INTERNATIONAL ECONOMIC REVIEW
Vol. 11, No. 2, June, 1970

ESTIMATION OF DISTRIBUTED LAGS*

BY PHOEBUS J. DHRYMES, LAWRENCE R. KLEIN
AND KENNETH STEIGLITZ[1]

1. PRELIMINARIES

THE USE OF DISTRIBUTED LAGS in econometric research is quite old. However, the current intensive interest in the subject dates back to the relatively recent work of Koyck [9]. Since then, a number of extensions have been made to the basic geometric lag distribution and the associated method of estimation taken up by Koyck. A number of such studies are summarized in Amemiya and Fuller [1]. More recently a search technique has been proposed by Dhrymes [2] for the case of a geometric lag distribution occurring in a relation characterized by a first order Markov process in its error term.

Jorgenson [7] has employed in empirical research rational distributed lags although he has not given a full treatment of the estimation problems of the parameters involved. Dhrymes [3] suggested a technique of estimating in a consistent and asymptotically efficient fashion the parameters of a rational distributed lag by the use of spectral techniques thus extending the results obtained by Hannan [5] in the case of the simple geometric lag structure.

At the same time, however, electrical engineers have been interested in much the same problems. In many instances they have produced the elements of a satisfactory solution to the problem of estimating the parameters of the rational distributed lag although their approach has not always been explicitly grounded on a statistical formulation and thus the properties of the resulting estimator were not clear. The present paper builds on an idea proposed by Steiglitz and McBride [13] in an engineering context.

Our purpose here is to give a rigorous formulation and solution to the problem of estimating, by maximum likelihood techniques, the parameters of a general lag structure, to point out the lines of research and terminology followed in the literature of electrical engineering and thus to make available this literature to econometricians. We believe that such contact will prove quite fruitful.

2. ENGINEERING MOTIVATION AND SOME FORMAL ANALOGIES IN ENGINEERING AND ECONOMETRIC RESEARCH

The determination of the dynamic characteristics of an electrical or mechanical system from observation records has been of interest to engineers for some time, especially with regard to the construction of adaptive or learning control systems. Most of the work in electrical engineering has been con-

* Manuscript received February 5, 1968; revised April 8, 1969.

[1] The work of Kenneth Steiglitz was supported in part by Army Research Office-Durham under Contract DA-HCO4-69-C-0012. Some of the computer facilities used were supported in part by NSF grant GP-579.

cerned with the linear stationary case, since many physical systems are operated in a fixed environment with relatively small excursions from quiescence.

Thus, assume that the postulated linear stationary system can be described completely at any time t by an n-dimensional vector w_t, called *state vector* at time t. Assume also that the system is excited by a single scalar variable x_t, called the *input*, and that the system response is determined by the value of the scalar variable y_t, called the output.[2] We may associate x_t with an exogenous or explanatory variable and w_t, y_t with endogenous variables, the general point of view being that x_t is determined outside the system and affects (or determines) w_t and y_t in a *causal way.*

Finally it will be assumed that output is a *linear combination of the components of the state vector and a scalar random variable* u_t, which represents the effect on the system of nonmeasurable or unknown exogenous variables. We many thus write

$$\begin{aligned} w_t &= Aw_{t-1} + bx_t \\ y_t &= c'w_t + u_t \, . \end{aligned} \tag{1}$$

In (1) A and b are respectively an $(n \times n)$ matrix and $n \times 1$ vector of constants, the assumption being that the transition from w_{t-1} to w_t is accomplished in a simple Markovian scheme, under the excitation induced by bx_t. For this reason the matrix A is called the *transition matrix* of the system, though it need not be a probability matrix.

Let z be complex, of unit modulus, and define

$$W(z) = \sum_{t=0}^{\infty} w_t z^{-t}, \; X(z) = \sum_{t=0}^{\infty} x_t z^{-t}, \; Y(z) = \sum_{t=0}^{\infty} y_t z^{-t}, \; U(z) = \sum_{t=0}^{\infty} u_t z^{-t} \, . \tag{2}$$

The first system of equations in (1) may now be written as

$$W(z) = Az^{-1}W(z) + bX(z) \, . \tag{2a}$$

Solving, we obtain

$$W(z) = (I - Az^{-1})^{-1}bX(z) \, . \tag{2b}$$

Finally substituting in the last equation of (1) we have

$$Y(z) = c'(I - Az^{-1})^{-1}bX(z) + U(z) \, . \tag{3}$$

Clearly $c'(I - Az^{-1})^{-1}b$ is a rational function of z^{-1} and as such it may be represented

$$c'(I - Az^{-1})^{-1}b = \frac{A(z^{-1})}{B(z^{-1})} \tag{3a}$$

where $A(z^{-1})$ and $B(z^{-1})$ are polynomials of suitable order. Clearly, the poles $(I - Az^{-1})^{-1}$ are the zeros of $B(z^{-1})$. Indeed, if λ_i are the roots of A then it can be easily seen that

[2] What follows can easily be extended to the multivariate case; i.e.. we can easily deal with the case in which x_t and y_t are vectors of suitable dimensions.

$$\frac{A(z^{-1})}{B(z^{-1})} = \sum_{i=1}^{n}\left[\frac{c_i^*}{\left(1-\frac{\lambda_i}{z}\right)}\right], \tag{4}$$

where c_i^* are suitable constants obtained through the process of expansion of the left hand side of (4) by partial fractions. Since the λ_i, $i = 1, 2, \cdots, n$ are less than unity in absolute value and z lies on the unit circle, we may expand

$$\frac{1}{1-\frac{\lambda_i}{z}} = \sum_{k=0}^{\infty}\left(\frac{\lambda_i}{z}\right)^k . \tag{4a}$$

Thus the system in (3a) can be written as

$$\begin{aligned}\sum_{t=0}^{\infty} y_t z^{-t} &= \sum_{i=1}^{n} c_i^* \sum_{t=0}^{\infty} z^{-t} \sum_{k=0}^{\infty} \lambda_i^k x_t z^{-k} + \sum_{t=0}^{\infty} u_t z^{-t} \\ &= \sum_{\tau=0}^{\infty}\left(\sum_{i=1}^{n} c_i^* \left(\sum_{k=0}^{\infty} x_{\tau-k}\lambda_i^k\right) z^{-\tau}\right) + \sum_{t=0}^{\infty} u_t z^{-t} ,\end{aligned} \tag{4b}$$

where the last member of (4b) is obtained by putting $\tau = t + k$. Equating like powers of z^{-1} on both sides we find

$$y_t = \sum_{i=1}^{n} c_i^* \sum_{k=0}^{\infty} \lambda_i^k x_{t-k} + u_t , \qquad t = 1, 2, \cdots, T , \tag{4c}$$

which shows that the system in (3a) represents a lag distribution. Moreover, this is a weighted sum of n simple geometric lag distributions with parameters λ_i, $i = 1, 2, \cdots, n$. Further, from (4b) we note that z^{-1} plays exactly the same role in engineering literature as the lag operator L in econometric and statistical literature.

In what follows we shall therefore use the lag operator exclusively. By definition

$$L^k x_t = x_{t-k} , \qquad L^0 = I , \qquad I x_t = x_t , \qquad k = 0, 1, 2, \cdots ; \tag{4d}$$

therefore, in virtue of (4) and (4a) we can write (4c) as

$$y_t = \sum_{i=1}^{n} c_i^* \sum_{k=0}^{\infty} (\lambda_i L)^k x_t + u_t = \sum_{i=1}^{n} \frac{c_i^* I}{(I-\lambda_i L)} \cdot x_t + u_t = \frac{A(L)}{B(L)} x_t + u_t , \tag{4e}$$

where $A(L)$ and $B(L)$ are polynomials of degree at most $n-1$ and n respectively. This is, of course, the notation of the standard rational distributed lag model discussed in the literature of econometrics. It is particularly striking, for example, that Jorgenson [7] and Steiglitz and McBride [13] use, in entirely different and unrelated contexts, exactly the same model (4e). This is an extreme instance of research convergence in econometrics and electrical engineering and should suggest to econometricians and electrical engineers the benefits to be derived from familiarity with certain aspects of the research in the two disciplines.

We conclude this section by giving a table of equivalent terminology.

Engineering	*Econometrics*
White-Gaussian error	Nonautocorrelated, normally distributed error
Colored-Gaussian	Autocorrelated, normally distributed error
Plant	Nonstochastic model
Record	Sample
Rational z transform	Rational lag distribution
Prefiltering	exponential weighting
Identification	Specification and estimation of a model

3. ESTIMATION OF THE GENERAL RATIONAL LAG MODEL

A. *Formulation.* In this section we shall deal with the problem of estimating the parameters of

$$(5) \qquad y_t = \frac{A(L)}{B(L)} x_t + u_t \,, \qquad t = 1, 2, \cdots, T \,,$$

where L is the lag operator defined in (4d) and

$$(5a) \qquad A(L) = \sum_{i=0}^{\mu} a_i L^i \,, \qquad B(L) = \sum_{j=0}^{\nu} b_j L^j \,, \qquad b_0 = 1 \,, \qquad \mu < \nu \,.$$

The independent variable x is assumed to be nonstochastic or, if stochastic, uncorrelated with the random term u_t. The latter has the specification

$$(5b) \qquad u \sim N(0, \sigma^2 I), \; u = (u_1, u_2, \cdots, u_T)' \,,$$

and is assumed to be independent of x for all t. We shall employ maximum likelihood methods. Thus the (log) likelihood function of the observations in (5) is given by

$$(6) \quad L(a, b, \sigma^2; y, x) = -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left(y - \frac{A(L)}{B(L)} x \right)' \left(y - \frac{A(L)}{B(L)} x \right)$$

where

$$y = (y_1, y_2, \cdots, y_T)' \,, \qquad x = (x_1, x_2, \cdots, x_T)'$$
$$a = (a_0, a_1, \cdots, a_\mu)' \,, \qquad b = (b_1, b_2, \cdots, b_\nu)' \,.$$

The first order conditions for a maximum are given by

$$(6a) \qquad \begin{aligned} \frac{\partial L}{\partial a_i} &= \frac{1}{\sigma^2} \sum_{t=\mu+\nu+1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right) \frac{L^j}{B(L)} x_t = 0 \,, & j = 0, 1, 2, \cdots, \mu \,, \\ \frac{\partial L}{\partial b_j} &= \frac{1}{\sigma^2} \sum_{t=\mu+\nu+1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right) \frac{A(L)}{B(L)^2} L^s x_t = 0 \,, & s = 1, 2, \cdots, \nu \,, \\ \frac{\partial L}{\partial \sigma^2} &= -\frac{T}{2} \frac{1}{\sigma^2} + \frac{1}{\sigma^4} \sum_{t=\mu+\nu+1}^{T} \left(y_t - \frac{A(L)}{B(L)} x_t \right)' \left(y_t - \frac{A(L)}{B(L)} x_t \right) = 0 \,, & \end{aligned}$$

We observe that while the equations in (6a) are highly nonlinear in the a_i and b_j, they are linear in a_i for given b_j and they are linear in σ^2. We can search the parameter space for estimates of a_i, given b_i or we can iterate for all parameter estimates simultaneously. Once we have estimates for a_i and b_j, however, an estimate for σ^2 will be easily obtained.

In what follows we shall concentrate our attention solely on the first two sets of equations which would correspond to the normal equations of least squares, although under the assumption of normality the resulting estimators would be maximum likelihood ones as well.

The strategy of our estimation procedure is to determine a consistent solution of the equations in (6a).

Since by a theorem of Huzurbazar [6], for large T there exists a unique consistent solution, and by a theorem of Wald [15] it corresponds to the global maximum of the likelihood function, we would then have found the maximum likelihood estimators of the parameters a_j, b_s, $j = 0, 1, 2, \cdots, \mu$, $s = 1, 2, \cdots, \nu$, which are asymptotically normal, unbiased, and efficient.

B. *An iterative algorithm.* Define

$$y_t^* = \frac{I}{B(L)} y_t\,, \qquad x_t^* = \frac{I}{B(L)} x_t\,, \qquad x_t^{**} = \frac{A(L)}{B(L)} x_t^*\,, \qquad t = 1, 2, \cdots, T\,, \tag{7}$$

and note that the first two systems in (6a) may be written as

$$\sum_{t=\mu+\nu+1}^{T} [B(L) y_t^* - A(L) x_t^*] x_{t-j}^* = 0\,, \qquad j = 0, 1, 2, \cdots, \mu\,, \tag{7a}$$

$$\sum_{t=\mu+\nu+1}^{T} [B(L) y_t^* - A(L) x_t^*] x_{t-s}^{**} = 0\,, \qquad s = 1, 2, \cdots, \nu\,. \tag{7b}$$

The equations in (7a) and (7b) are linear in the parameters provided we deal in the transformed variables y^*, x^*, x^{**}. But this suggests how we can solve the system: if an initial (consistent) estimator is given for $B(L)$ and $A(L)$, say $B^0(L)$, $A^0(L)$, it can be used to construct the variables y_t^*, x_t^*, x_t^{**}; then the system in (7a) and (7b) can be solved to provide another estimator $B^1(L)$, $A^1(L)$, etc., until the iteration converges, say at the k-th step; this would mean that for prescribed $\varepsilon > 0$

$$\max_j |a_j^k - a_j^{k+1}| < \varepsilon\,, \qquad \max_s |b_s^k - b_s^{k+1}| < \varepsilon\,. \tag{7c}$$

Let us see precisely what this algorithm entails. For computational convenience only, the various cross products involved will not contain the observations at times $r = 1, 2, \cdots, \nu$. To this effect, let

$$Y^* = \begin{bmatrix} y_{\mu+\nu}^*, & y_{\mu+\nu-1}^*, & \cdots, & y_{\mu+1}^* \\ y_{\mu+\nu+1}^*, & y_{\mu+\nu}^*, & \cdots, & y_{\mu+2}^* \\ \vdots & & & \vdots \\ y_{T-1}^*, & y_{T-2}^*, & \cdots, & y_{T-\nu}^* \end{bmatrix}, \quad X^* = \begin{bmatrix} x_{\mu+\nu+1}^*, & x_{\mu+\nu}^*, & \cdots, & x_{\nu+1}^* \\ x_{\mu+\nu+2}^*, & x_{\mu+\nu+1}^*, & \cdots, & x_{\nu+2}^* \\ \vdots & & & \vdots \\ x_T^*, & x_{T-1}^*, & \cdots, & x_{T-\mu}^* \end{bmatrix},$$
$$X^{**} = \begin{bmatrix} x_{\mu+\nu}^{**}, & x_{\mu+\nu-1}^{**}, & \cdots, & x_{\mu+1}^{**} \\ x_{\mu+\nu+1}^{**}, & x_{\mu+\nu}^{**}, & \cdots, & x_{\mu+2}^{**} \\ \vdots & & & \vdots \\ x_{T-1}^{**}, & x_{T-2}^{**}, & \cdots, & x_{T-\nu}^{**} \end{bmatrix}. \tag{8}$$

$$d = \begin{pmatrix} a \\ -b \end{pmatrix}, \quad c^* = (c_j^*), \quad c_j^* = \sum_{j=\mu+\nu+1}^{T} y_t^* x_{t-j}^*, \qquad j = 0, 1, 2, \cdots, \mu \tag{8a}$$

$$= \sum_{t=\mu+\nu+1}^{T} y_t^* x_{t+\mu-j}^{**}, \quad j = \mu+1, \mu+2, \cdots, \mu+\nu$$

$$W^* = \begin{bmatrix} X^{*\prime}X^* & X^{*\prime}Y^* \\ X^{**\prime}X^* & X^{**\prime}Y^* \end{bmatrix}. \tag{8b}$$

With the aid of the notation in (8), (8a) and (8b), the system in (7a) and (7b) may be written compactly as

$$W^*d = c^* . \tag{9}$$

Remark 1. We should observe that the definitions in (7), involving what engineers call prefiltering, are not as cumbersome as they appear. In particular, they need not involve power series expansions of the operators $I/B(L)$. To see this, note that by virtue of the convention in (5a), we can write

$$B(L) = I + \sum_{j=1}^{\nu} b_j L^j = I - B^*(L), \qquad B^*(L) = -\sum_{j=1}^{\nu} b_j L^j . \tag{10}$$

Thus, the first equation in (7) implies

$$y_t^* = y_t + B^*(L) y_t^* . \tag{10a}$$

If initial conditions for y_t^* are specified, we see that y_t^* can be computed recursively, given b, for as many *values of the index* as y_t is available.

Similar comments apply to x_t^*. In the case of x_t^{**} we see that since x_t^* is defined over the same range of index values as x_t, then clearly x_t^{**} is defined for $t = \mu+1, \mu+2, \cdots$. Thus, y_t^*, x_t^*, x_t^{**} are all defined for $t = \nu, \nu+1, \cdots, T$ which is the range of index values appearing in the matrices Y^*, X^* and X^{**}. Finally, a convenient initial condition for y_t^* and x_t^* is

$$x_{-i}^* = y_{-i}^* = 0 , \qquad i = 0, 1, 2, \cdots, \nu . \tag{10b}$$

Since

$$y_{-i}^* = \frac{I}{B(L)} y_{-i} , \qquad x_{-i}^* = \frac{I}{B(L)} x_{-i} \qquad i = 0, 1, 2, \cdots, \nu , \tag{10c}$$

(10b) is equivalent to stating that

$$x_{-i} = y_{-i} = 0 , \qquad i = 0, 1, 2, \cdots, \nu . \tag{10d}$$

This is not a serious handicap if the sample size T is large.

Remark 2. Observe that W^* and c^* are functions of d. In what follows we shall assume that for any admissible value of d, W^* is nonsingular. Now, it is clear from (9) that if an initial consistent estimator of d exists, say $\tilde{d}_0$, then W^* and c^* can be computed, say $\tilde{W}_0^*, \tilde{c}_0^*$, where the tilde is used to denote the fact that W^* is obtained on the basis of a consistent estimator of d and not on the basis of the true value of the parameter vector d.

In view of the above, we may iterate on d by solving, to obtain

$$\tilde{d}_1 = \tilde{W}_0^{*-1} \tilde{c}_0^* . \tag{11}$$

We may now evaluate W^* and c^* at $\tilde{d}_1$ and complete a second iteration

(11a) $$\tilde{d}_2 = \widetilde{W}_1^{*-1}\tilde{c}_1^* .$$

Two questions arise with respect to this procedure.

i. if $\tilde{d}_i$ is a consistent estimator of d, is it also the case that $\tilde{d}_{i+1}$ is a consistent estimator of d?

ii. does the iteration process converge, and, if so, under what circumstances?

To this effect we prove

LEMMA 1. *Under the hypotheses of the model as exhibited in* (5), (5a) *and* (5b), *if* $\tilde{d}_i$ *is a consistent estimator of* d *then so is* $\tilde{d}_{i+1}$, *the latter being defined by*

(12) $$\tilde{d}_{i+1} = \widetilde{W}_i^{*^{-1}}\tilde{c}_i^* .$$

PROOF. We note that by definition

(12a) $$y_t^* = \frac{I}{B(L)}y_t = \frac{I}{B(L)}\left[\frac{A(L)}{B(L)}x_t + u_t\right] = x_t^{**} + u_t^*$$

the starred quantities in (12a) having the obvious meaning. Thus

(12b) $$Y^* = X^{**} + U^*$$

where U^* is constructed in exactly the same fashion as Y^*. Define

(12c) $$W^{**} = \begin{bmatrix} X^{*\prime}X^* & X^{*\prime}\ X^{**} \\ X^{**\prime}X^* & X^{**\prime}X^{**}\end{bmatrix} .$$

Since

(12d) $$X^{*\prime}Y^* = X^{*\prime}X^{**} + X^{*\prime}U^* ,$$

then

(12e) $$\operatorname*{plim}_{T\to\infty}\frac{X^{*\prime}Y^*}{T} = \operatorname*{plim}_{T\to\infty}\frac{X^{*\prime}X^{**}}{T}$$

since x_t is nonstochastic (or independent of u_t).

Since $\tilde{d}_i$ is a consistent estimate of d, it follows that

(13) $$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\widetilde{W}_1^* = \lim_{T\to\infty}\frac{1}{T}W^{**} .$$

We also note that, by the same argument,

(13a) $$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\tilde{c}_i^* = \lim_{T\to\infty}\frac{1}{T}c^{**}$$

where

(13b) $$\begin{aligned} c^{**} = (c_j^{**}), c_j^{**} &= \sum_{j=\nu+1}^{T} x_t^{**}x_{t-j}^* , & j &= 0, 1, 2, \cdots, \mu , \\ &= \sum_{j=\nu+1}^{T} x_t^{**}x_{t+\mu-j}^{**} , & j &= \mu+1, \mu+2, \cdots, \mu+\nu . \end{aligned}$$

Since

(14) $$\tilde{d}_{i+1} = \left(\frac{\tilde{W}_i^*}{T}\right)^{-1} \frac{\tilde{c}_i^*}{T}$$

we conclude

(14a) $$\operatorname*{plim}_{T\to\infty} \tilde{d}_{i+1} = \lim_{T\to\infty}\left[\left(\frac{W^{**}}{T}\right)^{-1} \frac{c^{**}}{T}\right].$$

The model obeys

(15) $$B(L)y_t^* = A(L)x_t^* + u_t\,, \qquad t = 1, 2, \cdots, T\,.$$

Putting

(15a) $$y^\nu = (y_{\nu+1}, y_{\nu+2}, \cdots, y_T)'\,, \qquad u^\nu = (u_{\nu+1}, u_{\nu+2}, \cdots, u_T)'$$

we can write (15) in the slightly altered form

(15b) $$(X^*, Y^*)d = y^\nu - u^\nu\,.$$

Let

(15c) $$Z = (X^*, X^{**})$$

and consider

(15d) $$Z'(X^*, Y^*)d = Z'y^\nu - Z'u^\nu\,.$$

We note

(16) $$\lim_{T\to\infty} \frac{Z'(X^*, Y^*)}{T} = \lim_{T\to\infty} \frac{W^{**}}{T}\,,$$

(16a) $$\operatorname*{plim}_{T\to\infty} \frac{Z'y^\nu}{T} = \lim_{T\to\infty} \frac{1}{T}c^{**}\,, \qquad \operatorname*{plim}_{T\to\infty} \frac{Z'u^\nu}{T} = 0\,.$$

From (15d) we therefore obtain

(16b) $$\lim_{T\to\infty} \frac{W^{**}}{T}d = \lim_{T\to\infty} \frac{1}{T}c^{**}\,.$$

Comparing (14a) and (16b) we conclude

(16c) $$\operatorname*{plim}_{T\to\infty} d_{i+1} = d\,. \qquad \text{Q.E.D.}$$

The question of convergence for this procedure is rather difficult to settle definitively. Since asymptotically W^* converges to a positive definite matrix for every admissible d one would surmise that the iteration process will converge, at least for large T. Assuming that the process above is convergent, we are thus able to locate the consistent root of the maximum likelihood equations. By the theorem of Huzurbazar [6] we have therefore found, for large T, the global maximum of the likelihood function. Since the probability structure of the error term in (5) is regular, we conclude that such estimators are consistent, asymptotically efficient, and distributed as

(17) $$\sqrt{T}(\hat{d} - d) \sim N\left(0, \left[-\frac{1}{T}E\frac{\partial^2 L}{\partial d \partial d}\right]^{-1}\right).$$

One can handle rather easily the case of autocorrelated error terms porvided the autocorrelation is first order Markov. Thus, if

(17a) $$u_t = \rho u_{t-1} + \varepsilon_t,\ t = 1, 2, \cdots, T,\ |\rho| < 1,\ \varepsilon \sim N(0, \sigma_\varepsilon^2 I)$$

where

(17b) $$\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T)'\ ,$$

then one can employ a scanning (search) technique. As shown by Dhrymes [2] in a slightly different but relevent context the resulting estimators of $\begin{pmatrix} d \\ \rho \end{pmatrix}$, obtained by the procedure above coupled with a search on ρ, are consistent.

4. A MODEL WITH SEVERAL DISTINCT GEOMETRIC LAGS

It is interesting that the techniques of the previous discussion are easily applicable to the model

(18) $$y_t = \sum_{i=1}^{m} \frac{\alpha_i I}{I - \lambda_i L} x_{ti} + u_t\ . \qquad t = 1, 2, \cdots, T\ ,$$

which has been found intractable in previous economic applications. When the number of lags is small, say two, then the search technique given in Dhrymes [2] can easily be extended to produce maximum likelihood estimators in a relatively simple manner. If, however, $m > 2$, then the search technique is, realistically, nonapplicable and should resort to the estimation scheme discussed above.

Let us see precisely what this entails. As before we shall assume

(18a) $$u \sim N(0, \sigma^2 I)\ , \qquad u = (u_1, u_2, u_3, \cdots, u_T)'$$

and that the x_i, $i = 1, 2, \cdots, m$ are either nonstochastic or eventually independent of the error terms of (18). The (log) likelihood function of the observations is

(18b) $$L(\alpha, \lambda, \sigma^2; y, X) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{m}\frac{\alpha_i I}{I - \lambda_i L}x_{ti}\right)^2 .$$

The maximizing equations with respect to the α_i and λ_i are given by

(19) $$\begin{aligned} \frac{\partial L}{\partial \alpha_k} &= \sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{m}\frac{\alpha_i I}{I - \lambda_i L}x_{ti}\right)\frac{I}{I - \lambda_k L}x_{tk} = 0\ , \\ \frac{\partial L}{\partial \lambda_k} &= \sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{m}\frac{\alpha_k I}{I - \lambda_i L}x_{ti}\right)\frac{\alpha_i I}{(I - \lambda_k L)^2}x_{t-1,k} = 0\ . \end{aligned} \qquad k = 1, 2, \cdots, m\ ,$$

If we now define

(19a) $$x_{ti}^{*} = \frac{I}{I - \lambda_i L}x_{ti}\ , \qquad x_{ti}^{**} = \frac{I}{I - \lambda_i L}x_{ti}^{*}\ , \qquad y_{ti}^{*} = \frac{I}{I - \lambda_i L}y_t$$

the system in (19) may be written as

(19b) $$\sum_{i=1}^{m}\alpha_i \sum_{t=1}^{T} x_{ti}^{*}x_{tk}^{*} + \lambda_k \sum_{t=1}^{T} y_{t-1,k}^{*}x_{tk}^{*} = \sum_{t=1}^{T} y_{tk}^{*}x_{tk}^{*}$$

$$\sum_{i=1}^{m} \alpha_i \sum_{t=1}^{T} x_{ti}^{*} x_{t-1,k}^{**} + \lambda_k \sum_{t=1}^{T} y_{t-1,k}^{*} x_{t-1,k}^{**} = \sum_{t=1}^{T} y_{tk}^{*} x_{t-1,k}^{**} .$$

Two aspects of (19b) should be pointed out: first as an identity we may write, for any k,

$$y_t = \frac{I - \lambda_k L}{I - \lambda_k L} y_t = (I - \lambda_k L) y_{tk}^{*} = y_{tk}^{*} - \lambda_k y_{t-1,k}^{*} . \tag{19c}$$

Second, although the summation over t has the range $(1, T)$ we take $y_{0k}^{*} = x_{0k}^{*} = 0$ all k so that no problem arises.

The estimation scheme here is exactly the same as in previous sections; thus if consistent initial estimators exist for the λ_i, $i = 1, 2, \cdots, m$, say $\tilde{\lambda}_i^0$, then the quantities $\tilde{x}_{tk}^{*}$, $\tilde{x}_{tk}^{**}$, $\tilde{y}_{tk}^{*}$ can be computed from the expressions in (19a) where in lieu of λ_i we make use of the $\tilde{\lambda}_i^0$. Hence from (19b) we shall obtain estimators, say, $\tilde{\alpha}_i^1$, $\tilde{\lambda}_i^1$; using the $\tilde{\lambda}_i^1$ we can recompute the quantities $\tilde{x}_{tk}$, $\tilde{x}_{tk}^{**}$, $\tilde{y}_{tk}^{*}$ from the expressions in (19a) and from (19b) obtain another set of estimators, say $\tilde{\alpha}_i^2$, $\tilde{\lambda}_i^2$ and so on until convergence is obtained, i.e., until at the s-th step we find

$$\max \{ | \tilde{\lambda}_i^s - \tilde{\lambda}_i^{s-1} | , \quad | \tilde{\alpha}_i^s - \tilde{\alpha}_i^{s-1} | \} < \varepsilon \tag{19d}$$

where ε is a preassigned (small) positive constant.

5. AN ILLUSTRATION

Here we briefly examine the geometric lag distribution which has found extensive applications in econometrics. In this case

$$A(L) = \alpha I , \qquad B(L) = I - \lambda L , \qquad |\lambda| < 1 . \tag{20}$$

The model in (5) becomes

$$y_t = \frac{\alpha I}{I - \lambda L} x_t + u_t , \qquad t = 1, 2, \cdots, T . \tag{20a}$$

The equations in (7a) and (7b) become

$$\begin{aligned} &\sum_{t=1}^{T} [(I - \lambda L) y_t^{*} - \alpha x_t^{*}] x_t^{*} = 0 \\ &\sum_{t-1}^{T} [(I - \lambda L) y_t^{*} - \alpha x_t^{*}] \alpha x_{t-1}^{**} = 0 , \end{aligned} \tag{20b}$$

where

$$y_t^{*} = \frac{I}{I - \lambda L} y_t , \qquad x_t^{*} = \frac{I}{I - \lambda L} x_t , \qquad x_t^{**} = \frac{\alpha I}{I - \lambda L} x_t^{*} . \tag{20c}$$

After some rearrangement we can rewrite (20b) as

$$\begin{aligned} &\alpha \sum_{t=2}^{T} x_t^{*2} + \lambda \sum_{t=2}^{T} x_t^{*} y_{t-1}^{*} = \sum_{t=2}^{T} x_t^{*} y_t^{*} \\ &\alpha \sum_{t=1}^{T} x_{t-1}^{**} x_t^{*} + \lambda \sum_{t=2}^{T} x_{t-1}^{**} y_{t-1}^{*} = \sum_{t=2}^{T} x_{t-1}^{**} y_t^{*} . \end{aligned} \tag{21}$$

If we take an initial consistent estimator of λ and α, say $\tilde{\lambda}_0, \tilde{\alpha}_0$, then we can compute the prefiltered variables y_t^*, x_t^* and x_t^{**} recursively as follows:

$$y_t^* = y_t + \tilde{\lambda}_0 y_{t-1}^* , \qquad x_t^* = x_t + \tilde{\lambda}_0 x_{t-1}^* , \qquad x_t^{**} = \alpha x_t^* + \tilde{\lambda}_0 x_{t-1}^{**} . \tag{21a}$$

We can then solve the system in (21) to obtain another estimator, say $\tilde{\lambda}_1, \tilde{\alpha}_1$. We again compute the prefiltered variables in (20c) using the new estimators and continue until the iteration process converges, i.e., until

$$\max \{ | \tilde{\lambda}_{i+1} - \tilde{\lambda}_i | , \quad | \tilde{\alpha}_{i+1} - \tilde{\alpha}_i | \} < \varepsilon , \tag{21b}$$

where ε is some preassigned small quantity. An initial consistent estimator for α and λ can be obtained by instrumental variable techniques. In particular we can take the estimator proposed by Liviatan [10] which is obtained by solving

$$\begin{aligned} \tilde{\alpha}_0 \sum_{t=2}^{T} x_t^2 + \tilde{\lambda}_0 \sum_{t=2}^{T} x_t y_{t-1} &= \sum_{t=2}^{T} x_t y_t \\ \tilde{\alpha}_0 \sum_{t=2}^{T} x_t x_{t-1} + \tilde{\lambda}_0 \sum_{t=2}^{T} x_{t-1} y_{t-1} &= \sum_{t=2}^{T} x_{t-1} y_t . \end{aligned} \tag{21c}$$

Efficient estimation of the geometric lag distribution has been the subject of extensive research; a part of this literature was referred to in the introduction. In this connection, it should be noted that a recent Monte Carlo study by Morrison [12] compares a number of proposed estimators of the parameters of the model (20a) where the error terms are assumed to have the classical properties. He finds that the estimator proposed by Liviatan [10] and Hannan [5], as interpreted in the time domain by Amemiya and Fuller [1], on the whole do not do very well. The estimators proposed by Steiglitz and McBride [13], a variant of which was discussed above, does extremely well for large samples (50 observations); that proposed by Dhrymes [4], [2] performs relatively better than the Steiglitz and McBride estimator for smaller sample size, although for larger samples the two estimators perform equally well.

Finally, this is a convenient juncture to consider Malinvaud's comments [11] on the estimation of the geometric lag in the face of autocorrelated errors. Thus, suppose our model is

$$y_t = \frac{\alpha I}{I - \lambda L} x_t + u_t , \qquad u_t = \sum_{i=1}^{k} \rho_i u_{t-i} + \varepsilon_t , \qquad t = 1, 2, \cdots T . \tag{22}$$

where

$$\varepsilon \sim N(0, \sigma^2 I) , \qquad \varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T)' . \tag{22a}$$

We note that

$$y_t - \sum_{i=1}^{k} \rho_i y_{t-i} = \frac{\alpha I}{I - \lambda L}\left(x_t - \sum_{i=1}^{k} \rho_i x_{t-i} \right) + \varepsilon_t . \tag{22b}$$

We may then put

$$y_t' = \frac{\alpha I}{I - \lambda L} x_t' + \varepsilon_t , \qquad y_t' = y_t - \sum_{i=1}^{k} \rho_i y_{t-i} , \qquad x_t' = x_t - \sum_{i=1}^{k} \rho_i x_{t-i} . \tag{22c}$$

Malinvaud then claims that if we estimate α and λ by the method given in Klein [8] with y'_t, x'_t replacing y_t and x_t respectively, then the resulting estimators of α and λ are inconsistent. The iteration considered by Malinvaud begins with an inconsistent estimator of the parameters α, λ, ρ. It is simple enough to use Liviatan-type or other consistent estimators to start the interations.

Suppose that we have consistent estimators of the ρ_i, say $\hat{\rho}_{i0}$. Then we may define

$$y'_t = y_t - \sum_{i=1}^{k} \hat{\rho}_{i0} y_{t-i} , \qquad x'_t = x_t - \sum_{i=1}^{k} \hat{\rho}_{i0} x_{t-i} . \tag{23}$$

If an initial consistent estimator of α and λ are also available, then we can apply the scheme of this section with $\tilde{y}'_t$ and $\tilde{x}'_t$ replacing y_t and x_t in (20a). Thus we obtain estimators $\tilde{\alpha}_1, \tilde{\lambda}_1$. Using these we can compute

$$\tilde{u}_t - \tilde{\lambda}_1 \tilde{u}_{t-1} = y_t - \tilde{\lambda}_1 y_{t-1} - \tilde{\alpha}_1 x_t . \tag{23a}$$

From the left hand side of (23a) we can obtain recursively the $\tilde{u}_t$, $t = 1, 2, \cdots, T$, on the assumption, say, that

$$\tilde{u}_0 = 0 . \tag{23b}$$

The consequences of this assumption are minimal if the sample is at all large. Then we can regress $\tilde{u}_t$ on $\tilde{u}_{t-i}$, $i = 1, 2, \cdots, k$, to obtain another set ρ_{i1}, $i = 1, 2, \cdots, k$, and repeat the process. It is easily verified that this procedure will yield consistent estimators. Actually, in the empirically relevant case $k = 1$, one easily obtains a rather simply executed estimator which is consistent, asymptotically unbiased, and efficient. An alternative procedure if $k > 1$ may be as follows: Disregard the specification on u_t in (20) and obtain consistent estimators for α and λ by searching on λ. This may be done by using the form given in Klein [8]

$$y_t = \lambda^t \eta_0 + \alpha \sum_{i=1}^{t-1} \lambda^i x_{t-i} + u_t , \tag{24}$$

and employing ordinary least squares.

The resulting estimators of α, λ, say $\tilde{\alpha}_0, \tilde{\lambda}_0$, are consistent. Use the scheme of equations (23a) and (23b) to obtain the residuals $\tilde{u}_1, \tilde{u}_2, \cdots, \tilde{u}_T$. Then regress $\tilde{u}_t$ on $\tilde{u}_{t-i}$, $i = 1, 2, \cdots, k$, to obtain initial estimators of ρ_i say, $\hat{\rho}_{i0}$, $i = 1, 2, \cdots, k$. These are consistent estimators. Compute the quantities $\tilde{y}'_t, \tilde{x}'_t$ of (23) using the estimator $\hat{\rho}_{i0}$ above. Then consider

$$\tilde{y}'_t = \lambda^t \eta_0 + \alpha \sum_{i=1}^{t-1} \lambda^i \tilde{x}'_{t-i} + \tilde{\varepsilon}_t . \tag{24a}$$

This is asymptotically equivalent to

$$y_t - \sum_{i=1}^{k} \rho_i y_{t-i} = \lambda^t \eta_0 + \alpha \sum_{i=1}^{t-1} \lambda^i \left(x_t - \sum_{i=1}^{t-1} \rho_i x_{t-i} \right) + \varepsilon_t . \tag{24b}$$

Thus applying the search technique to (24a) in a least squares context yields asymptotically the maximum likelihood estimators of α and λ. One may, of course, iterate the procedure.

6. AN EXAMPLE

Here we shall apply the techniques developed in the previous sections to the problem of estimating the parameters of an investment function. Our purpose is not to give yet another theory of investment but rather to illustrate that the procedures developed have useful applications, and to indicate the extent of the variation in empirical results one might expect due to differences in estimation procedures. The example also demonstrates feasibility and convergence of the computational methods suggested. To this effect, we have chosen the investment function suggested by Jorgenson[2] with respect to the durable manufacturing sector. Our data are somewhat different from his, chiefly in that our sample period is 1948 (first qurater) to 1965 (fourth quarter) while his begin with 1948 and end with 1959. Aside from this both sets of data are comparable, and our results should be compared with the first row of Jorgenson's Table 2.2 in the work cited above. In Table 1 below I_t is Jorgenson's variable, investment at time t minus .0279 times capital stock at time $t-1$, and X_t is Jorgenson's variable $\Delta[p_t x_t/c_t]$, i.e., the change in the value of output divided by user cost.

TABLE 1

ESTIMATED INVESTMENT FUNCTION DURABLE MANUFACTURING, 1948.I-1965.IV

OLS:	$I_{t+3} = \dfrac{.007906 + .007944L + .0003197L^2}{1 - 1.541705L + .575882L^2} X_t$
Jorgenson*:	$I_{t+3} = \dfrac{.00096 + .00080L + .00034L^2}{1 - 1.29501L + .42764L^2} X_t$
Modified M. L.**+ (Instrumental variable) estimators	$I_{t+3} = \dfrac{.0023863 - .0007789L - .0012922L^2}{1 - 1.965438L + .972074L^2} X_t$
Maximum likelihood+ estimators	$I_{t+3} = \dfrac{.0018426 + .0001095L - .0015530L^2}{1 - 1.945464L + .952775L^2} X_t$

* Jorgenson's sample covers only 1948–1959.

** We shall explain the meaning of this below.

\+ The criterion of convergence employed in these computations has been the insensitivity of the residual sum of squares about its minimum.

The point estimates of the parameters of the hypothesized model might appear from Table 1 to be quite close no matter how we estimate them. However, their implications in terms of meaningful economic theoretic constructs are rather substantially different. Before we explore this let us stress again that we do not advance our new estimates above as alternative empirical characterizations of investment behavior, rather as illustrations how alternative estimation techniques can lead to substantially differing conclusions.

[2] Jorgenson, D. W., "Anticipations and Investment Behavior" in J. S. Duesenberry, G. Fromm, L. R. Klein, E. Kuh eds., *The Brookings Quarterly Econometric Model of the U. S.* (Chicago: Rand McNally, 1965).

First let us ask: What is the long run response of investment to the independent variable X_t? The answer is obtained by evaluating the rational functions of the table after replacing L by unity. The conclusions are: OLS: .05573, Jorgenson: .01583, Modified M. L.: .04749, M. L.: .05458. Without trying to explain the magnitude of these numbers—which in part reflect the units in which the variables are measured—we observe that simply by changing the sample period we obtain a more than threefold increase in this quantity. This is so since our OLS estimator is exactly like Jorgenson's eatimator, the only difference being the sample period. On the other hand, OLS, modified M. L. and M. L. procedures yield roughly comparable quantities.

Now, if the denominator polynomial is written as

$$B(L) = I + b_1L + b_2L^2 = (I - \lambda_1 L)(I - \lambda_2 L) \tag{25}$$

we have the identification

$$\lambda_1 + \lambda_2 = -b_1 \qquad\qquad \lambda_1\lambda_2 = b_2 \; . \tag{25a}$$

The four sets of results given in Table 1 imply the following estimators for λ_1, λ_2 respectively. OLS: .9043, .6347; Jorgenson: $.6475 \pm .1825i$ ($|\lambda|^2 = .4525$); Modified M. L.: $.9827 \pm .1234i$ ($|\lambda|^2 = .9809$); M. L.: $.9727 \pm .1616i$ ($|\lambda|^2 = .9722$). These results indicate considerable variation in the conclusions to be derived from the four sets of estimators. First, by enlarging the period of the sample we do not have oscillations in the lag coefficients, i.e., OLS yields real roots while Jorgenson results yield complex roots. Second the modified M. L. and M. L. estimators yield complex roots; moreover their modulus is very close to unity. In addition to that, in the last two sets we may well obtain *negative lag coefficients* due to the negative point estimators in the numerator polynomials. Of course, we have not appraised the statistical significance of these results, nor have we experimented with the order of the numerator polynomial so as to obtain the "best fitting" result as was the case with Jorgenson's study.

Finally, if we standardize the lag coefficients so that they add to unity we can obtain the implied mean lag as follows: Let

$$W(s) = \frac{\sum_{i=0}^{\mu} a_i s^i}{\sum_{j=0}^{\nu} b_j s^j} = \frac{A(s)}{B(s)} \tag{26}$$

be the lag generating function; it is apparent that

$$W(1) = \frac{A(1)}{B(1)} \tag{26a}$$

represents the sum of the lag coefficients. If *all lag coefficients are positive*, as must be the case in Jorgenson's model, then it makes perfectly good sense to divide the lag coefficients by $W(1)$ so that they lie in the interval $[0, 1]$ and sum to unity. Thus, they have all the characteristics of a set of probabilities, and we may define the mean lag in the same way as we define the mean of a random variable. In this case we obtain

(26b) $$\text{Mean lag} = \frac{A'(1)}{A(1)} - \frac{B'(1)}{B(1)}$$

where $A'(1), B'(1)$ indicate respectively the derivatives of $A(s), B(s)$ evaluated at $s = 1$. This measure is not useful in the case of the modified M. L. and M. L. estimators—at least not in the present case.

The mean lag for OLS is 15.16 quarters; for Jorgenson it is 7.02 quarters. This is indeed a very substantial variation and one that we might not expect to materialize simply by the enlargement of the sample period. However, it is not our purpose here to comment on this substantive aspect.

To conclude our discussion let us elucidate two aspects. First, by modified M. L. estimators we mean the following. The maximum likelihood (M. L.) estimators are obtained by (iteratively) solving the equations (7a) and (7b). If, however, we replace the quantities x_{t-s}^{**} by y_{t-s}^{*} then, in fact, we lighten the computational burden without losing consistency. Indeed, in view of the assumptions we make concerning the error term, the quantities y_{t-s}^{*} are not correlated with the error term and thus the estimators obtained (by iteration) from

(27) $$\begin{bmatrix} X^{*\prime}X^{*} & X^{*\prime}Y^{*} \\ Y^{*\prime}X^{*} & Y^{*\prime}Y^{*} \end{bmatrix} d = \bar{c}^{*}$$

where X^{*}, Y^{*} are as defined in (8) and

(27a) $$\begin{aligned} \bar{c}^{*} = (\bar{c}_j^{*}) \,, \quad \bar{c}_j^{*} &= \sum_{t=\mu+\nu+1}^{T} y_t^{*} x_{t-j}^{*} \,, \qquad j = 0, 1, 2, \cdots, \mu \\ &= \sum_{t=\mu+\nu+1}^{T} y_t^{*} y_{t+\mu-j}^{*} \,, \qquad j = \mu+1, \mu+2, \cdots, \mu+\nu \,, \end{aligned}$$

have an interpretation as instrumental variable estimators.[3] The advantage of making calculations with y_{t-s}^{*} instead of x_{t-s}^{**} is that the moment matrices of unknown coefficients (see equation (27)) are for each iteration symmetric and positive definite.

Second, we may obtain initial consistent estimators by an obvious extension of Liviatan-type methods or by using as initial instruments a suitable number of the principal components of *a set of lags in the independent variables*. This will have the effect of ameliorating the multicollinearity problems that are induced by using as instruments successive lags of the independent variable as Liviatan's method would suggest.

University of Pennsylvania and Princeton University, USA

REFERENCES

[1] Amemiya, T. and W. Fuller, "A Comparative Study of Alternative Estimators in a Distributed Lag Model," *Econometrica*, XXXV (July-October, 1967), 509-29.

[3] It should, of course, be noted that this is a less efficient estimator than the M. L. one.

[2] DHRYMES, P. J., "Efficient Estimation of Distributed Lags with Autocorrelated Error," *International Economic Review*, X (February, 1969), 41-67.

[3] ———, "Estimation of the General Rational Lag Structure by Spectral Techniques," Discussion Paper No. 35, Department of Economics, University of Pennsylvania.

[4] ———, "On the Treatment of Certain Recurrent Nonlinearities in Regression Analysis," *Southern Economic Journal*, XXXII (October, 1966), 187-96.

[5] HANNAN, E. J., "The Estimation of Relationships Involving Distributed Lags," *Econometrica*, XXXIII (January, 1965), 206-24.

[6] HUZURBAZAR, V. S., "The Likelihood Function, Consistency and the Maxima of the Likelihood Functions," *Annals of Eugenics*, XIV (1948), 185.

[7] JORGENSON, D. W., "Rational Distributed Lag Functions," *Econometrica*, XXXIV January, 1966), 135-49.

[8] KLEIN, L. R., "The Estimation of Distributed Lags," *Econometrica*, XXVI October, 1958), 553-65.

[9] KOYCK, L. M., *Distributed Lags and Investment Analysis* (Amsterdam: North Holland Publishing Company, 1954).

[10] LIVIATAN, N., "Consistent Estimation of Distributed Lags," *International Economic Review*, IV (January, 1963), 44-52.

[11] MALINVAUD, E., "The Estimation of Distributed Lags: A Comment," *Econometrica*, XXIX (July, 1961), 430-3.

[12] MORRISON, J. L., "Small Sample Properties of Selected Distributed Lag Estimators: A Monte Carlo Experiment," *International Economic Review*, IX (February, 1970), 13-32.

[13] STEIGLITZ, K. AND L. E. MCBRIDE, "A Technique for the Identification of Linear Systems," *IEEE Transactions on Automatic Control* Vol. AC-10, No. 4 (October, 1965), 461-4.

[14] ———, "Iterative Methods for Systems Identification," Technical Report No. 15, June, 1966, Department of Electrical Engineering, Princeton University.

[15] WALD, A., "Note on the Consistency of the Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, XX (1949), 595-601.

LINKED CITATIONS

Estimation of Distributed Lags
Phoebus J. Dhrymes; Lawrence R. Klein; Kenneth Steiglitz
International Economic Review, Vol. 11, No. 2. (Jun., 1970), pp. 235-250.
Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28197006%2911%3A2%3C235%3AEODL%3E2.0.CO%3B2-S

References

[1] **A Comparative Study of Alternative Estimators in a Distributed Lag Model**
Takeshi Amemiya; Wayne A. Fuller
Econometrica, Vol. 35, No. 3/4. (Jul. - Oct., 1967), pp. 509-529.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196707%2F10%2935%3A3%2F4%3C509%3AACSOAE%3E2.0.CO%3B2-7

[4] **On the Treatment of Certain Recurrent Non-Linearities in Regression Analysis**
Phoebus J. Dhrymes
Southern Economic Journal, Vol. 33, No. 2. (Oct., 1966), pp. 187-196.
Stable URL:
http://links.jstor.org/sici?sici=0038-4038%28196610%2933%3A2%3C187%3AOTTOCR%3E2.0.CO%3B2-D

[5] **The Estimation of Relationships Involving Distributed Lags**
E. J. Hannan
Econometrica, Vol. 33, No. 1. (Jan., 1965), pp. 206-224.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196501%2933%3A1%3C206%3ATEORID%3E2.0.CO%3B2-R

[7] **Rational Distributed Lag Functions**
Dale W. Jorgenson
Econometrica, Vol. 34, No. 1. (Jan., 1966), pp. 135-149.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196601%2934%3A1%3C135%3ARDLF%3E2.0.CO%3B2-6

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

[8] **The Estimation of Distributed Lags**
L. R. Klein
Econometrica, Vol. 26, No. 4. (Oct., 1958), pp. 553-565.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28195810%2926%3A4%3C553%3ATEODL%3E2.0.CO%3B2-1

[10] **Consistent Estimation of Distributed Lags**
Nissan Liviatan
International Economic Review, Vol. 4, No. 1. (Jan., 1963), pp. 44-52.
Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28196301%294%3A1%3C44%3ACEODL%3E2.0.CO%3B2-C

[11] **The Estimation of Distributed Lags: A Comment**
E. Malinvaud
Econometrica, Vol. 29, No. 3. (Jul., 1961), pp. 430-433.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196107%2929%3A3%3C430%3ATEODLA%3E2.0.CO%3B2-M

[12] **Small Sample Properties of Selected Distributed Lag Estimators**
J. Lawton Morrison, Jr.
International Economic Review, Vol. 11, No. 1. (Feb., 1970), pp. 13-23.
Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28197002%2911%3A1%3C13%3ASSPOSD%3E2.0.CO%3B2-D

[15] **Note on the Consistency of the Maximum Likelihood Estimate**
Abraham Wald
The Annals of Mathematical Statistics, Vol. 20, No. 4. (Dec., 1949), pp. 595-601.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28194912%2920%3A4%3C595%3ANOTCOT%3E2.0.CO%3B2-7