COMMITTEE ON ECONOMIC THEORY AND ECONOMETRICS

Working Paper No. 132

TRANSFORM METHODS FOR THE STUDY OF DISTRIBUTED LAGS*

by

Robert E. Hall

June 1968

*This research was supported by the Office of Emergency Planning through contract DACA31-67-C-0141, U.S. Army Corps of Engineers. This is Project Report No. 9 under that contract.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH

CENTER FOR RESEARCH IN MANAGEMENT SCIENCE

University of California, Berkeley

Transform Methods for the Study of Distributed Lags

by

Robert E. Hall

Transform methods for studying time-invariant dynamic systems have proven to be so useful for both theoretical and empirical work in other sciences that it is somewhat surprising that econometricians have made relatively little use of these methods. The association of transform methods with spectral analysis, which has yet to achieve any marked success in econometrics, may account for part of this unpopularity. The purpose of this paper is to present a general approach to the analysis of distributed lags, based on the use of Fourier transform methods. We hope to show that this approach gives the econometrician certain advantages in specifying and estimating distributed lag relationships.

Our emphasis is on the problem of obtaining the most precise estimates possible of the lag structure in the standard econometric distributed lag model. The plan of the paper is the following: In the first section, we present the basic Fourier transform method which we will use throughout the paper, in both continuous and discrete time. We state and prove the well-known convolution theorem, which shows that a distributed lag model has a much simpler structure in the frequency domain than in the time domain. We discuss the relation between the continuous-time formulation of a distributed lag model, which is convenient for some analytical purposes, and the discrete-time formulation, which is essential for practical empirical work. The two formulations are shown to be equivalent if the model in continuous time meets certain requirements of smoothness.

In the second section we turn to the problem of the specification of distributed lag functions. Some advantages of the Fourier transform approach in unifying the analytical and computational treatment of distributed lags are mentioned. A new lag specification, and some more flexible versions of old specifications, are discussed in terms of their Fourier transforms. There is no closed-form expression for these lag functions in discrete time, but they represent an interpolation of straightforward lag functions in continuous time.

The third section is devoted to the derivation of best estimators in the frequency domain for the case of a very simple stochastic specification. The problem of estimating nonlinear parameters is also touched upon. The section concludes with a discussion of certain derived estimates and their sampling properties.

In the fourth section we investigate the consequence of a more general stochastic specification. We call attention to some results on the efficiency of ordinary least squares when applied in the presence of autocorrelation; frequency domain methods are particularly convenient for studying this problem. Finally, we consider estimation methods which are more efficient than ordinary least squares. Our conclusion is skeptical of the value of elaborate methods of spectral analysis in estimating distributed lags. We finish with a proposal for a rather simple estimator which should provide quite efficient estimates of lag functions in the presence of disturbance processes of the kind usually found in econometrics.

1. Discrete and continuous time series

For reasons which should become clear in the course of this paper, we have chosen to write the basic distributed lag model in terms of variables which are measured continuously over time:

$$(1) \qquad y(t) = \int_0^T \beta(\tau)x(t - \tau)\, d\tau \quad .$$

Here $x(t)$ is the independent variable, $\beta(\tau)$ is the distributed lag function, and $y(t)$ is the dependent variable. We assume that we observe $x(t)$ and $y(t)$ over a period of observation $0 \leq t \leq T$. Furthermore, (here we depart from econometric convention) we assume that $x(t)$ and hence $y(t)$ are periodic functions of time, with period T, so that $x(t) = x(t - T) = x(t - 2T)....$ This assumption permits us to dispense with complicated corrections for end effects while retaining an exact theory. In empirical work it is justified if $x(0)$ is close to $x(T)$ and if $\beta(\tau)$ is close to zero for all values of greater than a small fraction of T. Both of these conditions are likely to be met in ordinary econometric work.

Next we define the Fourier transforms of these functions:

$$(1.2) \qquad X(\omega_j) = \int_0^T e^{-i\omega_j t} x(t)\, dt \quad ;$$

$$(1.3) \qquad B(\omega_j) = \int_0^T e^{-i\omega_j \tau} \beta(\tau)\, d\tau$$

$$(1.4) \qquad Y(\omega_j) = \int_0^T e^{-i\omega_j t} y(t)\, dt$$

for frequencies $\omega_o = 0$, $\omega_1 = \frac{2\Pi}{T}$, $\omega_{-1} = -\frac{2\Pi}{T}$, ..., $\omega_j = \frac{2\Pi j}{T}$, $\omega_{-j} = -\frac{2\Pi j}{T}$, ... The Fourier transform provides a useful alternative view of the function; its original form can be recovered from the Fourier transform by applying the inverse Fourier transform. For example, x(t) is given by

$$(1.5) \qquad x(t) = \frac{1}{T} \sum_{j=-\infty}^{\infty} e^{i\omega_j t} X(\omega_j)$$

The usefulness of the Fourier transform in the study of distributed lags is shown in the

Convolution Theorem

$$\text{If } y(t) = \int_0^T \beta(\tau) x(t - \tau)\, d\tau \quad ,$$

then

$$(1.6) \qquad Y(\omega_j) = B(\omega_j)\, X(\omega_j) \qquad \text{for all } j.$$

The proof is very simple.[1] First,

$$(1.7) \qquad \beta(\tau)x(t-\tau) = \frac{1}{T}\sum_{j=-\infty}^{\infty} e^{i\omega_j(t-\tau)}\beta(\tau)\,X(\omega_j) \quad .$$

Next we integrate both sides over τ to get $y(t)$:

$$(1.8) \qquad y(t) = \int_0^T \beta(\tau)x(t-\tau)$$

$$= \frac{1}{T}\sum_{j=-\infty}^{\infty} e^{i\omega_j t} X(\omega_j) \int_0^T e^{-i\omega_j\tau}\beta(\tau)\,d\tau$$

$$= \frac{1}{T}\sum_{j=-\infty}^{\infty} e^{i\omega_j t} X(\omega_j)\,B(\omega_j) \quad .$$

Thus $y(t)$ is seen to be equal to the inverse Fourier transform of $X(\omega_j)B(\omega_j)$, so its Fourier transform must be exactly $X(\omega_j)B(\omega_j)$, as asserted.

The next step in our approach is to reconcile the continuous view of time just presented to the empirical constraint that $x(t)$ and $y(t)$ are observed only at discrete intervals of time. For simplicity we will assume that $x(t)$ and $y(t)$ are sampled once per unit of time. We let $x_t = x(t)$ and $y_t = y(t)$ for $t = 1, 2, 3, \ldots, T$; from this point on we will assume that T is an odd integer. Our basic question is: Under what conditions is there a distributed lag model in discrete time,

---

[1]This proof is borrowed from Parzen [13], who discusses some of the ideas of this section from a rather different point of view.

$$(1.9) \qquad y_t = \sum_{\tau=0}^{T-1} \tilde{\beta}_\tau x_{t-\tau}$$

which is operationally equivalent to our original continuous-time model expressed in equation 1.1? In general, it is clear that such a discrete representation will not exist, since it fails to take account of the behavior of x(t) between the observed points, x_t. The requirement that the set of points x_t give a complete picture of x(t) is essentially a requirement that x(t) move smoothly and predictably as a function of time. One interpretation of smoothness which is particularly convenient for our present purpose is the assumption that x(t) be representable as the superposition of T harmonics:

$$(1.10) \qquad x(t) = \frac{1}{T} \sum_{j=-N}^{N} e^{i\omega_j t} \hat{X}(\omega_j) \quad ,$$

where $N = \frac{T-1}{2}$. It is easily verified that if x(t) has this property, its Fourier transform is

$$(1.11) \qquad X(\omega_j) = \hat{X}(\omega_j) \quad , \qquad |j| \leq N$$

$$= 0 \qquad \text{otherwise.}$$

Time series whose Fourier transforms vanish above a certain frequency are said to be bandlimited at that frequency. Our assumption, then,

is that x(t) is bandlimited at the frequency Π radians per unit of time.[1] We note as an immediate consequence of the convolution theorem that y(t) is bandlimited at the same frequency as x(t).

Next we define the discrete Fourier transform of the sequence x_t:

$$\hat{X}(\omega_j) = \sum_{t=1}^{T} e^{-i\omega_j t} x_t , \quad |j| \leq N . \tag{1.12}$$

An immediate question is: What is the relation between the discrete Fourier transform of x_t and the Fourier transform, $X(\omega_j)$, of x(t)? By substituting the inverse transform expression for x_t in equation 1.12, we get

$$\hat{X}(\omega_j) = \sum_{t=1}^{T} e^{-i\omega_j t} \frac{1}{T} \sum_{k=-\infty}^{\infty} e^{i\omega_k t} X(\omega_k) \tag{1.13}$$

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} X(\omega_k) \sum_{t=1}^{T} e^{i(\omega_k - \omega_j)t} .$$

Now

$$\sum_{t=1}^{T} e^{i(\omega_k - \omega_j)} = T \quad \text{if } \omega_k - \omega_j = 2\pi, \text{ where n is some integer} \tag{1.14}$$

$$= 0 \quad \text{for all other k,j.}$$

Thus

---

[1]One possible justification for this assumption is that if higher frequencies of an economic time series contained further information, data would be gathered more frequently. Thus national income is reported quarterly and stock prices are reported hourly.

$$(1.15) \qquad \tilde{X}(\omega_j) = \sum_{n=-\infty}^{\infty} X(\omega_j + 2\Pi n)$$

or the discrete transform is the sum of the values of the continuous transform at all frequencies which are indistinguishable if sampling is done only once per unit of time. The discrete transform of the sampled values does not give any useful information about $X(\omega_j)$ in the general case. But under our special assumption that x(t) is band-limited at the sampling frequency, all of the terms in the summation in equation 1.15 vanish except the term for n = 0, so we have

$$(1.16) \qquad \tilde{X}(\omega_j) = X(\omega_j) \qquad \text{for } |j| \leq N \ .$$

It is the fact that these two transforms are identical that establishes the crucial link between the underlying theoretical view of distributed lags, expressed in continuous time, and empirical work, which is necessarily carried out with discrete data.

Having derived the relation between the Fourier transforms of x(t) and x_t, we turn now to the relation between x(t) and x_t themselves. Under the bandlimiting assumption, x(t) is given by

$$(1.17) \qquad x(t) = \frac{1}{T} \sum_{j=-N}^{N} e^{i\omega_j t} X(\omega_j) \ .$$

For $X(\omega_j)$ we may substitute $\tilde{X}(\omega_j)$ as given in equation 1.12:

$$(1.18) \qquad x(t) = \frac{1}{T} \sum_{j=-N}^{N} e^{i\omega_j t} \sum_{\theta=1}^{T} e^{-i\omega_j \theta} x_\theta$$

$$= \frac{1}{T} \sum_{\theta=1}^{T} x_\theta \sum_{j=-N}^{N} e^{i\omega_j (t-\theta)}$$

Now

$$(1.19) \qquad \sum_{j=-N}^{N} e^{i\omega_j (t-\theta)} = \sum_{j=-N}^{N} \cos\,[\omega_j (t - \theta)]$$

$$= T \qquad \text{if } t = \theta$$

$$= 0 \qquad \text{if } t \text{ is an integer different from } \theta.$$

By defining the interpolation function, s(u), as

$$(1.20) \qquad s(u) = \frac{1}{T} \sum_{j=-N}^{N} \cos\, \omega_j u$$

we obtain the sampling theorem,

$$(1.21) \qquad x(t) = \sum_{\theta=1}^{T} s(t - \theta) x_\theta$$

That is, the original function x(t) can be reconstructed from a sample of T points by applying the weights given by the interpolation function. Except in the case of an integer value of t, all of the sample points

enter the expression for any single point of x(t).

The sampling theorem enables us to give a direct answer to the basic question raised earlier as to whether there is an operationally equivalent distributed lag model of the form

$$(1.9) \qquad y_t = \sum_{\tau=0}^{T-1} \tilde{\beta}_\tau x_{t-\tau} \ .$$

We start from the basic model,

$$(1.1) \qquad y(t) = \int_o^T \beta(\tau)\, x(t - \tau)\, d\tau$$

and substitute the expression for $x(t - \tau)$ from the sampling theorem to get

$$(1.22) \qquad y_t = y(t) \int_o^T \beta(\tau) \sum_{\theta=1}^{T} s(t - \tau - \theta) x_\theta \, d\tau$$

$$= \sum_{\theta=1}^{T} x_\theta \int_o^T \beta(\tau) s(t - \tau - \theta)\, d\tau \ .$$

Since we are restricting our attention to integer values of t, we can substitute $u = t - \theta$ to get

$$(1.23) \qquad y_t = \sum_{u=t-T}^{t-1} x_{t-u} \int_o^T \beta(\tau) s(u - \tau)\, d\tau$$

Now both x_t and $s(u)$ are periodic functions with period T, so we may add a multiple of T to u without affecting the value of y_t. If we add T to u for all values of u which are negative in equation 1.23 and change the order of summation, we get

$$(1.24) \qquad y_t = \sum_{u=0}^{T-1} x_{t-u} \int_0^T \beta(\tau)s(u - \tau)\, d\tau \quad .$$

Letting

$$(1.25) \qquad \tilde{\beta}_u = \int_0^T \beta(\tau)s(u - \tau)\, d\tau$$

we have the equivalent discrete form,

$$(1.26) \qquad y_t = \sum_{u=0}^{T-1} \tilde{\beta}_u x_{t-u} \quad .$$

The answer to our basic question, then, is in the affirmative--there is a discrete model which is exactly equivalent to our original continuous model.

We recall that the convolution theorem establishes that

$$(1.6) \qquad Y(\omega_j) = B(\omega_j) X(\omega_j)$$

It is possible to show that the convolution theorem also holds for discrete Fourier transforms; this establishes a similar result for the discrete model:

$$(1.27) \qquad \tilde{Y}(\omega_j) = \tilde{B}(\omega_j)\,\tilde{X}(\omega_j) \quad .$$

But since $\tilde{X}(\omega_j) = X(\omega_j)$ and $\tilde{Y}(\omega_j) = Y(\omega_j)$ for all $|j| \leq N$, and equation 1.27 is assumed to hold identically for all $\tilde{X}(\omega_j)$, we have

$$(1.28) \qquad \tilde{B}(\omega_j) = B(\omega_j) \quad , \qquad |j| \leq N \quad .$$

Thus $\tilde{\beta}_\tau$ is given by the inverse discrete Fourier transform of $B(\omega_j)$:

$$(1.29) \qquad \tilde{\beta}_\tau = \frac{1}{T} \sum_{j=-N}^{N} e^{i\omega_j \tau} B(\omega_j) \quad .$$

This is an alternative formula for $\tilde{\beta}_\tau$; it is clearly preferable for computational purposes since it replaces an integral with a finite summation. In addition, equation 1.29 shows that $\tilde{\beta}_\tau$ is a smoothed version of $\beta(\tau)$ obtained by dropping off the higher frequencies in the infinite sum of harmonics which describes $\beta(\tau)$. As we shall show in the next section, this formula gives a very natural answer to some questions of smoothing and interpolation which arise in the specification of distributed lags.

2. The specification of distributed lags

The recent proliferation of distributed lag specifications (a partial bibliography appears in [8]) has made it clear that a unified treatment of alternative specifications would have many analytical and computational advantages. Analytically, it would make it unnecessary to derive new formulas for deducing lag coefficients from estimated coefficients or new formulas for sampling properties, mean lags, and the like; that is, it would provide a general solution to the unscrambling problem. Computationally, it would make it unnecessary to develop a new computer program for each new lag specification.

The Fourier transform approach appears to provide both of these advantages, and several others as well. In particular, it offers a solution to the problem of interpolating between the integer values of certain parameters of discrete lag distributions. The simplest of such problems arises in the rectangular lag distribution, defined as

$$\beta_\tau = \alpha_1 \qquad \text{for } 0 \leq \tau \leq p-1 \tag{2.1}$$

$$= 0 \qquad \text{otherwise.}$$

Here the parameter, p, gives the number of periods covered by the distribution; it is restricted to integer values. Since ordinary estimation methods require that a parameter for estimation be able to take on any real value, it is conventional to assume a value for p rather than to estimate it.

If our underlying view of the lag mechanism is formulated in terms of continuous time, however, there is a natural solution to the problem of defining a real-valued parameter which is in some sense equivalent to p. In continuous time, the rectangular distribution is

$$\beta(\tau) = \alpha_1 \quad \text{if } 0 \leq \tau \leq \alpha_2 \tag{2.2}$$

$$= 0 \quad \text{otherwise.}$$

The parameter which is equivalent to p is labeled α_2 to indicate that it is to be estimated along with α_1. The Fourier transform of this distribution is

$$B(\omega_j) = \int_0^{\alpha_2} e^{-i\omega_j \tau} \alpha_1 \, d\tau \tag{2.3}$$

$$= \alpha_1 \frac{1 - e^{-i\omega_j \alpha_2}}{i\omega_j}$$

The equivalent discrete lag distribution, $\tilde{\beta}_\tau$, represents one solution to the interpolation problem. In terms of Fourier transforms, the rectangular distributed lag model is:

$$Y(\omega_j) = \alpha_1 \frac{1 - e^{-i\omega_j\alpha_2}}{i\omega_j} X(\omega_j) \tag{2.4}$$

For particular values of α_1 and α_2 we could calculate the corresponding discrete lag function, $\tilde{\beta}_\tau$, by applying the inverse discrete Fourier transform,

$$\tilde{\beta}_\tau = \frac{1}{T} \sum_{j=-N}^{N} [e^{i\omega_j\tau}][\alpha_1 \frac{1 - e^{-i\omega_j\alpha_2}}{i\omega_j}] \tag{2.5}$$

The relation between $\beta(\tau)$ and β_τ is shown in Figure 1 for a representative case.

The ability to estimate parameters which represent distances along the lag axis suggests a lag distribution specification of considerable flexibility. This is the step function,

$$\begin{aligned} \beta(\tau) &= \alpha_1 && \text{for } 0 \leq \tau \leq \alpha_2 \\ &= \alpha_3 && \text{for } \alpha_2 < \tau \leq \alpha_4 \\ &\vdots \\ &= \alpha_{M-1} && \text{for } \alpha_{M-2} < \tau \leq \alpha_M . \end{aligned} \tag{2.6}$$

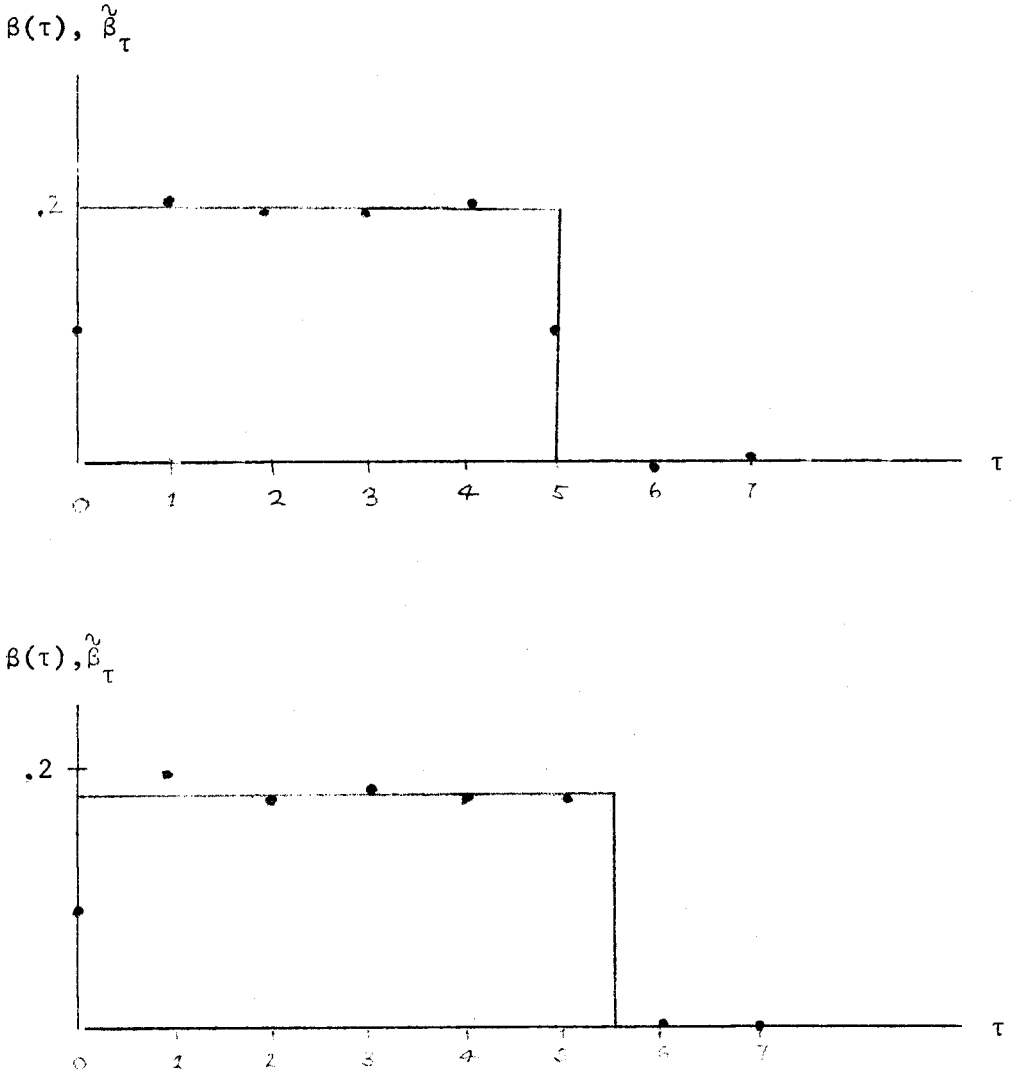

Figure 1. Continuous and discrete representations of the rectangular distributed lag function. Parameter values are T = 41, α_1 = .200, and α_2 = 5.0 in the upper plot and T = 41, α_1 = .182 and α_3 = 5.5 in the lower plot.

The Fourier transform of the step function distributed lag is

$$(2.7) \qquad B(\omega_j) = \alpha_1 \int_0^{\alpha_2} e^{-i\omega_j \tau}\, d\tau + \ldots + \alpha_{M-1} \int_{\alpha_{M-2}}^{\alpha_M} e^{-i\omega_j \tau}\, d\tau$$

$$= \frac{1}{i\omega_j} \left[\alpha_1 + (\alpha_3 - \alpha_1)\, e^{-i\omega_j \alpha_2} + \ldots \right.$$

$$+ (\alpha_{M-1} - \alpha_{M-3})\, e^{-i\omega_j \alpha_{M-2}}$$

$$\left. - \alpha_{M-1} e^{-i\omega_j \alpha_M} \right] .$$

Two advantages of this specification relative to existing specifications should be mentioned. First, the total length of the distribution, α_M, is a parameter which can be estimated along with the parameters governing the height of the lag distribution at each point. Second, the parameters giving the locations of the steps in the lag function, α_2, α_4, ..., α_{M-2}, can lie anywhere in the interval from 0 to α_M. This enables the step function specification to concentrate its flexibility in the region where the underlying empirical lag function requires it. By contrast, the well-known Almon specification [1], which might be used for the same kind of problem as this specification, has a fixed length and equal distribution of its flexibility along the lag axis.

All of the commonly-used distributed lag function specifications have counterparts in the present framework. The counterpart of the Koyck or geometric lag is exponential in form:

$$\beta(\tau) = \alpha_1 e^{-\alpha_2 \tau} \quad ; \tag{2.8}$$

its Fourier transform is

$$\mathcal{B}(\omega_j) = \frac{\alpha_1}{\alpha_2 + i\omega_j} \quad . \tag{2.9}$$

Similarly, the counterpart of the rational distributed lag has a Fourier transform which is a generalization of formula 2.9:

$$\mathcal{B}(\omega_j) = \frac{\alpha_1}{(\alpha_2 + i\omega_j)\ldots(\alpha_M + i\omega_j)} \quad ; \tag{2.10}$$

its special case, the Pascal distributed lag, has the following Fourier transform:

$$\mathcal{B}(\omega_j) = \frac{\alpha_1}{(\alpha_2 + i\omega_j)^K} \tag{2.11}$$

The Almon or polynomial lag function,

$$\beta(\tau) = \alpha_1 + \alpha_2\tau + \alpha_3\tau^2 + \ldots + \alpha_{M-1}\alpha^{M-2} \qquad (2.12)$$

$$\text{for } 0 \leq \tau \leq \alpha_M$$

$$= 0 \qquad \text{otherwise,}$$

has a Fourier transform which can be expressed conveniently in the following way. Let

$$\psi(x) = \int_0^{\alpha_M} e^{x\tau}\, d\tau \qquad (2.13)$$

$$= \frac{e^{\alpha_M x} - 1}{x} \quad .$$

Then

$$B(\omega_j) = \alpha_1 \int_0^{\alpha_M} e^{-i\omega_j\tau}\, d\tau + \alpha_2 \int_0^{\alpha_M} \tau e^{-i\omega_j\tau}\, d\tau \qquad (2.14)$$

$$+ \alpha_3 \int_0^{\alpha_M} \tau^2 e^{-i\omega_j\tau}\, d\tau + \ldots$$

$$+ \alpha_{M-1} \int_0^{\alpha_M} \tau^{M-2} e^{-i\omega_j\tau}\, d\tau$$

$$= \alpha_1\psi(-i\omega_j) + \alpha_2\psi'(-i\omega_j) + \alpha_3\psi''(-i\omega_j) + \ldots$$

$$+ \alpha_{M-1}\psi^{(M-2)}(-i\omega_j) \quad .$$

In most distributed lag studies the choice among alternative lag specifications is purely empirical. One of the advantages of the Fourier transform approach to distributed lag analysis is that the estimating equation,

$$Y(\omega_j) = B(\omega_j)\, X(\omega_j) + U(\omega_j) \tag{2.15}$$

has exactly the same form for all specifications. The alternatives can be compared to each other under precisely similar conditions, and with the same computer program. Furthermore, the estimating equation is of a form which is particularly convenient for nonlinear estimation--only $X(\omega_j)$ enters the formula for $Y(\omega_j)$, while in the time-domain distributed lag model, not only x_t but also x_{t-1}, x_{t-2}, ... enter the formula for y_t. It is impossible to set up a general procedure for estimating distributed lags in the time domain, but quite simple in the frequency domain.

3. Stochastic specification and estimation

The natural stochastic specification for our model expressed in discrete time is

$$(3.1) \qquad y_t = \sum_{\tau=0}^{T-1} \tilde{\beta}_\tau x_{t-\tau} + u_t \quad , \qquad 1 \leq t \leq T$$

where the usual hypotheses about the disturbances are assumed to hold: $V(u) = \sigma^2 I$, u being the vector of values of u_t. If we take the discrete Fourier transform of both sides of equation 3.1, we get

$$(3.2) \qquad Y(\omega_j) = B(\omega_j)\, X(\omega_j) + U(\omega_j) \quad , \qquad |j| \leq N \quad .$$

Suppose, for the time being, that the systematic part of the right hand side of equation 3.2 can be written as

$$(3.3) \qquad B(\omega_j)\, X(\omega_j) = \sum_{k=1}^{M} \alpha_k Z_k(\omega_j)$$

that is, the distributed lag specification underlying $B(\omega_j)$ has M parameters, $\alpha_1, \ldots, \alpha_M$, each of which enters linearly. For example, in the case of the Almon lag with fixed length,

$$(3.4) \qquad Z_k(\omega_j) = \psi^{(k-1)}(-i\omega_j)\, X(\omega_j) \quad .$$

Then if we let Y denote the vector of length T of values of $Y(\omega_j)$, α denote the vector of length M of values of α_k, Z denote the T by M matrix of values of $Z_k(\omega_j)$, and U denote the vector of length T of values of $U(\omega_j)$, our model is

$$(3.5) \qquad Y = Z\alpha + U \quad .$$

All of these variables except α have, in general, complex values.

To derive the minimum variance linear unbiased estimate of α, we first premultiply both sides by the matrix, F^{-1}, of the discrete inverse Fourier transform, to get

$$(3.6) \qquad F^{-1}Y = F^{-1}Z\alpha + u \quad .$$

The matrix F^{-1} is given by

$$(3.7) \qquad F^{-1} = \frac{1}{T} \begin{bmatrix} e^{i\omega_{-N}} & e^{i\omega_{-N+1}} & \cdots & 1 & \cdots & e^{i\omega_N} \\ e^{2i\omega_{-N}} & e^{2i\omega_{-N+1}} & \cdots & 1 & \cdots & e^{2i\omega_N} \\ \vdots & \vdots & & \vdots & & \vdots \\ e^{Ti\omega_{-N}} & e^{Ti\omega_{-N+1}} & \cdots & 1 & \cdots & e^{Ti\omega_N} \end{bmatrix} .$$

Under our stochastic specification, the MVLU estimator is

$$\hat{\alpha} = [Z'(F^{-1})'F^{-1}Z]^{-1}Z'(F^{-1})'F^{-1}y \quad . \tag{3.8}$$

It is easy to verify that $Z'(F^{-1})'F^{-1} = \frac{1}{T} Z^*$ where Z^* indicates the transpose of the complex conjugate of the matrix Z. Thus the estimator can be written more conveniently as

$$\hat{\alpha} = (Z^*Z)^{-1}Z^*y \quad . \tag{3.9}$$

If $B(\omega_j)$ is not linear in all of its parameters, the estimation problem is rather more complicated. One approach is to calculate least squares estimates with the aid of a nonlinear minimization algorithm, although relatively little can be said about the statistical properties of these estimates. A very simple algorithm for this purpose is provided by Gauss' method. For a vector of estimates $\alpha^{(k)}$ at the k'th iteration, we calculate a new set of estimates from the iteration equation,

$$\hat{\alpha}^{(k+1)} = (Z^*Z)^{-1}Z^*W \tag{3.10}$$

where Z is the matrix of derivatives of $B(\omega_j)\ X(\omega_j)$ with respect to $\hat{\alpha}^{(k)}$; its j'th row is given by

$$(3.11) \qquad Z(\omega_j) = X(\omega_j) \frac{\partial B(\omega_j)}{\partial \hat{\alpha}^{(k)}} \quad ,$$

and W is the vector Y adjusted for the implicit linear approximation:

$$(3.12) \qquad W(\omega_j) = Y(\omega_j) - B(\omega_j)X(\omega_j) + Z(\omega_j)\hat{\alpha}^{(k)}$$

with $B(\omega_j)$ evaluated at the parameter values $\hat{\alpha}^{(k)}$. Note that if the model is linear in α, the two adjustment terms cancel, and Gauss' method reduces to ordinary least squares.

We conclude this section with a brief discussion of the unscrambling problem in distributed lag estimation. Our goal is to obtain general formulas for the estimates of the distributed lag function, $\hat{\beta}_\tau$, the variance-covariance matrix of these estimates, an estimate of the sum of the lag coefficients and its variance, and an estimate of the mean lag.

Since the inverse Fourier transform is linear, the best estimates of the lag coefficients can be calculated from a best estimate, $B(\omega_j)$, of the Fourier transform:

$$(3.13) \qquad \hat{\beta} = F^{-1} \hat{B}$$

in this formula, F^{-1} is the matrix of the inverse transform and $\hat{B}$ is the vector of values of $\hat{B}(\omega_j)$. If $B(\omega_j)$ has M parameters, $\alpha_1, \ldots, \alpha_M$, and

the variance-covariance matrix of the vector of estimates of these, $\hat{\alpha}$, is denoted $V(\hat{\alpha})$, then an approximate variance-covariance matrix of $\hat{\beta}$ is given by

$$(3.14) \qquad V(\hat{\beta}) = (F^{-1} \frac{\partial \hat{B}}{\partial \hat{\alpha}}) V(\hat{\alpha}) (F^{-1} \frac{\partial \hat{B}}{\partial \hat{\alpha}})' \quad .$$

This formula is exact if the parameters enter linearly.

An estimate of the sum of the lag coefficients is provided by the Fourier transform evaluated at frequency zero:

$$(3.15) \qquad \hat{S} = \hat{B}(0) \quad ;$$

an estimate of its variance is given by

$$(3.16) \qquad V(\hat{S}) = (\frac{\partial \hat{B}(0)}{\partial \hat{\alpha}}) V(\hat{\alpha}) (\frac{\partial \hat{B}(0)}{\partial \hat{\alpha}})' \quad .$$

To get a convenient form for the estimate of the mean lag, we make the following observation:

$$(3.17) \qquad \hat{B}'(0) = -i \sum_{\tau=0}^{T-1} e^{-i\omega_j \tau} \tau \hat{\beta}_\tau \quad .$$

Thus an estimate of the mean lag is given by

$$(3.18) \qquad \hat{\mu} = \frac{\sum_{\tau=0}^{T-1} \tau\hat{\beta}_\tau}{\hat{S}} = \frac{i\hat{B}'(0)}{\hat{B}(0)} \quad .$$

The approximation to the variance of $\hat{\mu}$ based on differentiation of this formula has proven to be too unreliable for general use.

Specification of the Fourier transform $B(\omega_j)$ lies at the heart of the proposed approach to distributed lag estimation and interpretation. Significantly, for every well-known lag specification, $B(\omega_j)$ can be written in a simple expression involving only elementary arithmetic operators and functions. This in turn implies that a computer program endowed with the ability to handle simple functions analytically could estimate and unscramble almost any distributed lag function. The lag specification would be presented to the computer in a simple algebraic statement of its Fourier transform. Since a number of econometric computer systems currently in operation have the capability for analytical operations upon functions, including the most important operation, analytical differentiation, this is an entirely realistic proposal. In fact, it is the potential computational flexibility which provides the main justification for the use of transform methods in distributed lag work.

4. More general stochastic specifications and the consequences of errors in the stochastic specification

Experience indicates that the simple stochastic hypothesis discussed in the previous section is not, in fact, a sufficiently general view of the disturbance process in distributed lag models. Intertemporal dependence seems to characterize the disturbances in almost all time series models. For this reason, we find it necessary to adopt a somewhat more general view of the disturbance process. By analogy with our assumption about the systematic part of the distributed lag model, we will assume that a time-invariant process generates the disturbances:

$$U(\omega_j) = \mathcal{D}(\omega_j)\, V(\omega_j) \quad , \tag{4.1}$$

where $V(\omega_j)$ is the Fourier transform of a vector of disturbances, v, which meets the Gauss-Markoff hypotheses[1]: $E(v) = 0$ and $E(vv') = I$ (the variance parameter, σ^2, can be absorbed into $\mathcal{D}$). We note at this point that the equivalent statement of these hypotheses in terms of the Fourier transform is the following: $E(V) = 0$ and $E(VV^*) = I$, where V is the vector of values of $V(\omega_j)$ for $-N \leq j \leq N$.

The best-known stochastic specification of this type is the first-order autocorrelation model,

---

[1]Here we are implicitly assuming that the disturbance is periodic with period T. Once again, this assumption is made to achieve exact results at a small cost in realism.

$$u_t = \rho u_{t-1} + v_t \quad . \tag{4.2}$$

By taking Fourier transforms of both sides, we get

$$U(\omega_j) = \rho e^{-i\omega_j} U(\omega_j) + V(\omega_j) \tag{4.3}$$

or

$$U(\omega_j) = \frac{1}{1 - \rho e^{-i\omega_j}} V(\omega_j) \quad . \tag{4.4}$$

Thus for the first-order autocorrelation model,

$$\mathcal{D}(\omega_j) = \frac{1}{1 - \rho e^{-i\omega_j}} \quad . \tag{4.5}$$

Under our general stochastic specification, the complete distributed lag model is

$$Y(\omega_j) = \mathcal{B}(\omega_j) X(\omega_j) + \mathcal{D}(\omega_j) V(\omega_j) \tag{4.6}$$

Writing the model in this form suggests that the problems of specifying and estimating the distributed lag function $\mathcal{B}(\omega_j)$ and the disturbance lag function $\mathcal{D}(\omega_j)$ are symmetric and entirely comparable. Further examination shows, however, that this is not the case. For the moment, suppose that

$\mathcal{D}(\omega_j)$ were known. Then we could divide both sides of equation 4.6 by $\mathcal{D}(\omega_j)$ to get a model which satisfies the Gauss-Markoff hypotheses:

$$(4.7) \qquad \frac{Y(\omega_j)}{\mathcal{D}(\omega_j)} = \mathcal{B}(\omega_j) \frac{X(\omega_j)}{\mathcal{D}(\omega_j)} + V(\omega_j) \quad .$$

But there is another transformation which also yields a model satisfying these hypotheses. To show this we first define the real-valued function,

$$(4.8) \qquad S(\omega_j) = \mathcal{D}^*(\omega_j)\mathcal{D}(\omega_j) \quad .$$

Then we divide each side of the equation of the distributed lag model by the square root of this function to get

$$(4.9) \qquad \frac{Y(\omega_j)}{\sqrt{S(\omega_j)}} = \mathcal{B}(\omega_j) \frac{X(\omega_j)}{\sqrt{S(\omega_j)}} + \frac{\mathcal{D}(\omega_j)}{\sqrt{S(\omega_j)}} V(\omega_j) \quad .$$

A typical element of the variance-covariance matrix of the transformed disturbance is

$$(4.10) \qquad E\left[\frac{\mathcal{D}(\omega_j)}{\sqrt{S(\omega_j)}} V(\omega_j) \frac{\mathcal{D}^*(\omega_k)}{\sqrt{S(\omega_k)}} V^*(\omega_k)\right] = \frac{\mathcal{D}(\omega_j)\mathcal{D}^*(\omega_k)}{\sqrt{S(\omega_j)S(\omega_k)}} E[V(\omega_j)V^*(\omega_k)]$$

$$= 0 \qquad \text{if } j \neq k$$

$$= 1 \qquad \text{if } j = k \quad ,$$

as required. That is, knowledge of the real-valued function $S(\omega_j)$ is just as good, from the point of view of efficient estimation, as knowledge of the complex-valued function $\mathcal{D}(\omega_j)$. The function $S(\omega_j)$ is the spectrum of the disturbances, u_t. We have demonstrated the following important proposition: In a model with a stationary (time-invariant) disturbance process, the minimum variance linear unbiased estimates of the distributed lag parameters (assuming they enter linearly) are given by weighted least squares in the frequency domain; the weights are the inverses of the spectrum of the disturbances. In other words, by taking a Fourier transform, the problem of intertemporal dependence of the disturbances can be reduced to the problem of heteroscedasticy.[1]

We will be able to offer a fairly complete discussion of the problem of efficient estimation in the one-parameter linear model,

$$Y(\omega_j) = \alpha Z(\omega_j) + \mathcal{D}(\omega_j)\, V(\omega_j) \quad . \tag{4.11}$$

The rectangular distributed lag function with fixed length, p, provides an example of this case; α then gives the height of the lag function and $Z(\omega_j)$ is given by

$$Z(\omega_j) = \frac{1 - e^{-i\omega_j p}}{i\omega_j} X(\omega_j) \quad . \tag{4.12}$$

---

[1]This property of stationary disturbances was pointed out by Grenander [5] many years ago. Hannan [10] seems to have been the first to propose estimation methods taking advantage of it. The weighted least squares interpretation and an alternative estimation method were offered by Duncan and Jones [2].

Investigation of the extent to which the results for the one-parameter case can be generalized to several parameters is a topic for further research.

Now the ordinary least squares (OLS) estimator of α is

$$\hat{\alpha}_{OLS} = \frac{\sum_{j=-N}^{N} Z^*(\omega_j)\, Y(\omega_j)}{\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)} \quad ; \tag{4.13}$$

its variance is

$$V(\hat{\alpha}_{OLS}) = \frac{\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j) S(\omega_j)}{\left[\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\right]^2} \quad . \tag{4.14}$$

On the other hand, the efficiently weighted least squares (EWLS) estimator is

$$\hat{\alpha}_{EWLS} = \frac{\sum_{j=-N}^{N} \frac{Z^*(\omega_j)\, Y(\omega_j)}{S(\omega_j)}}{\sum_{j=-N}^{N} \frac{Z^*(\omega_j)\, Z(\omega_j)}{S(\omega_j)}} \tag{4.15}$$

and its variance is

$$(4.16) \qquad V(\hat{\alpha}_{EWLS}) = \frac{1}{\sum_{j=-N}^{N} \frac{Z^*(\omega_j)\, Z(\omega_j)}{S(\omega_j)}} \quad .$$

The emphasis of our discussion will be on the problem of obtaining efficient estimates of α, but it should be mentioned that the problem of estimating $\mathcal{D}(\omega_j)$ for its own sake is important in, for example, short-run prediction.

Our first question is: How serious an error do we commit by using the OLS estimator instead of the EWLS estimator?[1] For this purpose we define the efficiency function,

$$(4.17) \qquad \Phi(Z,S) = \frac{V(\hat{\alpha}_{EWLS})}{V(\hat{\alpha}_{OLS})} = \frac{\left[\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\right]^2}{\left[\sum_{j=-N}^{N} \frac{Z^*(\omega_j)\, Z(\omega_j)}{S(\omega_j)}\right]\left[\sum_{j=-N}^{N} S(\omega_j)\, Z^*(\omega_j)\, Z(\omega_j)\right]} \quad .$$

As a first step in answering our question, we state a theorem which provides a lower bound on the efficiency of ordinary least squares:

Theorem: For a given spectrum, $S(\omega_j)$,

$$(4.18) \qquad \Phi(Z,S) \geq \frac{4}{2 + \frac{S_{max}}{S_{min}} + \frac{S_{min}}{S_{max}}}$$

where S_{max} is the largest value of $S(\omega_j)$ and S_{min} is the smallest value of $S(\omega_j)$.

---

[1]This question was first investigated by Grenander [6]. Our discussion parallels Malinvaud ([12], pp. 437-439) in some respects. We also use a result of Watson [15] and Hannan [11] to provide a lower bound on the efficiency of OLS.

The proof is presented in Hannan ([11], p. 111].[1] We note from that proof that the lower bound is attained in the "worst case,"

$$(4.19) \qquad Z(\omega_j) = 1 \quad \text{if } \omega_j = \omega_{max} \text{ or } \omega_j = \omega_{min}$$
$$= 0 \quad \text{otherwise,}$$

where $S(\omega_{max}) = S_{max}$ and $S(\omega_{min}) = S_{min}$. In the typical case, the disturbance has its highest power at zero frequency and least power at the highest frequency, so $\omega_{max} = 0$ and $\omega_{min} = \Pi \frac{T - 1}{T}$. Then in the time domain, a worst-case right-hand variable is

$$(4.20) \qquad Z_t = 1 + \cos(\Pi \frac{T - 1}{T} t) \quad .$$

We will also be interested in evaluating the efficiency of OLS for a more typical case. The efficiency can be seen to depend only on the spectrum $Z^*(\omega_j)\, Z(\omega_j)$ of the right-hand variable and on the spectrum of the disturbances. A great many economic time series have roughly the same spectrum, as C. W. J. Granger's very careful investigation has shown. The "typical spectral shape" found by Granger is shown in Figure 2; it can be represented mathematically as

$$(4.21) \qquad Z^*(\omega_j)\, Z(\omega_j) = \frac{1}{1 - 2\delta \cos \omega_j + \delta^2} \quad .$$

---

[1]Hannan indicates that the same bound is approximately correct in the case of several right-hand variables.

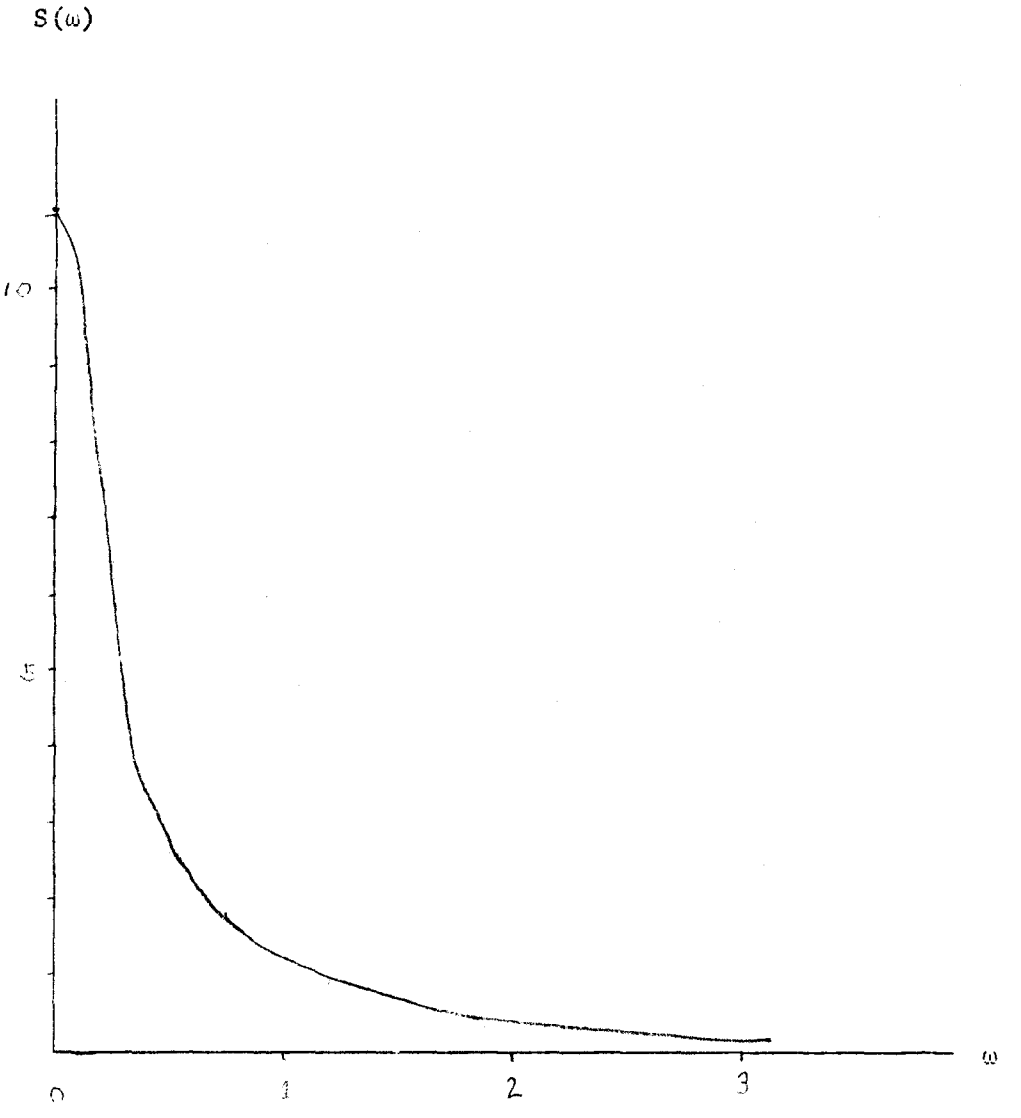

Figure 2. Spectrum of a typical economic variable, with autocorrelation δ equal to 0.7.

From formula 4.5 we can see that this is the same as the spectrum of a first-order autoregressive process with autocorrelation δ. The tendency of the right-hand variable in a distributed lag regression to have roughly this shape is particularly strong since the variable $Z(\omega_j)$ is derived from the transform of a time series, $X(\omega_j)$ by multiplying by a function which attenuates higher frequencies. This function is $\frac{1 - e^{-i\omega_j p}}{i\omega_j}$ in the case of the rectangular distributed lag. For this reason, we will be particularly concerned with the efficiency of ordinary least squares when the right-hand variable has this spectral shape with parameter δ close to one.

The lower bound and some typical values for the efficiency of ordinary least squares when applied in the presence of positive first-order autocorrelation are presented in Table 1. The lower bound is calculated by substituting the minimum and maximum values of the spectrum,

$$(4.22) \qquad U^*(\omega_j)U(\omega_j) = S(\omega_j) = \frac{1}{1 - 2\rho \cos \omega_j + \rho^2}$$

into formula 4.18; the minimum occurs at the highest frequency, $\omega_N = \pi\frac{T - 1}{T}$:

$$(4.23) \qquad S_{min} = \frac{1}{1 - 2\rho \cos \frac{T - 1}{T}\pi + \rho^2}$$

and the maximum occurs at frequency zero:

$$(4.24) \qquad S_{min} = \frac{1}{1 - 2\rho + \rho^2}$$

$$= \frac{1}{(1 - \rho)^2} \quad .$$

Table 1

Efficiency of ordinary least squares for various values of the autocorrelation of the disturbances (ρ) and of the right-hand variable (δ). The top line gives the lower bound for the efficiency over all possible spectra for the right-hand variable. The number of observations, T, is taken to be 75, but the results are not at all sensitive to its value.

	$\rho = 0$	$\rho = .3$	$\rho = .5$	$\rho = .7$	$\rho = .9$
Lower bound	1	.698	.361	.116	.011
$\delta = 0$	1	.835	.600	.342	.105
$\delta = .5$	1	.851	.600	.311	.079
$\delta = .8$	1	.914	.714	.389	.083
$\delta = .9$	1	.951	.813	.503	.104
$\delta = .95$	1	.974	.894	.649	.149
$\delta = .99$	1	.998	.991	.956	.625

The results shown in Table 1 suggest that the practical importance of autocorrelation in distributed lag models has been somewhat exaggerated. In the leading case, $\rho = .5$ and $\delta = .95$, the potential gain in efficiency from an autoregressive transformation of the variables (that is, weighted least squares in the frequency domain) is about 10%. In actual empirical work, even this small gain is unattainable, since the parameter ρ is unknown and must be estimated from a first stage regression.

The following explanation for this observation may be offered: The substantial potential gain in efficiency from an autoregressive transformation, as measured by the lower bound, is achieved by giving large weights to higher frequencies and small weights to lower frequencies. But if the right-hand variable has significant power only at low frequencies, as is almost inevitably the case in distributed lag estimation, then the weights applied to high frequencies are essentially irrelevant. The advantage of weighted least squares can be obtained only with a right-hand variable with power at both high and low frequencies.[1]

This observation may explain the puzzling finding of several authors[2] that in the application of a two-stage method for correcting for autocorrelation (such as Durbin's method, [3]), the results from the second stage, after an autoregressive transformation, are almost identical to the ones from the first stage.

This conclusion has a limited application. Table 1 also shows that if the right-hand variable does have significant power at high frequencies

---

[1]This point has been made by Grenander [6] for the case of polynomial or trigonometric right-hand variables.

[2]Including the present author, in [7].

(for example, if it looks like white noise, with $\delta = 0$), then the gain in efficiency from weighted least squares may be substantial. Fortunately, the spectrum, $Z^*(\omega_j)\ Z(\omega_j)$ can be calculated in advance, to find to what extent the investigator need be concerned about the stochastic specification. Perhaps it should be mentioned that it is precisely in cases where z_t has power at high frequencies and the stochastic specification is important that it is realistic to expect to estimate the fine details of the lag structure, since these are essentially high-frequency attributes. If the right-hand variable has the typical spectral shape, only the broad outline of the lag function can be discerned.

With this warning, we turn to a second important question about the stochastic specification: What is the potential loss in estimation efficiency in using a weighted least squares estimator derived from a crude approximation to the true stochastic specification instead of the efficiently weighted estimator? This question gains relevance in view of the proposals of Phillips [14] and others to estimates relatively complicated disturbance processes in econometric time series models. If in fact a crude approximation does almost as well, the effort devoted to estimating a complicated disturbance process may be wasted insofar as it is directed toward the goal of increasing the efficiency of the coefficient estimates. Our method in investigating this question will be to calculate the efficiency of arbitrarily weighted least squares (AWLS) relative to the EWLS estimator for the true disturbance process.

For an arbitrary set of weights, $W(\omega_j)$, the AWLS estimator is

$$(4.25) \qquad \hat{\alpha}_{AWLS} = \frac{\sum_{j=-N}^{N} Z^*(\omega_j)\, Y(\omega_j)\, W(\omega_j)}{\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\, W(\omega_j)}$$

Its variance is

$$(4.26) \qquad V(\hat{\alpha}_{AWLS}) = \frac{\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\, S(\omega_j)[W(\omega_j)]^2}{\left[\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\, W(\omega_j)\right]^2} .$$

The efficiency of the AWLS estimator is

$$(4.27) \qquad \Phi = \frac{\left[\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\, W(\omega_j)\right]^2}{\left[\sum_{j=-N}^{N} \frac{Z^*(\omega_j)\, Z(\omega_j)}{S(\omega_j)}\right]\left[\sum_{j=-N}^{N} Z^*(\omega_j)\, Z(\omega_j)\, S(\omega_j)[W(\omega_j)]^2\right]} .$$

Once again, a lower bound on the efficiency of the estimator can be obtained; it is

$$(4.28) \qquad \Phi \geq \frac{4}{2 + \frac{\bar{S}_{max}}{\bar{S}_{min}} + \frac{\bar{S}_{min}}{\bar{S}_{max}}} ,$$

where $\bar{S}_{max} = \max_j S(\omega_j)\, W(\omega_j)$ and $\bar{S}_{min} = \min_j S(\omega_j)\, W(\omega_j)$.

This follows from the substitution

$$\overline{Z}(\omega_j) = Z(\omega_j)\sqrt{W(\omega_j)} \tag{4.29}$$

and

$$\overline{S}(\omega_j) = S(\omega_j)\,W(\omega_j) \tag{4.30}$$

and the direct application of the earlier theorem to this transformed model.

The example which we will consider is a step function approximation to the first-order autoregressive disturbance process. The step function has two steps, and is equal to the reciprocal of the value of the spectrum of the autoregressive disturbance at the midpoint of each step. The first step covers the lower quarter of the frequencies and the second step covers the remaining upper three quarters:

$$W(\omega_j) = 1 - 2\rho\cos\frac{\pi}{8} + \rho^2 \quad , \quad -\frac{\pi}{4} \leq \omega_j \leq \frac{\pi}{4} \tag{4.31}$$

$$= 1 - 2\rho\cos\frac{5\pi}{8} + \rho^2 \quad , \quad \frac{\pi}{4} < |\omega_j| \leq \pi \quad .$$

Values of the efficiency of this two-step weighted least squares estimator are given in Table 2. The improvement over the ordinary least squares estimator is quite dramatic, especially in the case of moderately severe autocorrelation of the disturbances. When ρ is 0.7, the new estimator

Table 2

Efficiency of the two-step weighted least squares estimator.

	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
Lower bound	1	.824	.593	.369	.007
$\delta = 0$	1	.952	.897	.852	.809
$\delta = 0.5$	1	.955	.881	.804	.629
$\delta = 0.8$	1	.976	.914	.796	.452
$\delta = 0.9$	1	.987	.944	.812	.327
$\delta = 0.95$	1	.993	.968	.853	.288
$\delta = 0.99$	1	.999	.997	.980	.663

is always at least 79% efficient (if the right-hand variable has a spectrum of the autocorrelation shape), whereas the efficiency of ordinary least squares can drop to 31%. The efficiency of the new estimator is low for the most serious case of autocorrelation, $\rho = 0.9$, but econometric models with this much autocorrelation rarely deserve serious consideration.

This example suggests that effort devoted to careful calculation of weights for least squares estimation of distributed lag models would be misplaced. The two-step approximation to the disturbance spectrum will usually give estimates which are almost indistinguishable from the efficient estimates, not only in the case of autocorrelation, but for any spectrum which declines (or rises) smoothly from low to high frequencies. This includes, for example, the second-order autoregressive process and some moving average processes. The approximation would perform very poorly if the disturbance spectrum had a very low value for a group of intermediate frequencies at which the right-hand variables had significant power. Then the efficient estimates would be dominated by these observations, and would have a correspondingly low variance, whereas the two-step approximation would be unable to approximate the true weights, and would have a much larger variance. But this does not appear to be a likely occurence in distributed lag analysis.

In practical econometric work, of course, the disturbance spectrum is not known in advance, and the only estimator among those just discussed which can actually be used is the ordinary least squares estimator. Many two-stage estimation methods have been proposed for taking advantage of the increased efficiency which it is thought can be obtained by using weighted least squares

in the frequency domain or an equivalent transformation in the time domain.[1] The first stage in most of these methods involves the application of ordinary least squares to the original model, and the subsequent estimation of the disturbance spectrum (often in the form of estimating the auto-correlation coefficient) from the residuals of this regression. In the second stage, weighted least squares estimates are obtained, using as weights the reciprocals of the estimated points of the spectrum. It is well known that if the first stage gives a consistent estimate of the spectrum, the estimates of the regression coefficients from the second stage are asymptotically efficient [see Malinvaud ([12], p. 440)].

One interpretation we might offer of our observation that a crude approximation to the spectrum provides weights that are almost as good as the true weights is that two-stage estimation methods which involve the estimation of the fine details of the spectrum in the first stage are not likely to be significantly more efficient than a method in which a very rough estimate of the spectrum is made in the first stage. One such method is suggested by the two-step approximation to the spectrum mentioned above. The first stage in this method to the calculation of separate estimates for low frequencies and for high frequencies; in the simple case of one linear coefficient, these are:

$$\hat{\alpha}_1 = \frac{Z_1^* y_1}{Z_1^* Z_1} \tag{4.32}$$

---

[1] See Malinvaud ([12], chapter 13) for a review of time domain methods for handling autocorrelation in econometric models. Hannan [10] and [9] and Duncan and Jones [2] have developed more general non-parametric frequency domain methods.

and

$$\hat{\alpha}_2 = \frac{Z_2^* y_2}{Z_2^* Z_2} \quad . \tag{4.33}$$

In these formulas, Z_1 and y_1 are the vectors of observations on the variables for the $2\theta + 1$ lower frequencies and Z_2 and y_2 are the corresponding vectors for the $T - 2\theta - 1$ higher frequencies. As in the previous case, we might take $\theta = N/4$.

Estimates of the two steps in the spectrum are provided by the estimates of the residual variances in these two regressions:

$$S_1 = \frac{1}{2\theta} (Z - \hat{\alpha}_1 y)^*(Z - \hat{\alpha}_1 y) \tag{4.34}$$

and

$$S_2 = \frac{1}{T - 2\theta - 2} (Z - \hat{\alpha}_2 y)^*(Z - \hat{\alpha}_2 y) \quad . \tag{4.35}$$

We note that these estimates are unbiased if the spectrum does, in fact, have two steps.

Then the second stage estimate of α is

$$\hat{\alpha} = \frac{\frac{Z_1^* y_1}{S_1} + \frac{Z_2^* y_2}{S_2}}{\frac{Z_1^* Z_1}{S_1} + \frac{Z_2^* Z_2}{S_2}} \quad . \tag{4.36}$$

The efficiency of this estimator is unlikely to be much less than that of the more complicated alternatives in the literature, and may easily exceed the efficiency of some. One incidental advantage of this method is that if the disturbances are normally distributed, the statistic S_1/S_2 has the F-distribution with 2θ and $T - 2\theta - 2$ degrees of freedom, on the null hypothesis of equal values of the disturbance spectrum, or no autocorrelation. Thus it provides an alternative to the Durbin-Watson statistic in testing for departures from the simple stochastic hypothesis of intertemporal independence. In contrast to the Durbin-Watson statistic, its sampling distribution is known exactly. We should note, however, that just as in the case of the Durbin-Watson test, failure to reject the null hypothesis of intertemporal independence does <u>not</u> imply that there is no dependence or that there is no gain in efficiency from applying the second stage of the estimator.

References

[1] Shirley Almon, "The Distributed Lag Between Capital Appropriations and Expenditures," Econometrica, Vol. 33, January 1965, pp. 178-196.

[2] David B. Duncan and Richard H. Jones, "Multiple Regression with Stationary Errors," Journal of the American Statistical Association, Vol. 61, December 1966, pp. 917-928.

[3] J. Durbin, "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, ser. B, Vol. 22, 1960, 139-153.

[4] C. W. J. Granger, "The Typical Spectral Shape of an Economic Variable," Econometrica, Vol. 34, January 1966, pp. 150-161.

[5] Ulf Grenander, "On Empirical Spectral Analysis of Stochastic Processes," Arkiv for Matematik, Vol. 1, 1951, pp. 503-531.

[6] _____, "On Estimation of the Regression Coefficients in the Case of an Autocorrelated Disturbance," Annals of Mathematics and Statistics, Vol. 25, 1954, pp. 252-272.

[7] Robert E. Hall and Dale W. Jorgenson, "Estimating the Effects of Recent Investment Stabilization Policies," forthcoming in a Brookings Conference volume on the effects of tax incentives on investment.

[8] Robert E. Hall and Richard C. Sutch, "A Flexible Infinite Distributed Lag," Working Paper No. 124, Institute of Business and Economic Research, University of California, Berkeley, February 1968.

[9] E. J. Hannan, "The Estimation of Relationships Involving Distributed Lags," Econometrica, Vol. 33, January 1965, pp. 206-224.

[10] E. J. Hannan, "Regression for Time Series," in Symposium on Time Series Analysis, M. Rosenblatt (ed.), New York, John Wiley & Sons, 1963, Chapter 2.

[11] _____, Time Series Analysis, New York, John Wiley & Sons, 1960.

[12] Edmond Malinvaud, Statistical Methods of Econometrics, Chicago, Rand McNally, 1966.

[13] Emanuel Parzen, Empirical Time Series Analysis (forthcoming).

[14] A. W. Phillips, "Estimation of Systems of Difference Equations with Moving Average Disturbances," Paper read at the December 1966 meetings of the Econometric Society, San Francisco.

[15] G. S. Watson, "Serial Correlation in Regression Analysis I," Biometrika, Vol. 42, 1955, pp. 327-341.