Annals of Economic and Social Measurement, 2/1, 1973

A MONTE CARLO STUDY OF COMPLEX FINITE DISTRIBUTED LAG STRUCTURES

BY MALCOLM COHEN, ROBERT GILLINGHAM, AND DALE HEIEN*

This article uses Monte Carlo methods to assess the effectiveness of estimation techniques for determining lag structures. We conclude that for the power estimator approximation of the polynomial, both the $\bar{R}^2$ criterion and the t statistics are of only marginal help in determining the correct length or shape of the lag structure. Restriction of some weights to zero provides some indication of the overall correctness of the specification. However, without a priori *knowledge, it is impossible to determine whether the restriction compensates for misspecification in the variable set or the length of the lag.*

I. INTRODUCTION

Early works involving distributed lag functions made highly restrictive assumptions regarding the form of the lagged response. Fisher [5] assumed declining arithmetic weights. Koyck [8] specified declining geometric weights. Recent researchers have avoided these over-restrictive assumptions in order to fit the real world into the framework of a lag model. Jorgenson [7] developed an estimation technique which requires only that the distributed lag function be a ratio of two polynomials. Almon [1] developed a technique which requires the function to be finite. Despite their increased generality, these new techniques still require prior specification regarding the *true* shape and length of the lag under consideration. However, econometricians usually have very little theoretical justification for assuming any particular shape and length of lag structure.[1]

The purpose of this paper is to assess, using Monte Carlo methods, whether estimation techniques can be used to detect the true shape and length of a particular type of underlying lag structure. A common but cumbersome technique used to estimate lags is the Almon technique which involves Lagrangian interpolation. A power series approximation has been suggested as an alternative but simpler means of estimating distributed lags. This alternative and some computational problems in its use are discussed in the next section. The third section provides a discussion of the design of our Monte Carlo experiments. Section four presents an analysis of the results.

II. THE POWER ESTIMATOR

Almon [1] has suggested an estimation technique for approximating any finite distributed lag using Lagrangian interpolation. A simple and equivalent method is outlined here.[2]

Let,

$$y_t = \gamma_1 \sum_{i=0}^{N-1} w_i x_{t-i} + u_t \qquad t = 1, \ldots, T \tag{1}$$

* The authors are indebted to Lester Taylor for comments on an earlier draft of the paper. However, the authors alone are responsible for any errors.

[1] See Griliches [6] for a brief discussion of this problem.

[2] This presentation is drawn largely from Tinsley [10].

where T is the number of observations on x and y, N is the length of the lag, u is a random error term with zero mean, constant variance (σ^2) and zero covariance, and the w_i's and γ_1 are unknowns to be estimated subject to the constraint that,

$$\sum_{i=0}^{N-1} w_i = 1. \tag{2}$$

Let

$$w_i^* = \gamma_1 w_i. \tag{3}$$

We take the w_i^*'s to lie on a function which can be approximated by the Kth order polynomial,

$$w_i^* = m_0 + m_1 i + m_2 i^2 + \ldots + m_K i^K \qquad i = 0, \ldots, \ldots, N-1 \tag{4}$$

Given (4) equation (1) can be rewritten in matrix notation as,

$$y = X\,Pm + v \tag{5}$$

where X is a $T \times N$ matrix of current and successive lagged values of x; y is a $T \times 1$ vector of observations on y, v is a $T \times 1$ vector of random errors, m is a $K \times 1$ vector of polynomial coefficients, P is a power matrix defined as

$$p_{ij} = i^j \qquad i = 0, 1, \ldots, N-1; \qquad j = 0, 1, \ldots, K.^3$$

Applying ordinary least squares to (5) results in

$$\hat{m} = (P'X'XP)^{-1}P'X'y \tag{6}$$

and

$$\hat{w}^* = P\hat{m} \tag{7}$$

where $w^* = Pm$. If the weights defined by (4) lie exactly on the approximating polynomial and all the customary least squares assumptions are met, then

$$E(\hat{w}^*) = w^*. \tag{8}$$

The variance covariance matrix of Pm will be

$$\text{var}\,(w^*) = \sigma^2 P(P'X'XP)^{-1}P'. \tag{9}$$

Generalization of this presentation to allow for a constant term and additional variables is straightforward.

In general, the vectors of the matrix XP will be highly multicollinear and of substantially different magnitudes. Hence $P'X'XP$ will be difficult to invert accurately, with a danger that rounding errors may obscure the results. Tests suggested by Longley [9] and others confirmed the ill-conditioning of our transformed matrix. To overcome this problem we used the Gram–Schmidt orthonormalization process and high precision arithmetic to estimate the inverse of the

[3] We employ the convention that $0^0 = 1$.

transformed matrix.[4] We caution other researchers engaged in distributed lag estimation using the power estimator that problems of numerical accuracy are likely to be acute. Wampler [12] provides a description of the relative accuracy of the most popular inversion algorithms. Rounding errors in the Almon technique which involves even more computations are likely to be as severe.

III. Design of the Monte Carlo Experiments

In order to perform the Monte Carlo experiments, we compute for any given lag structure and length of lag a set of observations on the dependent variable according to

$$y_t = \gamma_0 + \gamma_1 \sum_{i=0}^{N-1} w_i x_{t-i} + \gamma_2 z_t, \qquad t = 1, \ldots T. \tag{10}$$

The variables x and z are time series data on Corporate Profits After Taxes and Gross National Product. The coefficients γ_0, γ_1 and γ_2 were obtained from an unlagged regression of Gross Private Nonresidential Investment on x_t and z_t. Their values are $\gamma_0 = 69.8133$, $\gamma_1 = 0.3826$, and $\gamma_2 = 0.832$. In the experiments, t runs quarterly from 1951 IV to 1967 IV with additional observations on x depending on the length of the lag. Corporate Profits was chosen as the lagged variable since it is less autocorrelated than GNP. Distributed lags on non-autocorrelated variables should be easier to detect since the effect of each individual weight will be more pronounced. Next a set of random error terms were added to the dependent variable in (10), for each replication within a given Monte Carlo experiment. These pseudo-random errors were generated by a computer subroutine written at the Bureau of Labor Statistics. The method used is derived from an algorithm developed by Behrenz [2], while the approach suggested by Box and Muller [4] is used to obtain a normal distribution. The first two sample moments of the generated numbers are tested within the program to insure normality. The number of replications was set at 100 after determining that this number would provide stable estimates of the parameters to be analyzed. Two series of experiments using two distinct lag structures were performed. In each series, true weights were chosen which lie on a polynomial of known degree K and which satisfy the condition $\sum_{i=0}^{N-1} w_i = 1$. Series A, using a quartic lag, was designed to test for the effect of misspecifying the degree of the approximating polynomial. The assumed degree is denoted as k to differentiate it from the true

[4] The GS process substitutes a sophisticated form of elimination procedure for the solution of the matrix of inner products as in more conventional regression techniques. Basically, it operates directly on the vectors of X, taking each in turn and eliminating its influence from the remaining vectors. It differs from straightforward elimination procedures in that it does not take the dependent variable (y) into account at all stages, but rather first transforms the vectors of X to have zero intercorrelations—in this way dealing with the problem of multicollinearity. From a numerical standpoint, the electronic computer cannot carry the same number of decimals in all stages of calculation. It thus appears beneficial when there is wide size variance in the vectors to precondition the matrix by subtracting out integer means converted to floating point, negative integer means being inserted in the first row of the identity matrix. This step greatly reduces the incidence of computer round-off error. The interested reader should see Longley [9], Walsh [11], Wampler [12], and Yule and Kendall [13], for more details. The use of GS as the inversion algorithm is equivalent to using the transformation suggested by Tinsley [10] for generating an orthonormal set of weighting vectors.

degree K. Throughout this series the assumed length of lag (n) was correctly specified as the true length (N). Series B, using quadratic lag structure, was designed to test for the effect of misspecifying the length of lag. Throughout this series the assumed polynomial degree was correctly specified as 2. Both series of experiments were run using two different variances for the Monte Carlo error vector. The two values of σ^2 were designed to provide, under correct specification, average R^2's of approximately 0.95 and 0.80. For each variance, experiments A and B were run for all combinations of (a) restricting (R) and not restricting (NR) the nth weight to zero and (b) not misspecifying (NM) and misspecifying (M) the variable set by omitting z.

In each experiment we calculated from the 100 replications the mean, standard error, and t-ratio for $\hat{\beta}$, $\hat{m}$, $\hat{\gamma}$, and the $\hat{w}_i$'s. In addition we report the mean R^2 and a goodness of fit statistic defined as:

$$G = \sum_{i=0}^{N-1} \left(\frac{\bar{w}_i^*}{\bar{\gamma}_1} - w_i \right)^2 \tag{11}$$

where $\bar{w}_i^*$ is the mean of the non-normalized weights over the 100 replications and $\bar{\gamma}_1$ is the mean of γ_1 computed from the 100 replications, and w_i is the true weight. Since $w_i^* = \gamma_1 w_i$ and $\sum w_i = 1$, $\bar{\gamma}_1 = \sum_{i=0}^{N-1} \bar{w}_i^*$. The calculation in (11) is more stable than one involving

$$\hat{w}_i = \hat{w}_i^* \Big/ \left(\sum_{i=0}^{N-1} \hat{w}_i^* \right)$$

computed for each regression. Both the numerator and denominator in (11) are computed across all 100 regressions. The smaller G, the better the fit. We also computed $\bar{R}^2$ and a Durbin Watson statistic. These statistics are discussed when they have special importance.

Before reporting the results, note that applying the restriction $w_n = 0$ to the estimated relation can result in a zero variance for a particular weight, or a constant difference between two weights. For example, from (4):

$$\hat{w}_i^* = \sum_{j=0}^{k} \hat{m}_j i^j \tag{12}$$

and

$$\hat{w}_n^* = \sum_{j=0}^{k} \hat{m}_j n^j \tag{13}$$

If $w_n = 0$, we can subtract (12) from (13) to eliminate m_0, leaving

$$\hat{w}_i^* = \sum_{j=1}^{k} \hat{m}_j (i^j - n^j). \tag{14}$$

If, for example, $k = 2$ and $n = 4$ then

$$\sum_{i=0}^{4} \hat{w}_i^* = -10\hat{m}_1 - 50\hat{m}_2 \qquad \hat{w}_1^* = -3\hat{m}_1 - 15\hat{m}_2 \tag{15}$$

$$\hat{w}_2^* - \hat{w}_3^* = -\hat{m}_1 - 5\hat{m}_2$$

and thus $w_1 = 0.3$ and $w_2 - w_3 = 0.1$, each having zero variance.

IV. Analysis of the Results

Experiment A. Experiment A deals with the problems which arise from misspecifying the degree of the approximating polynomial. The power estimator reduces the number of independent variables to be estimated when $k + 1 < n$. However if $k < K$ the approximation may result in an unsatisfactory fit. Experiment A is designed to detect the consequences of such misspecification. The results of specifying $k \geq K$ also are noted. Table I provides a tabulation of summary statistics for the Monte Carlo runs of Experiment A.

TABLE I

Experiment A—Varying Degree of Approximating Polynomial

Summary Statistics

	$k = 2$		$k = 4$		$k = 6$	
*	R^2	G	R^2	G	R^2	G
NM–NR	0.946	0.0224	0.946	0.0033	0.950	0.0174
NM–NR	0.817	0.0190	0.823	0.0134	0.828	0.0240
NM–R	0.943	0.0884	0.946	0.0007	0.949	0.0053
NM–R	0.812	0.0893	0.815	0.0017	0.821	0.0055
M–NR	0.913	0.1078	0.919	0.1281	0.921	0.1638
M–NR	0.789	0.1068	0.792	0.1322	0.799	0.1698
M–R	0.912	0.1019	0.915	0.0959	0.917	0.1190
M–R	0.781	0.0999	0.791	0.0919	0.801	0.1135

* NM — The equation was not misspecified.
NR — The last weight was not restricted to zero.
M — The equation was misspecified.
R — The last weight was restricted to zero.
$K = 4$
$N = 11$
$n = 11$
G — is defined in the text equation (11).

The combination of experiments shown in Table I was run for three different specifications for the degree of the approximating polynomial. The detailed results for these experiments are presented in Table II. We would expect, and do in fact, obtain, the best results (as measured by the lowest G) when $k = K = 4$ and with a full set of independent variables. In addition, a very good fit is obtained when $k = 6$, the equation is not misspecified, and w_n is restricted to zero. Lowering the error variance has little effect on the goodness of fit unless the equation is appropriately specified. Here, however, the percentage improvement of G is large but the absolute decrease in G is small. None of the approximating polynomials give good estimates of the complicated lag structure when a variable is omitted from the equation. Restricting w_n to zero improves the goodness of fit when $k \geq 4$ and the lag structure thus has sufficient degree to simulate the polynomial.

The truly interesting question from Tables I and II is whether the regression statistics produced by the experiments can lead us to the correct specification of the lag structure without prior knowledge of K and N. Table I demonstrates the basic insensitivity of the R^2 criterion to changes in either the parameter k or restriction

TABLE II

EXPERIMENT A—VARYING DEGREE OF APPROXIMATING POLYNOMIAL

True and Estimated Weights*

Specification			NM	NM	NM	NM	MS	MS	MS	MS
Restriction			NR	NR	R	R	NR	NR	R	R
σ^2			15	60	15	60	15	60	15	60
Pd.	True wts.									
0	0.1511		0.2204	0.1992	0.2446	0.2209	0.1694	0.1776	0.1630	0.1692
1	0.2059	$N = 11$	0.2290	0.2098	0.2011	0.1860	0.1510	0.1562	0.1492	0.1531
2	0.2389	$n = 11$	0.2262	0.2096	0.1619	0.1541	0.1335	0.1361	0.1352	0.1372
3	0.2430	$K = 4$	0.2118	0.1987	0.1269	0.1252	0.1169	0.1174	0.1209	0.1214
4	0.2160	$k = 2$	0.1859	0.1770	0.0961	0.0992	0.1012	0.1001	0.1065	0.1057
5	0.1606		0.1485	0.1446	0.0697	0.0761	0.0864	0.0841	0.0919	0.0902
6	0.0845		0.0995	0.1015	0.0474	0.0560	0.0725	0.0695	0.0771	0.0748
7	0.0000		0.0391	0.0476	0.0294	0.0389	0.0595	0.0562	0.0621	0.0596
8	−0.0754		−0.0328	−0.0171	0.0157	0.0247	0.0474	0.0443	0.0468	0.0445
9	−0.1194		−0.1163	−0.0924	0.0062	0.0135	0.0361	0.0338	0.0314	0.0295
10	−0.1050		−0.2113	−0.1785	0.0010	0.0053	0.0258	0.0247	0.0158	0.0147
0	0.1511		0.1225	0.2154	0.1473	0.1500	0.2031	0.1927	0.1439	0.1321
1	0.2059	$N = 11$	0.2260	0.1718	0.2066	0.1954	0.1096	0.1089	0.1543	0.1592
2	0.2389	$n = 11$	0.2701	0.1836	0.2374	0.2236	0.0968	0.1000	0.1477	0.1574
3	0.2430	$K = 4$	0.2648	0.2053	0.2371	0.2274	0.1147	0.1183	0.1313	0.1388
4	0.2160	$k = 4$	0.2214	0.2062	0.2068	0.2038	0.1289	0.1311	0.1105	0.1131
5	0.1606		0.1519	0.1709	0.1511	0.1547	0.1201	0.1203	0.0895	0.0872
6	0.0845		0.0690	0.0987	0.0783	0.0865	0.0840	0.0829	0.0709	0.0656
7	0.0000		−0.0137	0.0038	0.0004	0.0100	0.0320	0.0305	0.0557	0.0503
8	−0.0754		−0.0816	−0.0844	−0.0674	−0.0592	−0.0097	−0.0102	0.0436	0.0406
9	−0.1194		−0.1194	−0.1219	−0.1057	−0.1011	0.0001	0.0020	0.0328	0.0332
10	−0.1050		−0.1110	−0.0495	−0.0920	−0.0911	0.1199	0.1234	0.0198	0.0225
0	0.1511		0.2036	0.1317	0.1434	0.1177	0.2327	0.2629	0.1929	0.1549
1	0.2059		0.1516	0.2247	0.2582	0.2123	0.0228	0.0088	0.1044	0.1414
2	0.2389	$N = 11$	0.2464	0.2760	0.2412	0.2201	0.1333	0.1084	0.1016	0.1172
3	0.2430	$n = 11$	0.2571	0.2333	0.2093	0.2224	0.1738	0.1647	0.1289	0.1163
4	0.2160	$K = 4$	0.1911	0.1608	0.1858	0.2221	0.1229	0.1305	0.1413	0.1226
5	0.1606	$k = 6$	0.1203	0.1206	0.1535	0.1948	0.0717	0.0794	0.1186	0.1118
6	0.0845		0.0808	0.1142	0.0937	0.1232	0.0707	0.0649	0.0678	0.0759
7	0.0000		0.0414	0.0811	0.0071	0.0164	0.0784	0.0663	0.0179	0.0315
8	−0.0754		−0.0558	−0.0430	−0.0794	−0.0884	0.0136	0.0214	0.0033	0.0110
9	−0.1194		−0.1846	−0.2109	−0.1227	−0.1395	−0.0906	−0.0530	0.0389	0.0371
10	−0.1050		−0.0520	−0.0885	−0.0900	−0.1012	0.1707	0.1457	0.0844	0.0803

* See Table I for glossary of abbreviations.

of the last weight to zero, although it is helpful at the more basic level of specifying the variable set. The R^2 criterion is of little help in choosing the correct degree of the approximating polynomial.

The standard errors for both the polynomial interpolation coefficients (the $\hat{m}_i$'s) and the non-normalized weights ($\hat{w}_i^*$'s) are derived from the variance covariance matrix for the interpolation coefficients. The multicollinearity of the independent variables (to be expected in time series analyses) causes the elements of this matrix and thus both sets of standard errors to be relatively large. This effect is especially pronounced for the interpolation coefficients. When $k = K = 4$, they exceed their standard errors only when there is an omitted variable, and, even in this case for only four out of the twenty coefficients. The non-normalized weights are more often significantly different from zero. High t-values are more

predominant in the first (positive) portion of the lag structure. However, with $k = K = 4$, an estimated weight exceeds twice its standard error only once, unless the standard error is decreased by either (1) restriction of w_n to zero, or (2) omission of an independent variable. In other words, the size of the standard errors makes it difficult to determine even the sign of either an interpolation coefficient or a lag weight.

TABLE III

EXPERIMENT A—VARYING DEGREE OF APPROXIMATING POLYNOMIAL

Coefficients, Weights and t Values

$\sigma^2 = 15$

		Not misspecified Restricted		
	True	$k = 2$	$k = 4$	$k = 6$
Interpolation Coefficients				
1		0.1189	0.0550	0.0573
		(2.97)	(0.55)	(0.38)
2		0.0222	0.0268	0.1025
		(−1.38)	(0.17)	(0.14)
3		0.0010	−0.0042	−0.0783
		(0.88)	(−0.07)	(0.12)
4			−0.0005	0.0254
			(−0.07)	(0.11)
5			0.0001	−0.0041
			(0.18)	(−0.11)
6				0.0003
				(0.11)
7				−0.0000
				(−0.10)
Non-normalized Weights				
0	0.0578	0.1189	0.0550	0.0573
		(2.97)	(0.55)	(0.38)
1	0.0788	0.0977	0.0771	0.1031
		(3.62)	(2.63)	(0.68)
2	0.0914	0.0787	0.0886	0.0963
		(4.35)	(1.78)	(1.18)
3	0.0930	0.0617	0.0885	0.0836
		(4.12)	(2.09)	(0.97)
4	0.0826	0.0467	0.0771	0.0742
		(2.79)	(2.71)	(1.19)
5	0.0614	0.0339	0.0564	0.0613
		(1.73)	(1.86)	(0.97)
6	0.0323	0.0230	0.0292	0.0374
		(1.08)	(0.87)	(0.58)
7	0.0000	0.0143	0.0001	0.0028
		(0.67)	(0.00)	(0.06)
8	−0.0288	0.0076	−0.0251	−0.0317
		(0.39)	(−0.97)	(−0.41)
9	−0.0457	0.0030	−0.0394	−0.0490
		(0.20)	(−1.00)	(−0.95)
10	−0.0402	0.0005	−0.0343	−0.0360
		(0.05)	(−0.83)	(−0.32)
Long run Coefficient	0.3826	0.4860	0.3730	0.3995

Given the above described difficulty in assigning confidence intervals for coefficients, we might still ask if information can be gained from the *relative* significance of either the interpolation coefficients or the non-normalized weights. As pointed out above, neither R^2 nor the standard error of estimate is sensitive to misspecification in k. Therefore, the relative size of the standard errors as k is changed is determined by changes in either $(P'X'XP)^{-1}$ or $P(P'X'XP)^{-1}P'$. The standard errors therefore increase with the number of intercorrelated regressors included in the equation, rather than reacting to the correctness of the specification. In general, the lower the number of regressors the lower the standard errors and thus the higher the t-values. Furthermore, the relative significance of an additional interpolation coefficient is not systematically related to whether $k \lessgtr K$.

Table III illustrates the difficulty of trying to determine correct lag specification from regression results. All the t-values vary inversely with k. Thus the t-values for $k = 2$ are highest even though this specification results in the most biased weight estimates and a serious bias in the long run coefficient of the lagged variable. On the other hand, when $k(= 6)$ is overestimated with correspondingly high variances, the expected fit of the equation is superior to when $k = 2$. Without *a priori* knowledge of the true lag structure, there is no reason to select $k = 4$ as the correct specification.

Experiment B. Experiment B deals with the consequences of misspecifying the estimated length of the lag, n. K and k both were set at 2. N was 11 and experiments were run for $n = 7$, $n = 11$, and $n = 15$. The results of these experiments are shown in Table IV through VI. Interpretation of these results when n, k, and the variable set are correctly specified is less straightforward than interpretation of the correctly specified quartic lag structure of experiment A. The general lag shape is faithfully reproduced only when the error variance is low. When the variance is high, the estimator misses the mode of the distribution. When a variable is omitted, the lag structure is convex to the x-axis unless w_n is restricted to zero. Surprisingly, the high variance, restricted equation (R) results in a very accurate estimate of the weight structure.

TABLE IV

EXPERIMENT A—VARYING LENGTH OF LAG

Summary Statistics

*	$n = 7$		$n = 11$		$n = 15$	
	R^2	G	R^2	G	R^2	G
NM–NR	0.941	0.0338	0.942	0.0004	0.938	0.0027
NM–NR	0.803	0.0395	0.802	0.0021	0.790	0.0034
NM–R	0.938	0.0344	0.940	0.0000	0.940	0.0048
NM–R	0.793	0.0331	0.802	0.0022	0.799	0.0052
M–NR	0.896	0.0676	0.907	0.0113	0.916	0.0256
M–NR	0.759	0.0576	0.771	0.0080	0.783	0.0224
M–R	0.885	0.0273	0.904	0.0002	0.911	0.0108
M–R	0.745	0.0281	0.769	0.0000	0.775	0.0100

* Definitions same as in Table I
$K = 2$
$N = 11$
$k = 2$

TABLE V

EXPERIMENT B—VARYING LENGTH OF LAG

True and Estimated Weights

Specification		NM	NM	NM	NM	MS	MS	MS	MS*
Restriction		NR	NR	R	R	NR	NR	R	R
σ^2		15	60	15	60	15	60	15	60
Pd.	True wts.								
0	0.1129	0.1957	0.2073	0.0619	0.0677	0.2686	0.2575	0.1515	0.1613
1	0.1173 $N = 11$	0.1419	0.1415	0.1337	0.1362	0.1530	0.1547	0.1721	0.1763
2	0.1188 $n = 6$	0.1098	0.1023	0.1786	0.1786	0.0816	0.0906	0.1786	0.1786
3	0.1173 $K = 2$	0.0994	0.0897	0.1966	0.1949	0.0545	0.0653	0.1710	0.1682
4	0.1120 $k = 2$	0.1108	0.1037	0.1878	0.1853	0.0715	0.0788	0.1494	0.1452
5	0.1056	0.1438	0.1442	0.1520	0.1495	0.1327	0.1310	0.1136	0.1095
6	0.0953	0.1986	0.2213	0.0895	0.0878	0.2382	0.2221	0.0639	0.0611
7	0.0821								
8	0.0660								
9	0.0469								
10	0.0249								
0	0.1129	0.1102	0.0971	0.1134	0.1446	0.1412	0.1399	0.1223	0.1116
1	0.1173 $N = 11$	0.1130	0.1025	0.1176	0.1375	0.1202	0.1218	0.1233	0.1165
2	0.1188 $n = 11$	0.1136	0.1057	0.1189	0.1291	0.1028	0.1066	0.1219	0.1184
3	0.1173 $K = 2$	0.1120	0.1068	0.1173	0.1196	0.0891	0.0940	0.1180	0.1172
4	0.1129 $k = 2$	0.1081	0.1056	0.1128	0.1089	0.0791	0.0842	0.1117	0.1131
5	0.1056	0.1021	0.1021	0.1054	0.0969	0.0727	0.0772	0.1030	0.1059
6	0.0953	0.0937	0.0964	0.0951	0.0838	0.0699	0.0729	0.0919	0.0958
7	0.0821	0.0832	0.0884	0.0819	0.0694	0.0708	0.0713	0.0783	0.0826
8	0.0660	0.0704	0.0783	0.0658	0.0539	0.0753	0.0725	0.0624	0.0665
9	0.0469	0.0554	0.0659	0.0468	0.0371	0.0835	0.0765	0.0440	0.0473
10	0.0249	0.0382	0.0512	0.0248	0.0192	0.0954	0.0831	0.0232	0.0252
0	0.1129	0.1321	0.1222	0.1451	0.1160	0.0987	0.1001	0.0846	0.0874
1	0.1173 $N = 11$	0.1243	0.1223	0.1314	0.1101	0.0896	0.0916	0.0870	0.0891
2	0.1188 $n = 15$	0.1162	0.1205	0.1183	0.1039	0.0816	0.0839	0.0883	0.0897
3	0.1173 $K = 2$	0.1077	0.1169	0.1057	0.0974	0.0746	0.0771	0.0884	0.0892
4	0.1129 $k = 2$	0.0988	0.1114	0.0938	0.0907	0.0686	0.0711	0.0874	0.0877
5	0.1056	0.0896	0.1041	0.0824	0.0838	0.0637	0.0660	0.0853	0.0851
6	0.0953	0.0800	0.0949	0.0716	0.0765	0.0598	0.0619	0.0819	0.0815
7	0.0821	0.0700	0.0839	0.0613	0.0691	0.0569	0.0585	0.0775	0.0767
8	0.0660	0.0597	0.0710	0.0516	0.0613	0.0551	0.0561	0.0718	0.0709
9	0.0469	0.0490	0.0563	0.0425	0.0533	0.0544	0.0545	0.0650	0.0640
10	0.0249	0.0380	0.0398	0.0340	0.0451	0.0547	0.0538	0.0571	0.0560
11	0.0000	0.0266	0.0214	0.0261	0.0366	0.0560	0.0540	0.0480	0.0470
12	0.0000	0.0149	0.0012	0.0187	0.0278	0.0584	0.0550	0.0377	0.0368
13	0.0000	0.0028	−0.0209	0.0119	0.0188	0.0618	0.0569	0.0263	0.0256
14	0.0000	−0.0097	−0.0448	0.0057	0.0095	0.0662	0.0597	0.0137	0.0134

* See Table 1 for glossary of abbreviations.

As in experiment A, we wish to determine if the output from the distributed lag regression provides us with information for correctly specifying the distributed lag. In this experiment R^2 is sensitive only to a misspecification of the variable set and not to misspecification of the length of the lag. The Durbin Watson is also somewhat sensitive to misspecification of the variable set (especially when σ^2 is low), but is useless for assessing the estimated length of lag.

The standard errors and t-values of the interpolation coefficients and nonnormalized weights provide little more information for specifying the degree of the

approximating polynomial. As in experiment A, t-values for the interpolation coefficients are in the main insignificantly different from zero. They are increased when either a variable is omitted or w_n is restricted, but not to the same extent as are the t values of the weights. The t-values for the non-normalized weights may provide a small amount of judgmental information about the length of the lag. When $n = 15$ and the equation is otherwise specified correctly, the t-value for the weights in periods 12 through 15 are extremely low. However, when $n = 11$ or 15, weights in periods 9, 10, and 11 also have low t-values, so one cannot determine the actual length of the lag but only infer from the long tail of relatively insignificant weights that the length may be misspecified. In addition, when omission of a variable resulted in decreased standard errors for the non-normalized weights, even the spurious weights were significantly different from zero when σ^2 was small.

Restriction of w_n to zero also provides some indication of the overall correctness of specification. The zero restriction can, *ceteris paribus*, alter the shape of the lag distribution and make it appear much more reasonable. Table VI illustrates two such cases. In only one of the two cases in the table, however, was the more reasonable appearing lag structure also correct and that occurred when the misspecification was in the variable set rather than in the length of lag. Without *a priori* knowledge of the true lag shape, it is impossible to determine for what type of misspecification the zero restriction is compensating. When the equation is correctly specified, the effect of the restriction is far less.

TABLE VI

EXPERIMENT B—VARYING LENGTH OF LAG

Coefficients, Weights and t Values

		Full variable set $\sigma^2 = 60$		Variable Omitted $\sigma^2 = 60$	
	True Weights	NR $n = 6$	R $n = 6$	NR $n = 11$	R $n = 11$
Normalized Weights					
0	0.1129	0.2073	0.0677	0.1399	0.1116
1	0.1173	0.1415	0.1362	0.1218	0.1165
2	0.1188	0.1023	0.1786	0.1066	0.1184
3	0.1173	0.0897	0.1949	0.0940	0.1172
4	0.1129	0.1037	0.1853	0.0842	0.1131
5	0.1056	0.1442	0.1495	0.0772	0.1059
6	0.0953	0.2113	0.0878	0.0729	0.0958
7	0.0821			0.0713	0.0826
8	0.0660			0.0725	0.0665
9	0.0469			0.0765	0.0473
10	0.0249			0.0831	0.0252

V. SUMMARY

The purpose of this paper is to assess whether the shape and length of lags can be estimated in the absence of *a priori* information. Monte Carlo experiments were run using the power estimator. The power estimator approximates the shape

of a polynomial by a power series expansion. By examining the measure of fit, G, we are able to answer the question—is it possible to estimate the true shape and length of lag by repeated regressions with different shapes and different lengths when the true shape and length are *not known a priori*? Our conclusion is that both the $\bar{R}^2$ (or R^2) criterion and the t statistics are of marginal help in determining either the correct length or shape of lag. Experiments were run omitting a non-lagged variable from the regression. This resulted in considerable deterioration in the closeness of the estimated and true weights, a lower R^2 or $\bar{R}^2$, and about the same difficulty in detecting either the shape or length of lag. The Durbin Watson statistic was not of much use except in possibly detecting a misspecified non-lagged variable.

The significance statistics for $\hat{m}$ and $\hat{w}^*$ do not provide a basis for determining K. The t-statistics are inversely related to the magnitude rather than the correctness of specification of k. Some possibility for detecting the correct length of the lag was suggested by examining the tail of the estimated lag. However, the possible information gain appears extremely slight. In general, restriction of the nth weight to zero has a more marked effect on the weight estimates when some form of misspecification has been made. Whether weight estimates are improved depends upon what type of misspecification is present.

Like any Monte Carlo studies these results may be peculiar to our particular model and may not hold in general. The experiment is applicable at most for those lag structures which are smooth in the sense that an approximating polynomial can be specified of degree K such that $K + 1 < N$.

University of Michigan
Bureau of Labor Statistics

References

[1] Almon, Shirley, "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica*, January 1965, pp. 178–196.
[2] Behrenz, Peter G., "Algorithm 133, Random," *Communications of the ACM*, p. 553.
[3] Bischoff, Charles W., "A Study of Distributed Lags and Business Fixed Investment," Cambridge, M.I.T. Ph.D. Thesis, unpublished, 1966.
[4] Box, G. E. and Muller, M. E., "A Note on the Generation of Random Normal Deviates," *The Annals of Mathematical Statistics*, vol. 29, 1958, pp. 610–11.
[5] Fisher, Irving, "Note on a Short-Cut Method for Calculating Distributed Lags," *Bulletin de l'Institut International de Statistique*, 1937, pp. 323–327.
[6] Griliches, Zvi, "Distributed Lags: A Survey," *Econometrica*, January 1967, pp. 16–49.
[7] Jorgenson, Dale, "Rational Distributed Lag Functions," *Econometrica*, January 1966, pp. 135–144.
[8] Koyck, L. M., *Distributed Lags and Investment Analysis*, Amsterdam: North-Holland Publishing Company, 1954.
[9] Longley, James W., "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User," *Journal of the American Statistical Association*, September 1967, pp. 819–841.
[10] Tinsley, Peter A., "On Polynomial Approximation of Distributed Lags," unpublished paper, 1967.
[11] Walsh, Philip J., "Algorithm 127," *Communications of the ACM*. Vol. 5, 1962, pp. 511–13.
[12] Wampler, Roy, "A Report on the Accuracy of Some Widely Used Least Squares Computer Programs," *Journal of the American Statistical Association*, 1970.
[13] Yule and Kendall, *An Introduction to the Theory of Statistics*, Chapter 12, New York: Hafner Publishing Company, 1950.