Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: http://www.nber.org/books/kmen80-1

Publication Date: 1980

Chapter Title: Multicollinearity and the Estimation of Low-Order Moments in Stable Lag Distributions

Chapter Author: Michio Hatanaka, T. Dudley Wallace

Chapter URL: http://www.nber.org/chapters/c11708

Chapter pages in book: (p. 323 - 337)

Multicollinearity and the Estimation of Low-Order Moments in Stable Lag Distributions

MICHIO HATANAKA

INSTITUTE OF SOCIAL AND ECONOMIC RESEARCH
OSAKA UNIVERSITY
SUITA OSAKA, JAPAN

and

T. DUDLEY WALLACE

DEPARTMENT OF ECONOMICS
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA

1. Introduction

The purpose of this paper is to present arguments that, for typical economic time series data, low-order moments of lag distributions in many cases can be estimated more precisely than short-run effects. Moreover, the transformation from short-run effects to moments, again for typical time series data in economics, orders the precision inherent in the design matrix so that one can expect to estimate the long-run effect (sum of ordinates in the lag distribution) with greatest precision, the first-order moment (mean lag) with next greatest precision, etc. Thus we argue that much can be learned with available data by redirecting attention away from short-run effects to a different parameter space. We believe that there are three reasons for this redirection of interest.

1. Low-order moments are of considerable policy interest. Often the long-run effect and mean lag are sufficient summary statistics for the investigator's purpose. For example, the total return from a marginal dollar spent on advertising along with some notion of the timing of the returns

323

flow may be sufficient for decision purposes. For another, the mean lags for fiscal versus monetary policy are of direct interest.

2. With the focus on estimation of short-run effects, priors of one kind or another on the shape of the lag distribution typically have been imposed on data, usually in *ad hoc* ways. Many quite flexible probability density functions require knowledge of only a few low-order moments for complete specification. If low-order moments can be obtained with precision in a first round pass, restrictions can be imposed in a second round which are not at variance with the data.

3. In cases where smoothness priors have been imposed *a priori*, least squares estimation of low-order moments of lag distributions serves as an additional check on specification. As is well known the mean lag, for example, can be quite sensitive to the form of the prior.

The mean lag usually can be estimated with considerable precision without priors, as we argue in this paper. Therefore, reestimation of a few low-order moments free of priors may reveal gross inadequacies of those priors.

The weaknesses of the procedure we outline include the following:

1. Except for the sum of coefficients (the "zero-order moment"), transformation from ordinates to moments makes sense only if the lag weights are of the same sign. Feedback that causes "overshooting" in a delay process destroys the useful analogy of a lag distribution to a probability density function. Our procedure works only for lagged adjustment processes where all short-run effects can be expected to have the same sign.

2. The procedure depends on the assumption that the number of nonzero lag coefficients is considerably smaller than the number of time series observations which is available. This leaves the problem of how many lags to include.

3. We assume exogeneity. The response variable depends only on contemporaneous and lagged values of an independent variable. This assumption can, of course, be tested in applications.

4. Even in those cases where the design matrix loads precision onto the low-order moments, the precision inherent in the data still may not be very satisfactory. An example application is given in Table 4 where the long-run effect can be estimated quite precisely (large t value) but the mean and higher-order moments are not estimated with as great precision as one might wish.[1]

In summary, the proposed procedure is to transform from ordinates to moments of a lag distribution. Since the ordinates of a lag distribution do

---

[1] Such is not always the case. In an investigation of the lag between consumer and wholesale prices, even the third-order moment (skewness), was estimated with a t value of 4.0 with monotonically larger t values for variance, mean, and sum of ordinates (Silver & Wallace, 1980).

not sum to one, the sum of the ordinates is called the zero-order moment.[2] The transformation from ordinates to moments is nonsingular so no precision is gained in a total sense. However, the precision inherent in typical economic design matrices is loaded onto the low-order moments.

Our theoretical arguments use modern time series concepts in order to exploit the theorem about approximate inverses of autocovariance matrices of general stationary stochastic processes. However, the general notion is that positive linear combinations of negatively correlated random variables can have lower variances than the original variables. In a distributed lag problem, the variance-covariance matrix of the estimated short-run coefficients is approximately proportional to the inverse of the autocorrelation matrix of the process generating the regressor. Most economic time series are positively autocorrelated; therefore, the inverse is dominated by negative correlations—the estimated regression coefficients are by and large negatively correlated with each other. Therefore, their sum has a variance smaller than the sum of their variances. Similarly, the mean lag is proportional to the weighted sum of the estimated short-run effects with the positive integers as weights, so that the negative correlations are again exploited. Some care has to be taken in setting up appropriate criteria because the sums of weights in the transformations are not unity, and the relevant question is whether or not (theoretical) t statistics are larger for the estimated moments than for estimated short-run effects.[3]

Turning more formally to the proposed method, consider the model

$$Y_t = \alpha + \sum_{s=0}^{N} \beta_s X_{t-s} + U_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $U_t \sim \text{iid}(0, \sigma^2)$ and independent of X_{t-s}, $s = 0, 1, \ldots, N$. We assume that the sample size is large enough so that for $N < T$ any lags beyond N can be neglected. We assume further that all the βs are of the same sign and also, until a later section, we assume that the X and Y processes are covariance-stationary.

Converting the estimation problem from short-run lag effects to moments of the lag distribution motivates the following transformation of the βs:

$$\begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & N \\ 0 & 1^2 & 2^2 & \cdots & N^2 \\ \vdots & & & & \\ 0 & 1^N & 2^N & \cdots & N^N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} \qquad \text{or} \qquad \mu = C\beta. \tag{2}$$

[2] In cases for which the implied conceptual experiment is valid, the zero-order moment can be interpreted as the long-run effect of a unit change in the control variable. Where the lag distribution arises due to expectations, a once and for all change in the control variable changes the structure by which the data were generated and therefore is an invalid concept.

[3] A theoretical t statistic is the square root of the inverse of coefficient of variation.

Nonsingularity of the transformation to the μs is assured since the first principal submatrix of C is a Vandermonde matrix.

Note that the μs are unadjusted moments in a nonnormalized lag distribution. That is, μ_0 is the zero-order moment, μ_1 is the mean of the nonnormalized lag distribution, μ_2 is the second moment about zero in the nonnormalized lag distribution, etc. What we mean by nonnormalized is that the βs do not sum to unity. To convert to a proper probability density function the β ordinates must be divided by μ_0, the sum of βs. Therefore, the moments of final interest require the transformations

$$M_0 = \mu_0, \qquad M_1 = \mu_1/\mu_0, \qquad M_2 = \mu_2/\mu_0 - M_1^2, \qquad \text{etc.,} \qquad (3)$$

where M_0 is the long-run effect of a change in X_t on Y_t, M_1 is the mean of the lag distribution, M_2 is its variance, etc.

In the sections to follow we focus first on estimation of the μs and for that purpose, define C_h' as the hth row of the matrix C in Eq. (2).

2. Objectives and Precision Criteria

The moments of the lag distribution, $\mu_0, \mu_1, \ldots$, are various linear combinations of the β vector. Let $\mu_h = C_h'\beta$, where C_h' is the hth row of the transformation matrix C of Eq. (2). Least squares estimators for μ_h are

$$\hat{\mu}_h = C_h'b \qquad \text{and} \qquad b = (X'X)^{-1}X'Y, \qquad h = 0, 1, \ldots, N, \qquad (4)$$

where the elements of X and Y are measured in terms of deviations from the respective sample means. The dimension of X is $T \times (N + 1)$ and Y is $T \times 1$; therefore, b, the least squares estimator for β, is $(N + 1) \times 1$, and $\hat{\mu}_h$ is the scalar estimate of μ_h.

We claim that (i) some of the μ_h can be estimated more precisely than β for typical economic data in a distributed lag context and (ii) the precision is ordered from low-order to higher-ordered moments; $\hat{\mu}_0$ is more precise than $\hat{\mu}_1$, $\hat{\mu}_1$ is more precise than $\hat{\mu}_2$, etc. Of course, the latter claim implies that those μ_h which can be estimated more precisely than the elements of β are the lower-order moments—those for which h is small. Our claims are based on the nature of the typical design matrix in economics for distributed lag problems—that autocorrelation in the regressor can be exploited by concentrating on the low-order moments rather than on the short-run effects represented by the βs.

To substantiate the claims we first define a criterion of relative precision. The index of precision for $\hat{\mu}_h$ is its squared coefficient of variation, $\text{SCV}(\hat{\mu}_h)$, the square of the inverse of the theoretical t statistic. From standard regres-

sion theory,

$$\text{SCV}(\hat{\mu}_h) = \sigma^2 \frac{C_h'(X'X)^{-1}C_h}{[C_h'\beta]^2}. \tag{5}$$

An analogous scalar reflecting the dispersion of b from β is the mean squared error of b divided by the length of the β vector. That is, define

$$\text{SCV}(b) = \frac{E(b - \beta)'(b - \beta)}{\beta'\beta} = \sigma^2 \frac{\text{tr}(X'X)^{-1}}{\beta'\beta}. \tag{6}$$

If (6) were large relative to (5), we would claim that μ_h is estimated more precisely than β because of the particular design matrix $X'X$, i.e., because of the properties of the exogenous variable. To pin down what is meant by "large" in this context, consider the case in which $X'X$ is a scalar k times an identity matrix. Since $X'X$ is proportional to an autocovariance matrix for large samples, zero multicollinearity is approximated by this case, and for no multicollinearity, the ratio of (5) to (6) is a function $\delta_1(h)$:

$$\delta_1(h) = \frac{C_h'C_h\beta'\beta}{(N + 1)[C_h'\beta]^2}. \tag{7}$$

Therefore the relative precision criterion that $\hat{\mu}_h$ is better than b is for the ratio of (5) to (6) to be smaller than the standard in (7), which requires a scalar function that we call $\delta_2(h)$ to be less than $(N + 1)^{-1}$. That is, the criterion is for

$$\delta_2(h) = \frac{C_h'(X'X)^{-1}C_h}{C_h'C_h \text{tr}(X'X)^{-1}} < \frac{1}{N + 1}. \tag{8}$$

The inequality in (8) is examined in Section 4 for typical economic data as the procedure for establishing the claim that μ_h for small h can be estimated more precisely than β. The ordering of precision from lower-to higher-order moments requires consideration of not only that $\delta_2(0) < \delta_2(1) < \delta_2(2)$, etc., but that the function $\delta_1(h)$ is similarly ordered, and $\delta_1(h)$ depends upon the nature of the lag distribution. All we can argue here is that the numerator of (7) increases quite rapidly with h, and for typically right skewed lag distributions, the declining β_j offset the increasing C_{jh} in the denominator, so for a large class of lag distribution one would expect $\delta_1(h)$ to be an increasing function of h. Therefore, we work henceforth with the criterion in (8) to establish both claims. Note, however, that only the ordering conjecture depends on $\delta_1(h)$. The conjecture that $\hat{\mu}_h$ can be estimated more precisely for some h than the β vector for typical design matrices in economics depends only on establishing the inequality in (8).

3. A Numerical Example Using $\delta_2(h)$ Criterion

To illustrate the criterion, consider quarterly U.S. GNP in current prices, seasonally adjusted, with $T = 60$.[4] We designate the raw data by X_{1t}; set $N = 19$, and construct $X'X$. Let V_1 be the tenth diagonal element of $X'X$ and $V_1^{-1/2}X_{1t}$ be the first of three data sets among which some comparisons are made. Derive $X_{2t} = X_{1t} - \exp\{\hat{\alpha} + \hat{\beta}t\}$, where $\hat{\alpha}$ and $\hat{\beta}$ are obtained by regressing $\log X_{1t}$ upon t. Again, $T = 60$ and $N = 19$, and the second data set examined is $V_2^{-1/2} X_{2t}$, where V_2 is obtained in the same way as V_1. Third, let $X_{3t} = X_{2t} - \hat{a} - \hat{b}X_{2,t-1}$, where $\hat{a}$, $\hat{b}$ are obtained by regressing X_{2t} upon $X_{2,t-1}$. In this case $T = 59$ and $N = 19$, and $V_3^{-1/2}X_{3t}$ is the third data set.

Strength of autocorrelation is successively reduced, moving from the first through the third sets of data. In the first set the estimated autocorrelation function declines slowly and steadily from 1.0 to .63, up to $N = 19$, while the third series has autocorrelations of .25, $-.18$, .16, $-.20$, $-.12$, $-.12$, 0, $-.18$, for lags 1, 2, ...

Row 1 of Table 1 confirms the comment that multicollinearity is reduced in the successive data sets. The standard for $\delta_2(h)$ mentioned earlier is $10^{-1} \times .5$, and the last three rows show that least squares estimators for μ_0, μ_1, and μ_2 would compare favorably with the standard, especially for the first two data series.[5]

TABLE 1

DATA SETS

	$V_1^{-1/2}X_{1t}$	$V_2^{-1/2}X_{2t}$	$V_3^{-1/2}X_{3t}$
$\mathrm{tr}(X'X)^{-1}$	$10^5 \times .81$	$10^3 \times .49$	$10^2 \times .72$
$\delta_2(0)$	$10^{-4} \times .18$	$10^{-2} \times .21$	$10^{-1} \times .84$
$\delta_2(1)$	$10^{-3} \times .45$	$10^{-2} \times .21$	$10^{-1} \times .98$
$\delta_2(2)$	$10^{-3} \times .87$	$10^{-2} \times .23$	$10^{-1} \times .95$

4. Theoretical Arguments Supporting Claims of Precision

To furnish theoretical support for the claims that low-order moments in lag distributions can be estimated more precisely than short-run effects for typical economic data we employ some concepts from modern time series

---

[4] The time period for the data is from the fourth quarter of 1960 through the third quarter of 1975.

[5] Scaling by $V_i^{-1/2}$ does not affect $\delta_2(h)$.

analysis later in this section. For the reader not familiar with those concepts we start this section with two examples that can be worked out using time domain approaches. For the first example we assume that the stochastic process generating the exogenous variable X_t is a first order autoregressive scheme and that the sample is large. In this case $X'X$, the critical matrix in the $\delta_2(h)$ function of Eq. (8), is proportional to the well known autocorrelation matrix for a first-order autoregressive process. We further restrict the example by assuming that there are only three short-run effects. β_0 is the contemporaneous effect of a unit change in X_t on the dependent variable, β_1 is the first lag effect, and β_2 is the second lag effect. All higher-order lags are assumed to have zero coefficients so that the basic underlying regression is Y_t regressed on X_t, X_{t-1}, and X_{t-2}.

Therefore, for a large sample

$$(X'X)^{-1} \propto \begin{bmatrix} 1 & -\rho & 0 \\ -\rho & 1 + \rho^2 & -\rho \\ 0 & -\rho & 1 \end{bmatrix}, \tag{9}$$

where $\propto$ indicates "proportional to" and ρ is the autocorrelation coefficient for the X_t process ($|\rho| < 1$).

Applying the $\delta_2(h)$ formula in (8), we have

$$\delta_2(0) = \frac{3 + \rho^2 - 4\rho}{3(3 + \rho^2)}, \quad \delta_2(1) = \frac{5 + \rho^2 - 4\rho}{5(3 + \rho^2)}, \quad \delta_2(2) = \frac{17 + \rho^2 - 4\rho}{17(3 + \rho^2)}. \tag{10}$$

Table 2 gives values for the $\delta_2(h)$ criterion for the example for $h = 0, 1, 2$ and various values of ρ. Note in Table 2 that for positive autocorrelation

TABLE 2

$\delta_2(h)$ Criterion Values
for Example 1[a]
and Various Values of ρ

		h	
ρ	0	1	2
$-.75$	.61	.48	.34
$-.50$	.54	.45	.35
$-.25$	.44	.40	.35
0	.33	.33	.33
.25	.22	.27	.31
.50	.13	.20	.28
.75	.05	.14	.24

[a] $h = 0, 1, 2$.

in the independent variable, the $\delta_2(h)$ values are less than .33 for all three moments and are monotonically smaller for larger positive values of ρ, thus substantiating the first claim made in Section 2. Note also that for positive ρ the zero-order moment is estimated with more precision than the first-order moment, etc., consistent with our second claim. Thus for positive ρ, the parameters $\mu_0 = \beta_0 + \beta_1 + \beta_2$, $\mu_1 = \beta_1 + 2\beta_2$, and $\mu_2 = \beta_1 + 4\beta_2$ can each be estimated more precisely than the β vector according to the inequality criterion (8); moreover, the precision is monotonically decreasing with h.[6]

For a second example assume that $X'X$ is proportional to

$$R = (1 - \rho)I + \rho J, \tag{11}$$

where J is an $N + 1$ order matrix of ones and $0 < \rho < 1$. The inverse of R is

$$aI + bJ = R^{-1}, \tag{12}$$

where $a = 1/(1 - \rho)$ and $b = -\rho/(1 - \rho)(1 + N\rho)$. For this example,

$$\delta_2(0) = \frac{1}{N + 1}\left[\frac{1 - \rho}{1 - \rho + N\rho}\right]. \tag{13}$$

Therefore, $\delta_2(0)$ is smaller than the standard for zero multicollinearity. That is autocorrelation in this particular process works in the direction of improving efficiency of the least squares estimate of the zero-order moment μ_0.

One can also show that for the preceding example

$$\delta_2(1) \cong \frac{1}{N + 1}\frac{1}{1 + N\rho - \rho}\left[1 + \frac{1}{4}N\rho\right]. \tag{14}$$

For large N the result in (14) indicates that μ_1 may be estimated about four times more precisely than individual β. Also note that

$$\delta_2(0)/\delta_2(1) \cong (1 - \rho)/(1 + \tfrac{1}{4}N\rho), \tag{15}$$

which can be of order of magnitude less than one, again substantiating in this case the ordering of the precision from lowest-to next lowest-order moment.

Similarily to the approximation given in (14) above, we derive

$$\delta_2(2) \cong \frac{1}{N + 1}\frac{1}{1 + N\rho - \rho}\left[1 + \frac{4}{9}N\rho\right]. \tag{16}$$

---

[6] The joint confidence interval in the μ_h parameter space translates one to one into a joint confidence interval in the β_j parameter space so that nothing is gained in a total sense by the transformation. The point was made earlier that the transformation is nonsingular, and therefore no precision is gained in a total sense. Our criterion is in terms of widths of marginal confidence intervals on the μ_hs and does not account for correlation among the μ_h estimators. This explains why, in the example, all three linear combinations of βs meet the standard in inequality (8).

Thus for μ_2 for the example, one can expect to beat the standard of zero multicollinearity by an efficiency factor of slightly larger than two. Also, the efficiency ordering conjecture is again substantiated by the results in (14) and (16).

A more general explanation for high efficiency of ordinary least squares (OLS) estimators of low-order moments of lag distributions can be given for covariance-stationary processes on the causal variable, using some concepts from modern time series analysis. In the discussion to follow we assume that the spectrum of the $\{X_t\}$ process has maximum power at the origin, displaying the typical shape for most economic time series.[7] Additionally, we assume that $(1/T)X'X$ is a consistent estimator of the auto-covariance matrix Γ for any $N + 1$ contiguous observation space on the process. The explanation is by way of approximate diagonalization of Γ.

From Fuller (1976) and Wahba (1968), if $\{X_t\}$ is a covariance-stationary process with an absolutely summable covariance function $\sigma(h)$, there exists an orthogonal matrix Q independently of $\sigma(h)$, such that

$$Q'\Gamma Q \cong \Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{N+1}\}, \tag{17}$$

where Γ is the $N + 1$ order autocovariance matrix. The matrix Q is defined here for N even ($N + 1$ odd). The other case is similar and the argument is not changed by consideration of the case in which $N + 1$ is even (see Fuller (1976)).

$$(N+1)^{1/2}2^{-1/2}Q' = \begin{bmatrix} 2^{-1/2} & 2^{-1/2} & 2^{-1/2} & \cdots & 2^{-1/2} \\[2mm] 1 & \cos\dfrac{2\pi}{N+1} & \cos\dfrac{2\pi\cdot 2}{N+1} & \cdots & \cos\dfrac{2\pi\cdot N}{N+1} \\[2mm] 0 & \sin\dfrac{2\pi}{N+1} & \sin\dfrac{2\pi\cdot 2}{N+1} & \cdots & \sin\dfrac{2\pi\cdot N}{N+1} \\[2mm] 1 & \cos\dfrac{4\pi}{N+1} & \cos\dfrac{4\pi\cdot 2}{N+1} & \cdots & \cos\dfrac{4\pi\cdot N}{N+1} \\[2mm] \vdots & & & & \\[2mm] 0 & \sin\dfrac{N\pi}{N+1} & \sin\dfrac{N\pi\cdot 2}{N+1} & \cdots & \sin\dfrac{N\pi\cdot N}{N+1} \end{bmatrix}. \tag{18}$$

From Fuller, as N increases, the elements of $Q'\Gamma Q$ converge to the elements of a diagonal matrix Λ, and those elements are

$$\lambda_1 = \sum_{h=-\infty}^{\infty} \sigma(h) = 2\pi f(0), \tag{19}$$

---

[7] See Granger (1966).

and

$$\lambda_{2j} = \lambda_{2j+1} = \sum_{h=-\infty}^{\infty} \sigma(h)e^{i2\pi hj/(N+1)} = 2\pi f\left(\frac{2\pi j}{N+1}\right), \quad j = 1, 2, \ldots, \frac{N}{2}. \quad (20)$$

where f is the spectral density function of the process and an approximation to the characteristic roots of Γ.

To make use of the diagonalization theorem in evaluating the $\delta_2(h)$ criterion, it is required that $Q'\Gamma^{-1}Q$ tend to Λ^{-1}. For invertible moving-average processes, the inverse of the covariance matrix of the process is approximately a covariance matrix for the same order autoregressive scheme. Hence the required convergence is assured for a wide class of processes with absolutely summable covariance functions. In general, it is required that Γ^{-1} be an autocovariance matrix with absolutely summable rows.[8]

We shall now evaluate the $\delta_2(h)$ criterion in terms of the Γ matrix and its inverse, first for the estimator $C(0)'b$, where $C(0)$ is an $N + 1$ vector of unit elements and b is the OLS estimator of the lag coefficients. In terms of Γ, and recognizing that the roots of Γ^{-1} correspond to λ_j^{-1}, $j = 1, \ldots, N+1$, we have

$$\delta_2(0) \simeq \frac{C(0)'\Gamma^{-1}C(0)}{(N+1)\,\mathrm{tr}\,\Gamma^{-1}} = \frac{\lambda_1^{-1}}{\sum \lambda_j^{-1}}. \quad (21)$$

Thus for stationary processes having maximum power at the origin, the δ_2 criterion is minimized for $C(0)$ over all other nontrivial linear combinations of OLS estimators of β, in the approximate sense made clear in the preceding argument.[9] Equation (21) also shows that $C_0'\beta$ is estimated by OLS with equal precision to individual coefficients for the case of identically and independently distributed processes on the independent variable.

For the remaining moments, $C_h'\beta$, note that the matrix Q' is such that $Q'C_h$ displays the coefficients of a Fourier representation of the vector C_h i.e., that

$$Q'C(h) = \phi(h), \qquad h = 1, \ldots, N+1, \quad (22)$$

where

$$\phi(h)' = [a_{0h}, a_{1h}, b_{1h}, a_{2h}, b_{2h}, \ldots, a_{Nh/2}, b_{Nh/2}],$$

---

[8] We have assumed that the $(N+1)$-order matrix $T^{-1}(X'X)$ converges in probability to Γ as $T \to \infty$; hence we assume that as $N \to \infty$ and $T \to \infty$ then $N/T \to 0$.

[9] For some processes the convergence of $Q'\Gamma Q$ to Λ is not fast (see Gallant & Goebel (1976)), and the approximation is compounded since we are dealing with a consistent estimator for Γ^{-1} rather than Γ itself.

and the as and bs are the Fourier coefficients associated with the cosine and
sine transforms of the elements of C_h at frequencies $0, 1, 2, \ldots, (N/2)$. Hence

$$C_h'\Gamma^{-1}C_h \cong \sum_{j=1}^{N+1} \lambda_j^{-1}\phi_j^2(h), \tag{23}$$

where the ϕ_j are the elements of ϕ.

Since the approximate roots of Γ come in pairs except for the first (see
Eq. (20)), Eq. (23) can be written as

$$
\begin{aligned}
C_h'\Gamma^{-1}C_h \cong{}& \lambda_1^{-1}(a_{0h}^2) + \lambda_2^{-1}(a_{1h}^2 + b_{1h}^2) + \lambda_4^{-1}(a_{2h}^2 + b_{2h}^2) \\
&+ \cdots + \lambda_N^{-1}(a_{Nh/2}^2 + b_{Nh/2}^2), \qquad h = 1, 2, \ldots;
\end{aligned} \tag{24}
$$

therefore, approximately,

$$\delta_2(h) = \sum \phi_j^2 \lambda_j^{-1} \bigg/ \phi'\phi \sum_{j=1}^{N+1} \lambda_j^{-1}. \tag{25}$$

Thus the size of $\delta_2(h)$ depends on a weighted average of the inverse of
the spectrum for $\{X_t\}$, where the weights are the intensities of the Fourier
transforms of the vectors C_h. Again, note that for iid. processes on $\{X_t\}$ the
standard is $(N + 1)^{-1}$.

Table 3 shows the periodigrams for the vectors $C_h, h = 1, 2, 3, 4$, where
$N = 25$. Note that the weights decline monotonically for higher frequencies

TABLE 3

PERIODOGRAMS FOR C_h [a]

Frequency[b]	C_1	C_2	C_3	C_4
0	.735	.545	.429	.354
1	.162	.288	.342	.358
2	.041	.068	.090	.113
3	.019	.030	.042	.053
4	.011	.018	.025	.031
5	.007	.012	.017	.021
6	.005	.008	.012	.016
7	.004	.007	.010	.012
8	.004	.005	.008	.010
9	.003	.005	.007	.009
10	.003	.005	.006	.008
11	.003	.004	.006	.008
12	.003	.004	.006	.007

[a] $h = 1, 2, 3, 4$; $N + 1 = 25$.
[b] Multiple of $2\pi/25$.

except for column 4, and they pile up fairly closely to the origin. Periodograms for other choices of $N + 1$ are similar.

Thus for $\{X_t\}$ processes with typical economic spectra, OLS estimators for the moments μ_h, $h = 0, 1, \ldots, N$, can be expected to be best for low-order lag moments, and on the δ_2 criterion OLS estimators of the early moments compare favorably with a situation of no multicollinearity.

These theoretical results are consistent with the empirical results given in Table 1. Note that the OLS estimator for μ_0 is on a stronger footing than the succeeding early moments. The theoretical results require only that the spectral density of the $\{X_t\}$ process has maximum power at the origin for $\hat{\mu}_0$ to be better than the OLS estimator of any other linear combination of the βs.[10] For $\hat{\mu}_1$, $\hat{\mu}_2$, etc., all the spectral weights are involved, and there is a trade-off between the spectral weights and the spikes in the periodograms for the C_h. Also, estimation of the long-run lag effect is a meaningful exercise even when the βs alternate in sign.

The previous explanation was based on the condition that the auto-covariance sequence of X_t is absolutely summable. However, we expect that the explanation may be extended beyond this condition. For example, the autocorrelation sequence 1, ρ, ρ, ... for lags 0, 1, 2, ... is not absolutely summable. Such a process arises when X_t is an aggregate of X_{it}, and X_{it} has an error component in the i direction.[11] We have evaluated this case as the second example in this section. In the next section we also take up an example in which the $\{X_t\}$ process is nonstationary.

5. Other Considerations

Although μ_0, the long-run effect of a change in the control variable, is of immediate interest, μ_1, μ_2, etc. are of less interest than the normalized lag parameters, μ_1/μ_0, μ_2/μ_0, etc. Some theoretical results on the normalized mean lag are available (Burdick & Wallace, 1976). In that article it is shown that for large samples the covariance matrix for the least squares estimators is

$$\mathrm{Cov}\{\hat{\mu}_0, \hat{\mu}_1\} = K \begin{bmatrix} 1 & N/2 \\ N/2 & (N/2)^2 + \delta \end{bmatrix}, \qquad (26)$$

where $\delta > 0$. Thus, assuming $\mu_1/\mu_0 > 0$,

$$V(\hat{\mu}_1 | \hat{\mu}_0) = \frac{K}{\mu_0^2} \left[\left\{ \frac{N}{2} - \frac{\mu_1}{\mu_0} \right\}^2 + \delta \right] < \frac{1}{\mu_0^2} V(\hat{\mu}_1), \qquad (27)$$

---

[10] Therefore, even with seasonals showing up as spikes at higher frequencies, one can expect $\hat{\mu}_0$ to be a good estimator for most time series data.

[11] See Wallace & Hussain (1969).

where $V(\cdot)$ is variance in the limiting distribution. Hence unless μ_0 is small, precise estimation of μ_1 implies precise estimation of μ_1/μ_0, at least in an asymptotic sense. Similar bounds for higher-order normalized moments have not been discovered.

Most of economic time series used as $\{X_t\}$ in our model (1) have trends, and a theoretical explanation on nonstationary $\{X_t\}$ is difficult. However, the importance of the problem would justify consideration on a special case, namely,

$$X_t = aX_{t-1} + b + \varepsilon_t, \qquad a > 1, \qquad t = 1, \ldots, T, \qquad (28)$$

where ε_t is iid with variance σ_ε^2. Following Quenouille (1957, p. 57), we rewrite (28) as

$$X_t = a^{-1}X_{t+1} - a^{-1}b - a^{-1}\varepsilon_{t+1}, \qquad t = T, T-1, \ldots, 1.$$

It can then be shown that if T is sufficiently large, $[(T - N)^{-1}X'X]^{-1}$ is approximately the autocovariance matrix of a noninvertible moving-average process,

$$\sigma_\varepsilon^{-2}\begin{bmatrix} 1+a^2 & -a & \cdot & \cdot & \cdot & 0 \\ -a & \cdot & & \cdot & & \vdots \\ \vdots & & \cdot & & \cdot & -a \\ 0 & \cdot & \cdot & \cdot & -a & 1+a^2 \end{bmatrix},$$

except at the $(1, 1)$ and $(N + 1, N + 1)$ elements. Ignoring the differences in the two exceptional elements and taking the Fourier transform, we obtain the diagonal matrix having $\sigma_\varepsilon^{-2}|1 - a\exp(-i2\pi(j - 1)/N + 1)|^2$, $j = 1, 2, 3, \ldots, (N/2)$, which corresponds to the distribution of λ_j^{-1} (not λ_j) in Section 4. The expression $|1 - a\exp(-i\omega)|^2$, $\pi \geq \omega \geq -\pi$, attains a minimum at $\omega = 0$. Hence the theoretical explanation in Section 4 is applicable to this special case of nonstationarity.

If one were to go directly for lag moments in a new data set with little or no prior information about the form of the lag distribution, there is a problem of choosing N, the number of lagged terms to include. If N is too small, biased estimates are a result, and if N is too large, variances become a problem. An empirical approach would be to vary N, looking for stability in the lag moment estimators.

We suggest that one of the uses of form free estimation of lag moments is that they provide another specification check. In situations for which *ad hoc* priors are employed, fitted lag distributions yield definite implications about the lag weights. Thus, since form free estimators of the low-order moments can be expected to be precise, some unconstrained regressions may reveal specification error not otherwise noticeable. In support of this argument, Table 4 gives unconstrained estimates of the long-run marginal propensity to consume and mean lag using U.S. quarterly data given by Griliches *et al.* (1962), first quarter, 1955, through the second quarter, 1961.

MICHIO HATANAKA AND T. DUDLEY WALLACE

TABLE 4

ESTIMATES OF LAG MOMENTS, U.S. QUARTERLY CONSUMPTION
FUNCTION

N	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\mu}_1/\hat{\mu}_0$	$\hat{\rho}$	DW	F	d.f.
6	.905	.856	.946	.7	2.26	1145.0	43
	(.034)[a]	(.521)	(.325)				
10	.935	1.741	1.861	.6	2.46	590.0	35
	(.029)	(1.03)	(1.20)				
14	.952	2.639	2.743	.4	1.83	648.0	27
	(.016)	(1.131)	(1.375)				
18	.961	6.231	6.484	.3	1.97	1281.0	19
	(.009)	(.901)	(.869)				
22	.954	5.160	5.408	.4	2.56	664.8	11
	(.014)	(1.657)	(2.938)				

[a] Figures in parentheses are standard errors.

The reported $\hat{\rho}$s are estimates of a first-order autocorrelation coefficient based on the Hildreth–Lu (1960) procedure, the DW column reports Durbin–Watson statistics, the d.f. column gives degrees of freedom for the regression, and F is the overall regression F statistic.

The estimates of long-run marginal propensity are quite stable for $N \geq 10$, but the mean lag estimates take a large jump between $N = 14$ and $N = 18$. The jumpiness may be due to seasonals, although Griliches et al. reported "deseasonalized" data. There is also the question about stability of parameters in the consumption function. However, the data probably represent a fairly stable period for price level expectations as compared to the late 1960s and early 1970s. Even with this very rough pass at the data, the moment estimates are interesting when compared to the Zellner–Geisel (1970) results, which are derived from the same data. Zellner and Geisel used the Koyck lag with various error assumptions and various estimation procedures, including strong priors on μ_0. In some cases they got estimates of the long-run marginal propensity to consume greater than unity, and in all cases got estimates which implied very short mean lags (as low as 1/3 of a quarter and no higher than 3 quarters). Also, standard errors for the estimates in Table 4 compare favorably with those obtained by Zellner and Geisel, even in cases where they imposed very tight priors on μ_0.

ACKNOWLEDGMENTS

REFERENCES

Burdick, D. S., & Wallace, T. D. A theorem on an inequality of two quadratic forms and an application to distributed lags. *International Economic Review*, 1976, **17**, 769–771.

Fuller, W. A. *Introduction to statistical time series.* New York: Wiley, 1976.

Gallant, A. Ronald, & Goebel, J. Jeffery. Nonlinear regression with autocorrelated errors. *Journal of the American Statistical Association*, 1976, **71**, 961–967.

Granger, C. W. J. The typical spectral shape of an economic variable. *Econometrica*, 1966, 34, 150–161.

Griliches, Z., Maddala, G. S., Lucas, R., & Wallace, N. Notes on estimated aggregate quarterly consumption functions. *Econometrica*, 1962, **30**, 491–500.

Hildreth, C., & Lu, J. Y. *Demand relations with autocorrelated disturbances.* Technical Bulletin 276, Michigan State University, East Lansing, Michigan, 1960.

Quenouille, M. H. *The analysis of multiple time series.* New York: Hafner, 1957.

Silver, L., & Wallace, T. D. The lag relationship between wholesale and consumer prices. *Journal of Econometrics*, 1980, **12**, in press.

Wahba, G. On the distribution of some statistics useful in the analysis of jointly stationary time series. *Annals of Mathematical Statistics*, 1968, **39**, 1849–1862.

Wallace, T. D., & Hussain, A. The use of error components models in combining cross section with time series data. *Econometrica*, 1969, **37**, 55–72.

Zellner, A., & Geisel, M. S. Analysis of distributed lag models with applications to consumption function estimation. *Econometrica*, 1970, **38**, 865–888.