed Lags with Autocorrelated Errors", International Economic Review, vol. 10 (1969), 47-67.
7. ________, "On the Strong Consistency of Estimators for Certain Distributed Lag Models with Autocorrelated Errors", vol. 12 (1971) 329-343.

8. ____________, Distributed Lags: Problems of Estimation and Formulation, Holden-Day, San Francisco, 1971.

9. Dhrymes, P. J., L. R. Klein and K. Steiglitz, "Estimation of Distributed Lags", International Economic Review, vol. 11 (1970), 235-250.

10. Fisher, I, "A Note on a Short cut Method for Calculating Distributed Lags", Bulletin de L'Institut Internationale de Statistique, vol. 29 (1937), 323-328.

11. Gikhman, I. I. and A. V. Skorokhod, Introduction to the Theory of Random Processes, Philadelphia, W. B. Saunderi, 1969.

12. Grether, D. and G. S. Maddala, "A comparison of Several Two Step Distributed Lag Estimators", to appear in International Economic Review.

13. Griliches, Z., "Distributed Lags: A Survey", Econometrica, vol. 35 (1967), 16-49.

14. Gupta, Y. P., "Least Squares Variant of the Dhrymes Two Step Estimation Procedure of the Distributed Lag Model", International Economic Review, vol. 10 (1969), 112-113.

15. Hannan, E. J., "The Estimation of Relationships Involving Distributed Lags", Econometrica, vol. 33 (1965), 409-418.

16. Hoeffding, W. and H. Robbins, "The Central Limit Theorem for Dependent Variables", Duke Mathematical Journal, vol. 15 (1948), 773-780.

17. Houthakker, H. S. and L. D. Taylor, Consumer Demand in the United States, 1929-1970, Harvard University Press, Cambridge, Mass., 1966.

18. Jennrich, R. I., "Asymptotic Properties of Non Linear Least Squares", Annals of Mathematical Statistics, vol. 40 (1969), 633-643.

19. Jorgenson, D. W., "Anticipations and Investment Behavior" in The Brookings Quarterly Econometric Model of the United States, J. S. Duesenberry, G. Fromm, L. R. Klein, and E. Kuh (Eds) Rand McNally and Co., Chicago, 1965.

20. Jorgenson, D. W., "Rational Distributed Lag Functions", Econometrica, vol. 34 (1966), 135-149.

21. Klein, L. R., "The Estimation of Distributed Lags", Econometrica, vol. 26 (1958), 553-556.

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

23. Maddala, G. S., "Generalized Least Squares with an Estimated Covariance Matrix", Econometrica, vol. 39 (1971), 23-34.

24. Malinvaud, E., "The Consistency of Non Linear Regression", Annals of Mathematical Statistics, vol. 41 (1970), 956-969.

25. Mann, H. B. and A. Wald, "On the Statistical Treatment of Stochastic Difference Equations", Econometrica, vol. 11 (1943), 173-220.

26. McCarthy, M. D., "The Wharton Quarterly Econometric Forecasting Model, Mark III", Studies in Quantitative Economics No. 6, Economics Research Unit Department of Economics, University of Pennsylvania, 1972.

27. Morrison, J. L., "Small Sample Properties of Selected Distributed Lag Estimators: A Monte Carlo Experiment", International Economic Review, vol. 11 (1970), 13-23.

28. Nerlove, M., The Dynamics of Supply: Estimation of Farmers' Response to Price, The Johns Hopkins Press, Baltimore, 1958.

29. Rasche, R. H. and H. T. Shapiro, "The MIT-FRB Model: Its Special Features", American Economic Review, vol. 58 (1968), 123-149.

30. Royden, H. L., Real Analysis, McMillan, New York, 1963.

31. Sims, C., "Discrete Approximations to Continuous Time Distributed Lags in Econometrics", Econometrica, vol. 39 (1971), 545-563.

32. Solow, R. M., "On a Family of Lag Distributions", Econometrica, vol. 28 (1960), 393-406.

33. Steiglitz, K. and L. E. McBride, "A Technique for the Identification of Linear System", IEEE Transactions on Automatic Control, AC-10, 1965, 461-464.

34. Wald, A., "A Note on the Consistency of the Maximum Likelihood Estimate", Annals of Mathematical Statistics, vol. 20 (1949), 5595-601.

35. Wolfowitz, J., "On Wald's Proof of the Consistency of the Maximum Likelihood Estimate", Annals of Mathematical Statistics, vol. 20 (1949), 601-602.