Faculty Working Papers

SPLINE LAGS: WHY THE ALMON LAG HAS GONE TO PIECES

Dale J. Poirier

#235

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 19, 1975

SPLINE LAGS: WHY THE ALMON LAG HAS GONE TO PIECES

Dale J. Poirier

#235

SPLINE LAGS: WHY THE ALMON LAG HAS GONE TO PIECES

by

Dale J. Poirier*

"We may be asking too much of our data. We want them to test our theories, provide us with estimates of important parameters, and disclose to us the exact time form of the interrelationships between the various variables. Progress in this area is likely to be slow until we have a much better theoretical base for imposing a time-lag structure on the data."

Zvi Griliches (1967, pp. 17-18)

1. Introduction

The distributed lag literature is voluminous to say the least. One indication of the continuing growth of the literature is that since the survey article of Griliches (1967), three subsequent survey works have come on the scene: Dhyrmes (1971), Nerlove (1971), and Sims (1973). The already extensive "classical" literature on the subject has been recently augmented by numerous studies taking Bayesian approaches (e.g., Chetty (1971), Leamer (1972), and Shiller (1973)). This influx of Bayesian ideas has been fruitful if for no other reason than to emphasize why conventional unconstrained least squares estimation techniques often perform so unsatisfactorily. Instead of bad data, the real culprit is revealed to

---

*The author is an assistant professor of economics at the University of Illinois at Urbana-Champaign. An earlier version of this study was presented at the NBER-NSF Workshop on Segmented and Switching Regressions held in Madison, Wisconsin June 3-4, 1974. The comments of its participants were most appreciated, however, none of them should be held responsible for any weaknesses that remain. Also the author wishes to acknowledge the computational assistance of Jeffrey Greenspan in preparing the empirical results of sections 4 and 5.

be inadequate prior information. The opening quotation — by Griliches (1967) — and the works of Leamer (19 2, 1973), Nerlove (1971), and Sims (1973) are aimed at awakening researches to this reality.

While there has been progress both in developing dynamic economic theory and in efforts to apply it to distributed lag models, its role, in the words of Sims (1973, p. 4), has "been for the most part not to provide us with solutions, but to help us better understand the limits of our ignorance." In practice the typical applied researcher incorporates prior information in a distributed lag model in an ad hoc fashion. Few attempts are made to derive the lag structure from an optimizing framework, and for the most part, whatever prior information is available is loosely used in selecting a lag parameterization.[1]

The imposition of any parameterization on the lag weights can be viewed from two distinct methodological vantage points. The first assumes that nonstochastic restrictions reflecting the lag parameterization are in fact literally true. (For example, the lag weights may be required to be on a cubic polynomial.) If these restrictions are false, then the classical estimation approach yields biased and inconsistent estimators of the lag weights. The requirement that the weights lie exactly (not even off by 10^{-6}) on the assumed lag parameterization is viewed much the same way the whole standard classical linear regression model is viewed — as a theoretical construct, adequate for the purposes at hand, which captures the "flavor" of the theoretical model, and which serves in a sense as a

---

[1]One classic counter-example is the derivation of the partial adjustment model from a cost minimization hypothesis. See Poirier (1973c) for a further extension.

place to "hang one's hat" and begin analysis. Obviously, such stringent requirements demand that all available prior information be examined to insure that the specified lag parameterization does not rule out any lag shapes which are theoretically possible.

From the second vantage point the lag parameterization restrictions are not viewed as being exactly true, but rather only as a "reasonable" approximation to the true unknown lag structure. If these restrictions are in fact only approximately true and if they are specified in a nonstochastic manner, then the resulting lag weight estimators will be biased and inconsistent — although they may still dominate other estimators in mean square error. Such approximate restrictions can yield consistent estimators if they are specified stochastically along the lines suggested by Theil and Goldberger (1961) and Theil (1963).[2] However, it seems (to this author at least) that once such "mixed" estimators are considered, the full-fledged Bayesian framework provides a much better vehicle in which to incorporate stochastic prior information.

In any case such estimation issues are <u>not</u> directly the subject of this study. Rather the issue here is the choice of lag parameterization. Clearly, from both vantage points it is desirable to incorporate all available prior information concerning the shape of the lag, and in the case of the second vantage point, it is also desirable to select a lag parameterization which is flexible enough to serve as an adequate approximation to other possible lag structures.

---

[2]See Yancey, Bock, and Judge (1972) and Mehta and Swamy (1974) for an analysis of some small sample properties of Theil's mixed estimator.

Fortunately, modern approximation theory offers a family of functions that not only serves as "the most successful approximating functions in use today,"[3] but which also can be used quite handily in representing prior information (which is likely to be available) concerning the shape of the lag. Members of this family are called spline functions and can be defined as follows.

> Definition. An nth degree polynomial spline function is a piecewise polynomial function made up of polynomials of degree at most n such that the spline and its derivatives up to and including the (n - 1)th are continuous.

A thorough discussion of the approximation properties of spline functions is beyond the scope of this study.[4] Intuitively, spline functions owe much of their approximating power to the discontinuities permitted in their nth derivative. In comparison to polynomials (i.e., Almon lags) which are a subclass of spline functions, this permits spline functions to better approximate lag functions with asymptotes. Since splines are made up of polynomials, they cannot have asymptotes. However, they can have flat segments as will be seen in section 3.

In this study the use of spline functions in lag parameterizations, hereafter referred to as spline lags, will be motivated as follows. Sections 2 and 3 will provide theoretical motivations. Often it is felt a

---

[3] de Boor and Rice (1968, p. 7).

[4] A concise introduction to the use of spline functions in approximation theory can be found in Handscomb (1966), and a thorough discussion of their mathematical properties can be found in Ahlberg, Nilson, and Walsh (1967). As for their use in economics, see Poirier (1973a, 1973b, 1973c, 1973d, 1975).

priori that the lag weights should be on a smooth curve, and section 2 gives an analytical interpretation to the concept of "smoothness," as well as showing that a cubic spline is in a sense the smoothest function. Section 3 weakens the stringent requirement of most lag parameterizations - namely, that the exact lag length be known - and replaces it with a less stringent requirement which in turn leads again to spline lags. In contrast to sections 2 and 3, section 4 provides "approximation" motivations for the use of spline lags by looking at its "robustness" compared to an Almon lag. Section 5 illustrates many of the ideas discussed in the preceding sections by empirical applications of both linear and cubic spline lags. Finally, section 6 provides a summary of results as well as comments on practical implementation.

2. Motivation I: Smoothness

Consider the distributed lag model

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_k x_{t-k} + \varepsilon_t \quad (t = 1, 2, \ldots, T)$$

or

$$y = X\beta + \varepsilon \tag{2.1}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_T \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 & x_0 & \dots x_{1-k} \\ x_2 & x_1 & \dots x_{2-k} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_T & x_{T-1} & \dots x_{T-k} \end{bmatrix}$$

We assume that the x_t's are fixed in repeated sampling, or if stochastic, distributed independently of ε and such that plim $(\frac{1}{T}X'X)$ is positive definite. We also assume that $\varepsilon \sim N(0, \sigma^2 I)$.

Throughout the distributed lag literature, it has been postulated that the lag coefficients should lie on a "smooth" curve and that "wild" changes in magnitude and sign are unlikely.[5] Specifically, Shiller (1973, p. 777) gives the following motivation for smoothness priors.

> "First degree smoothness priors can be likened to a draftsman's flexible curve. The favored shape is a straight line; moreover, the shapes that one can draw with a flexible curve will have small 'second differences' unless the curve is bent very hard."

---

[5] See Sims (1971, 1973) for a counter-argument in terms of discrete approximations to continuous-time distributed lag models. The relationship between the discrete lag distribution and the underlying continuous-time distribution depends on the local serial correlation of the independent variable. For locally smooth independent variables, the graph of the discrete distribution looks very much like the continuous-time distribution. However, the discrete distribution corresponding to a one-sided continuous-time distribution is unlikely to be smooth near a time lag of zero. Hence, this line of reasoning implies that the smooth cubic spline lag described in this section should not include the first couple of lag weights. Otherwise, the suggested techniques are relevant and the example in section 5 will illustrate this. Arguments for unconstrained ordinary least squares are also presented in Cargill and Meyer (1974).

Interestingly, Shiller's analogy to the draftsman's flexible curve immediately brings to mind spline functions since their name in fact stems from their similarity to the draftsman's spline.[6] The widespread popularity of spline functions in approximation theory is in large part due to the following mathematical property.

<u>Theorem 1</u>: Let $t_0 < t_1 < \ldots < t_m (m < k)$ be a set of abscissa values (called knots), and $\beta_0, \beta_1, \ldots, \beta_m$ an associated set of ordinates (lag weights). Then of all functions $f(t) \varepsilon C^2 [t_0, t_m]$ such that $f(t_j) = \beta_j (j = 0, 1, \ldots, m)$, the integral

$$\int_{t_o}^{t_m} [f''(t)]^2 dt \tag{2.2}$$

is minimized when f(t) is a natural cubic spline, i.e., a third degree polynomial spline whose second derivative at t_0 and t_m vanishes.[7]

Using (2.2) as a measure of smoothness, Theorem 1 indicates that the natural cubic spline is the "optimal" lag structure passing through (t_j, β_j) $(j = 0, 1, \ldots, m)$.[8] However, this optimality is conditional on the knot selection. Fortunately, as will be seen in the following section,

---

[6]See Ahlberg, Nilson, and Walsh (1967).

[7]For a proof see Ahlberg, Nilson, and Walsh (1967, pp. 75-77). An application of Theorem 1 in monetary theory is given in Poirier (1973d).

[8]A discussion of the optimality of spline functions in the smoothing of stochastic processes can be found in Kimeldorf and Wahba (1970). Since (2.2) is often a good approximation to the integral of the square of the curvature of the function y = f(t), Theorem 1 is often referred to as the <u>minimum curvative property</u> (see Ahlberg, Nilson, and Walsh (1967)).

the location of knots can often be dictated by theoretical considerations. In any case we delay discussion of knot selection until section 6.

To gain a better understanding of Theorem 1 and the use of cubic spline lags, consider the following.[9] Suppose we have a monthly distributed lag model with a one year (k=12) lag, and that five equally-spaced knots are chosen in "lag time," i.e., at 0, 3, 6, 9, and 12. Then it is possible to write β as a linear combination of the lag weights at the knots, i.e.,

$$\beta = W_1 \gamma, \tag{2.3}$$

where $\gamma = [\beta_0, \beta_3, \beta_6, \beta_9, \beta_{12}]'$ and W_1 (calculated from equation (2.14) in Poirier (1973a, p. 518)) is given by

$$W_1 = \begin{bmatrix} 1.0000 & .0000 & .0000 & .0000 & .0000 \\ .5873 & .5132 & -.1270 & .03175 & -.005291 \\ .2341 & .8915 & -.1587 & .03968 & -.006614 \\ .0000 & 1.0000 & .0000 & .0000 & .0000 \\ -.07804 & .7646 & .3862 & -.08730 & .01455 \\ -.05291 & .3545 & .8042 & -.1270 & .2116 \\ .0000 & .0000 & 1.0000 & .0000 & .0000 \\ .02116 & -.1270 & .8042 & .3545 & -.05291 \\ .01455 & -.08730 & .3862 & .7646 & -.07804 \\ .0000 & .0000 & .0000 & 1.0000 & .0000 \\ -.006614 & .03968 & -.1587 & .8915 & .2341 \\ -.005291 & .03175 & -.1270 & .5132 & .5873 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix}.$$

---

[9] For the remainder of this study we shall refer to a natural cubic spline as simply a cubic spline unless the context indicates otherwise. This convention reflects the strong implications of Theorem 1 favoring the use of a natural cubic spline. For a detailed discussion of other types of cubic splines, which may be of value in particular cases, see Poirier (1973a) and (1973b, Chapter 3).

Substituting (2.3) into (2.1) yields

$$y = (XW_1)\gamma + \varepsilon \qquad (2.4)$$

which is in the form of a standard unconstrained linear model. Theorem 1 tells us that of all possible lag parameterizations in $C^2[0, 12]$ passing through β_0, β_3, β_6, β_9, β_{12}, the cubic spline lag given by (2.3) is the smoothest in the sense of (2.2).

Exactly how the estimation procedure proceeds from here depends on the methodological beliefs of the individual. The strict classical statistician, having incorporated all the prior information he can, would proceed to estimate γ directly from (2.4), and then most likely, by means of hypothesis testing, check to see if his resulting estimates were "consistent" with maintained hypotheses. Statisticians with Bayesian tendencies would most likely use Theil's "mixed" approach discussed earlier or a full-Bayesian approach. In the latter case, like Chetty (1971) and Mouchart and Arsi (1974), but unlike Leamer (1972) and Shiller (1973), a parsimonious parameterization is introduced first to characterize the lag structure, then an informative conjugate prior can be used.

3. Motivation II: Variable Lag Length

In some situations the researcher may not feel that the lag structure need by smooth in the sense of Theorem 1, but only that it be continuous. One such simple formulation would be a continuous piecewise linear representation, i.e., a linear spline. One of the first distributed lag formulations, by Fisher (1937), assumed that the regression coefficients satisfy an arithmetic (linear) progression. While more recent distributed

lag techniques such as the Almon (1965) lag emphasize higher degree polynomials, non-monotonic lag formulations can also be represented by piecewise linear functions. Such formulations have received some recent attention by Sims (1971, p. 562) and Clark (1974). Interestingly, Clark (1974, p. 8) explains his choice by saying that connected broken line distributions are "closer to the present author's prior notions than Lagrange polynomials."

Besides being reasonable and flexible, the linear spline is also particularly useful in cases where the exact lag length is not known, buy only a set of conceivable lag lengths (including the true one) are known. For example, suppose that in a monthly model it is believed that the distributed lag ends at either two, three, or four quarters. Then a linear spline with knots at 0, 3, 6, 9, and 12 months permits (but does not require) an inverted V-shaped structure for lag periods 0-6, and allows for zero lag coefficients from 6-12 or 9-12. Furthermore, it is easy to test for such possible lag lengths.

To see this consider the following. Analogus to (2.3) the monthly lag weights can be expressed as a linear combination of the weights at the knots by

$$\beta = W_2 \gamma$$

where

$$W_2 = \frac{1}{3}\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Then the sets of hypotheses

$$\begin{aligned} &H_0 : \beta_6 = \beta_9 = \beta_{12} = 0 && H_0' : \beta_9 = \beta_{12} = 0 \\ &H_1 : \beta_6 \neq 0 \text{ or } \beta_9 \neq 0 \text{ or } \beta_{12} \neq 0 && H_1' : \beta_9 \neq 0 \text{ or } \beta_{12} \neq 0 \end{aligned} \tag{2.5}$$

test for the lag ending at 6 or 9 months, respectively. Furthermore, the hypotheses $H_0'' : \beta_{12} = 0$ and $H_1'' : \beta_{12} \neq 0$ test whether the lag ends at 12 months.

The importance of not restricting oneself to a fixed lag length when there is some uncertainty about the exact length should not be overlooked. Schmidt and Waud (20, pp. 12-13) have pointed out that the popular Almon lag is quite restrictive even when the true lag length is overspecified, since a polynomial of degree $p < k$ can have a most p zeros in an interval. The use of "endpoint conditions" further restricts the number of zeros in the interval [0, k]. Once again the continuity conditions on all the derivatives of a polynomial rule out flat segments permissible with a linear spline.

A cubic spline can also have flat segments, but the continuity of its first and second derivatives has important implications for the shape of the other segments.[10] For example, a natural cubic spline with knots at 0, 3, 6, 9 and 12, and which has flat segments along the horizontal axis over [6, 12], must be either convex or concave, but not both, over [0, 6]. This is easily seen by recalling that the second derivative of a cubic spline is a linear spline. If such a restriction is undesirable, then it can be weakened by either including more knots in the interval [0, 6], or by using a cubic spline with a left end condition different than a vanishing second derivative.[11]

---

[10] It should be noted that in the case of the cubic spline, hypotheses (2.5) must be augmented by the additional specification that the second derivative vanishes over the interval under consideration. See Poirier (1973a, p. 520) for details.

[11] See Poirier (1973b, Chapter 3).

Table 1 gives the effects of continuity in derivatives for a cubic spline with five segments, having knots at t_0, t_1, ..., t_5, and with a flat fifth segment along zero. Writing the equation of the cubic spline in the i th interval as

$$S(t) = a_i + b_i(t-t_i) + c_i\ (t-t_i)^2 + d_i(t-t_i)^3\ ,\ t_{i-1} \leq t \leq t_i,$$

with interval lengths $h_i = t_i - t_{i-1}$ (i = 1, 2, ..., 5), the coefficients for segments two through five are given in Table 1. The discontinuities in the third derivative imply one additional "degree of freedom" is picked up in each interval as we move away from the flat segment. Not until we are four segments away are the coefficients in some sense "free." If this segment is the first segment (as in this example), then the left end condition of the natural cubic spline will require $c_1 = 0$.

4. A Robustness Comparison of Almon and Spline Lags

The success spline functions have enjoyed in approximation theory, together with the best approximation property given in section 2, lead one to believe that spline lags should be less susceptable to specification error than Almon lags.[12] This section presents evidence which strongly supports this claim.

Suppose we restrict the lag coefficients of (2.1) to satisfy the linear restrictions

$$\beta = R\gamma, \tag{4.1}$$

where R is a (k+1) x m matrix of known constants with rank m, and γ is

---

[12]Section 3 has also illustrated this with respect to lag length.

Table 1

Cubic Spline Coefficients With a Flat Right-End Segment Along Zero

i	a_i	b_i	c_i	d_i
5	0	0	0	0
4	0	0	0	d_4
3	$-d_4h_4^3$	$3d_4h_4^2$	$-3d_4h_4$	d_3
2	$-d_4h_4[h_4^2 + 3h_4h_3 + h_3^2] - d_3h_3^3$	$3[d_4h_4(h_4 + 2h_3) + h_3^2d_3]$	$3h_3(d_3 + d_4h_4)$	d_2

a mxl column vector of unknown parameters. In the case of a cubic or linear spline lag, R is given by the matrices W_1 and W_2 respectively, discussed in sections 2 and 3.[13] In the case of an Almon lag, R may take on various forms, e.g., we may express the β's as a linear combination of ordinate values analogous to the cubic spline, or as a linear combination of conventional polynomial coefficients.[14]

Subject to (4.1) the ordinary least squares estimate of β in (2.1) is

$$\hat{\beta} = R\hat{\gamma} = R\left[(R'X'XR)^{-1}R'X'y\right] = Dy = DX\beta + D\varepsilon$$

where $D = R(R'X'XR)^{-1}R'X'$ and $DD' = R(R'X'XR)^{-1}R'$.

The mean square error matrix of $\hat{\beta}$ is

$$M = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$$= E\left[(DX - I)\beta + D\varepsilon\right]\left[(DX - I)\beta + D\varepsilon\right]'$$

$$= \sigma^2 DD' + (DX - I)\beta\beta'(DX - I)' \qquad (4.2)$$

It would be nice to have a simple scalar function (e.g., trace or determinant) of (4.2), however, (4.2) itself is much too cumbersome to work with. Instead we will follow the strategy of Amemiya and Morimune (1974, p. 379) and consider only the much-simplified function

---

[13] Of course there are other parameterizations available also, e.g., in terms of conventional polynomial coefficients and jump discontinuities in the nth derivative (n = 1 or n = 3) at the interior knots.

[14] See for example Cooper (1972).

$$\mathrm{tr}(MT^{-1}X'X) = \sigma^2T^{-1}\mathrm{tr}(DD'X'X) + \mathrm{tr}\left[(DX - I)\beta\beta'(DX - I)'T^{-1}X'X\right]$$

$$= \sigma^2T^{-1}\mathrm{tr}\left[R(R'X'XR)^{-1}R'X'X\right] + \mathrm{tr}\left[\beta'(DX - I)'T^{-1}X'X(DX - I)\beta\right]$$

$$= \sigma^2T^{-1}(k + 1) + \beta'\left[T^{-1}X'X - T^{-1}X'XR(RT^{-1}X'XR)^{-1}R'T^{-1}X'X\right]\beta. \quad (4.3)$$

One rationale for (4.3) is that it equals T^{-1} times the trace of the mean square error matrix of the estimator of $X\beta$, and hence it measures how well we can estimate the systematic part of the dependent variable.[15]

By computing (4.3) for various choices of R it is possible to compare in a sense the efficienty of an Almon lag with various spline lags. However, before doing this we shall make one more simplification of (4.3), again patterned after the approach of Amemiya and Moumume (1974). Assuming that x_t follows a stationary first-order autoregressive process with

$$E(x_t) = 0 \quad \text{and} \quad E(x_t x_{t+s}) = \frac{\sigma_x^2 \rho^{|s|}}{1-\rho^2},$$

we will take the probability limit of (4.3) to obtain

$$\mathrm{plim}\left[\mathrm{tr}(MT^{-1}X'X)\right] = \sigma_x^2\beta'\left[\Omega - \Omega R(R'\Omega R)^{-1}R'\Omega\right]\beta, \quad (4.4)$$

where

---

[15] Amemiya and Morimune (1974) are concerned with selecting the "optimal" order of the polynomial in an Almon lag. In their discussion they use a different parameterization of the Almon lag than (4.1). They report that they computed the trace of (4.2) for certain parameter values and that their results did not differ significantly from those using (4.3).

$$\Omega = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^k \\ \rho & 1 & \cdots & \rho^{k-1} \\ \vdots & \vdots & & \vdots \\ \rho^k & \rho^{k-1} & \cdots & 1 \end{bmatrix} .$$

Now letting R = A represent the Almon lag (AL) case, and R = S represent the spline lag (SL) case, we compute (4.4) for each case and form the ratio[16]

$$\lambda = \frac{\text{plim}[\text{tr}(M_{SL}T^{-1}X'X)]}{\text{plim}[\text{tr}(M_{AL}T^{-1}X'X)]} = \frac{\beta'[\Omega - \Omega S(S'\Omega S)^{-1}S'\Omega]\beta}{\beta'[\Omega - \Omega R(R'\Omega R)^{-1}R'\Omega]\beta} . \qquad (4.5)$$

---

[16] In their analysis, Amemiya and Morimune (1974) replace the second term of (4.3) with its probability limit, leaving the first term untouched. The resulting expression for efficiency can then be regarded as the asymptotic expectation of the essential part of the usual mean square prediction error (see Amemiya and Morimune (1974, p. 379)). If this procedure had been followed here, than instead of then instead of (4.5) we would have

$$\frac{\theta + \beta'[\Omega - \Omega S(S'\Omega S)^{-1}S'\Omega]\beta}{\theta + \beta'\ \Omega - \Omega A(A'\Omega A)^{-1}A'\Omega\ \beta}$$

where $\theta = (k + 1)\ \sigma^2/T\sigma_x^2$. In this context (4.5) is merely the limiting case as $\theta \to 0$. While under this measure the relative efficiencies of the two types of lags will tend to unity as θ increases, the difference in their absolute efficiencies will still remain constant.

In order to make meaningful statements concerning (4.5) in light of its many unknown parameters, we will proceed as follows. First, we will use the twelve sets of lag weights (β's) found in Amemiya and Morimune (1974, p. 380) and the three sets of weights found in Cargill and Meyer (1974 pp. 1036-1037).[17] Graphs (the ordering of which will become clear later) are shown in Figure 1. Graphs (a), (c), (f), (i), (j), (m), and (o) correspond to estimates from various empirical studies and come from Amemiya and Morimune (1974).[18] Graphs (b), (e), (k), (l), and (n) are also taken from Amemiya and Morimune (1974) and they correspond to second-order autoregressive models with different parameter values. Graphs (d), (g), and (h) in Figure 1 are taken from Cargill and Meyer (1974).

Second, we will choose k = 15. This corresponds to the maximum lag length of these beta sets. Since nine of the fifteen beta sets have lag lengths shorter than fifteen periods (e.g., the weights shown in graph (h) of Figure 1 have a lag length of five), this specification will permit a comparison of robustness with respect to lag length misspecification.

Third, we will choose Almon and spline lags with m = 4 unknown parameters. In the first case this corresponds to a cubic Almon lag — a choice which is likely to be representative of many of the applications found in the literature. In the second case this corresponds to a spline lag with four equally-spaced knots at 0, 5, 10, and 15.[19] Note that in

---

[17]The third set of weights taken from Cargill and Meyer (1974, p. 1037) correspond to their Model 3. This model contains distributed lags in two different explanatory variables, but only those weights corresponding to the second explanatory variable are used here.

[18]The sources for these seven sets of lag weights are given in Amemiya and Morimune (1974). None of these empirical studies used either Almon or spline lags.

[19]For the cubic spline lag the W-matrix was computed using equation (2.14) in Poirier (1973a, p.518) and is given in the Appendix.

Figure 1

Lag Weights

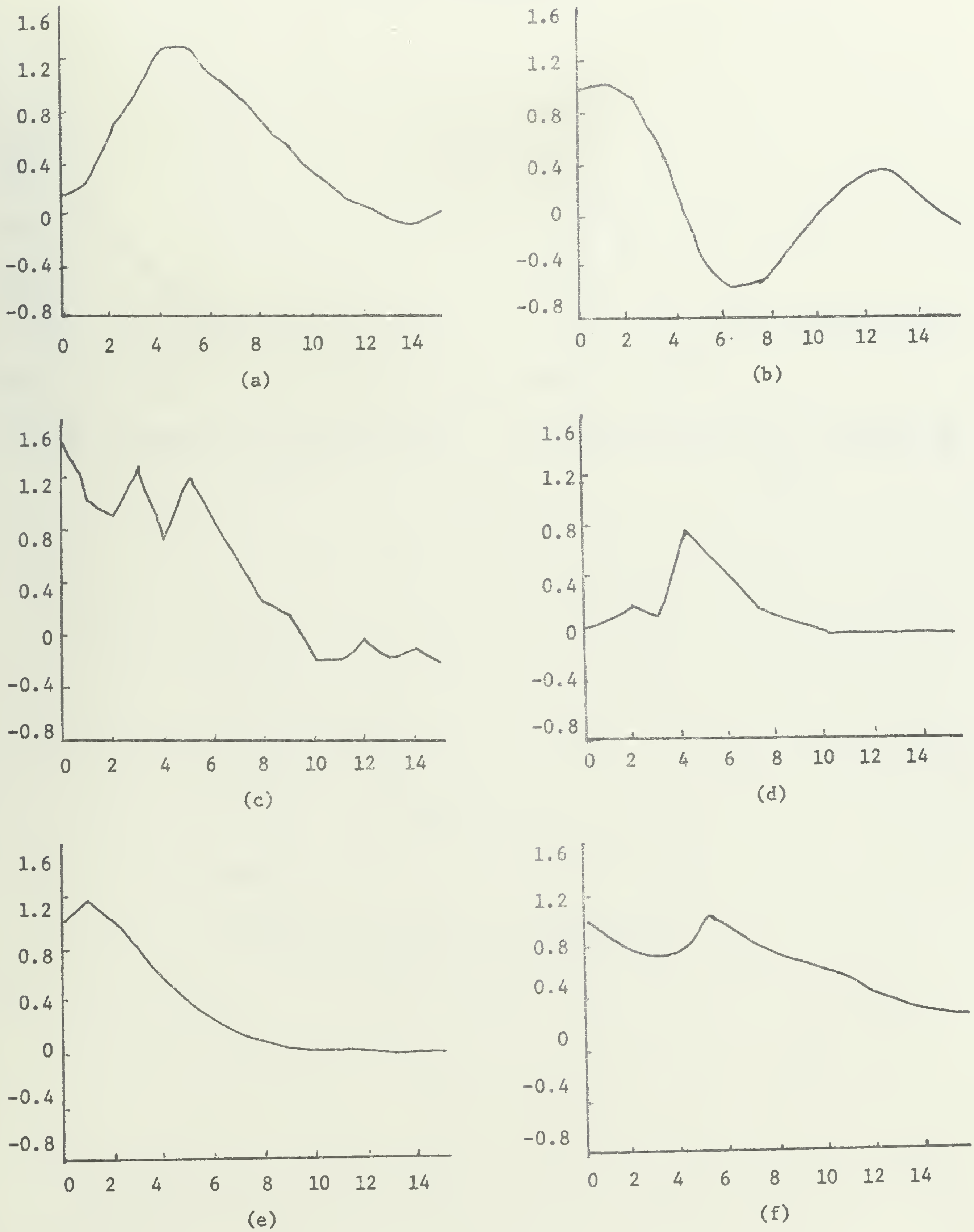


Figure 1 (continued)

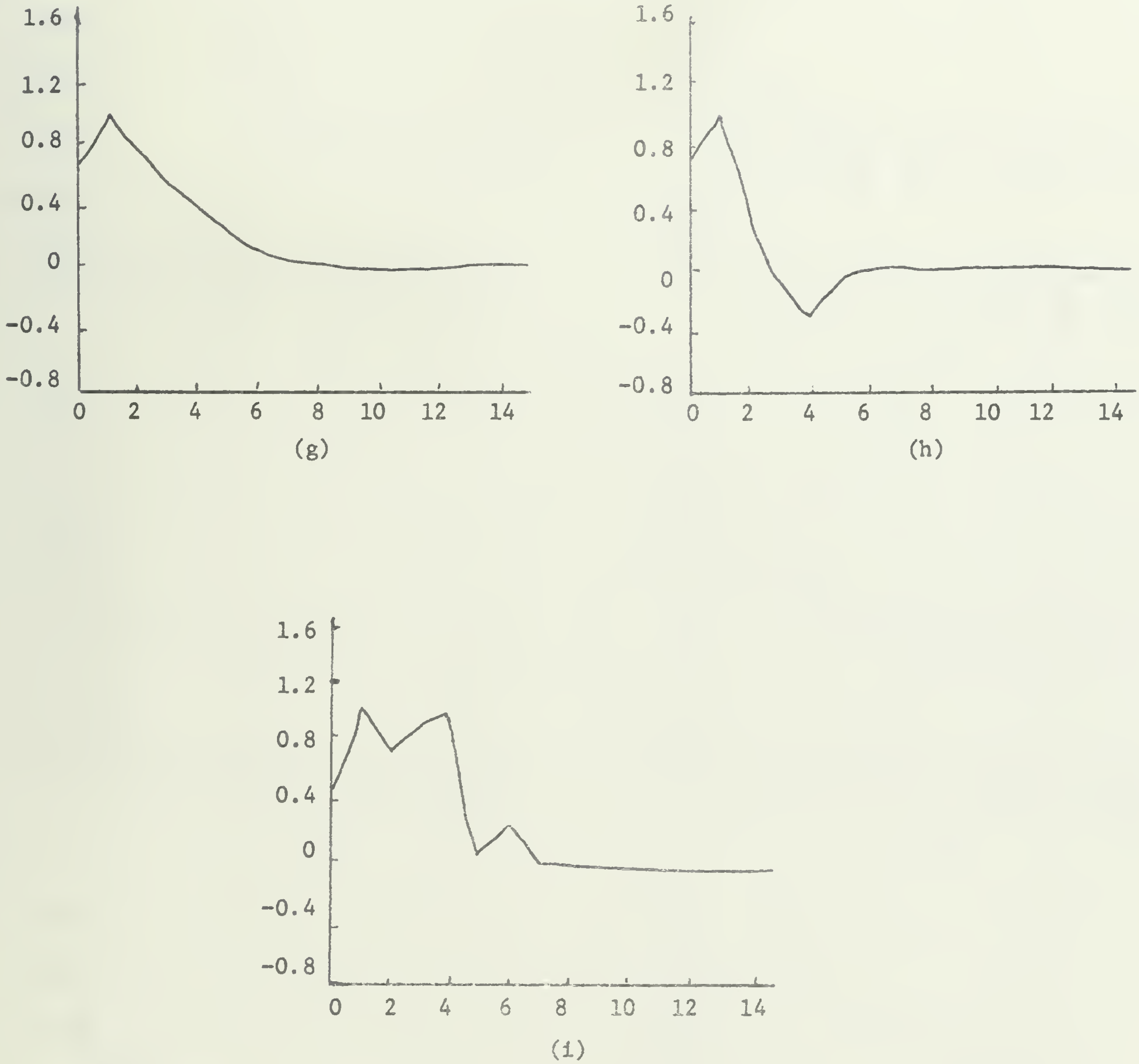
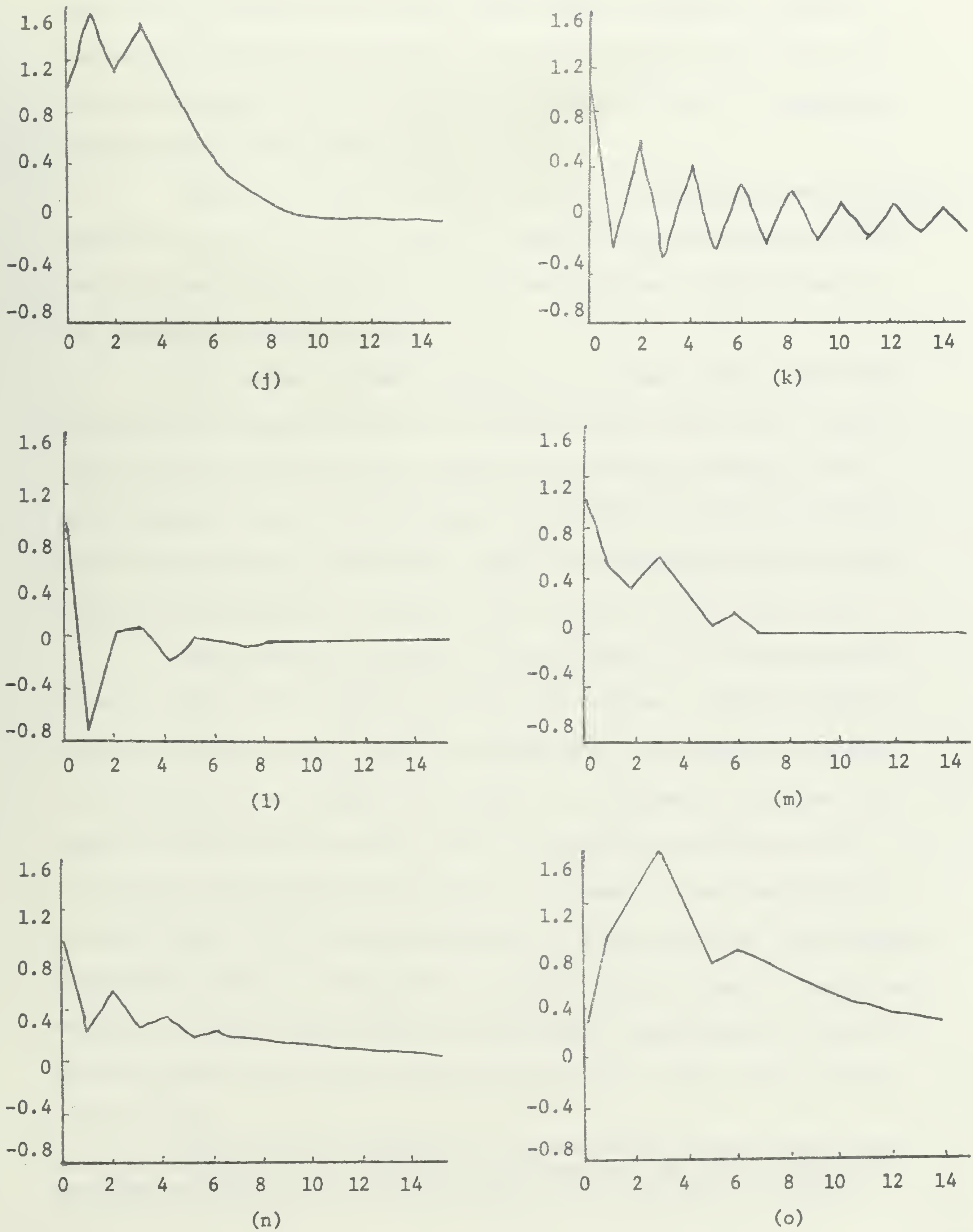


Figure 1 (continued)



order not to prejudice the results, these knots have not been chosen on the basis of possible lag lengths. To the extent that such choices can be feasibly made in pratice, the results presented here will understate the efficiency of the spline lag.

Fourth and finally, ratio (4.5) will be computed for the following values of ρ : .1, .3, .5, .7, and .9. These choices for ρ permit investigation of the effect on the two lag parameterizations of smoothness in the explanatory variable series.

The results of these computations are summarized in Table 2 for the cubic spline lag and in Table 3 for the linear spline lag. As can be seen from these tables, on the average the spline lags did quite well — with efficiency gains of approximately 9 percent for the cubic spline lag and 8 percent for the linear spline lag.[20] The efficiency gains for the spline lags are greatest (for most of the beta sets) when $.5 \leq \rho \leq .7$.

Unfortunately, the effect of the "nature" of the lag shape on λ is not so clear. For convenience the graphs have been ordered in such a way that proceeding from (a) to (o) corresponds to decreasing efficiency of the cubic spline. Graphs (a) through (i) correspond to the nine lag weight sets for which the cubic spline lag dominated the Almon lag (on the average), and graphs (j) through (o) correspond to the remaining six lag weight sets. The relative dominance was the same for the linear spline lag except in case (i). Basically, it seems (to this author at least) that in those lag shapes which seem most "reasonable," especially (a), (e), and (g), the spline lags do better than the Almon lag. Those cases in which

[20] Of course the relevance of these averages depends upon whether the equal weights each of the fifteen beta sets and the five values of ρ are given are indicative of their frequency of occurrence in practice.

Table 2

λ - Values for the Cubic Spline Lag

Beta set	ρ = .1	ρ = .3	ρ = .5	ρ = .7	ρ = .9	Column mean
(a)	.456	.441	.428	.416	.406	.429
(b)	.712	.709	.709	.713	.720	.713
(c)	.800	.750	.713	.694	.692	.730
(d)	.779	.748	.725	.712	.707	.734
(e)	.850	.821	.796	.781	.779	.805
(f)	.853	.835	.819	.809	.809	.825
(g)	.930	.901	.878	.865	.864	.888
(h)	.931	.951	.974	.990	.997	.969
(i)	1.02	1.00	.986	.975	.973	.990
(j)	1.03	1.02	1.00	.987	.983	1.00
(k)	1.04	1.05	1.06	1.07	1.08	1.06
(l)	1.06	1.06	1.06	1.06	1.07	1.06
(m)	1.13	1.13	1.14	1.15	1.17	1.14
(n)	1.11	1.13	1.14	1.16	1.18	1.14
(o)	1.21	1.20	1.19	1.19	1.19	1.20
Row mean	.928	.916	.908	.905	.907	.913

Table 3

λ - Values for the Linear Spline Lag

Beta Set	ρ = .1	ρ = .3	ρ = .5	ρ = .7	ρ = .9	Column mean
(a)	.527	.519	.516	.515	.514	.518
(b)	.650	.660	.674	.689	.699	.675
(c)	.614	.544	.494	.469	.465	.517
(d)	.600	.546	.511	.493	.485	.527
(e)	.833	.806	.793	.791	.796	.804
(f)	.689	.679	.676	.681	.692	.683
(g)	.907	.877	.860	.857	.865	.873
(h)	.954	.974	.991	1.00	1.00	.985
(i)	1.06	1.05	1.04	1.04	1.04	1.05
(j)	1.18	1.17	1.16	1.14	1.13	1.16
(k)	1.05	1.06	1.08	1.10	1.12	1.08
(l)	1.09	1.09	1.09	1.10	1.11	1.10
(m)	1.21	1.23	1.27	1.30	1.34	1.27
(n)	1.17	1.20	1.22	1.25	1.28	1.22
(o)	1.44	1.41	1.38	1.36	1.35	1.39
Row mean	.932	.921	.917	.919	.926	.923

the Almon lag dominates (i.e., (j) through (o)) tend to have large jumps in weights between lag periods zero and one.[21] Since a natural cubic spline requires the second derivative to vanish at zero, this smoothness requirement may in part explain its relative inefficiency, and hence give at support to the procedure of not including the weight at zero in the lag parameterization.

Exactly how far the successful results (in terms of the spline lags) of this section can be generalized is of course not clear. The particular lag wieght sets were chosen because (i) they had been used in recent studies for choosing and comparing lag parameterizations, (ii) they cover an extremely wide range of lag shapes, and (iii) they did not seem to in any way prejudice the results beforehand. As in any investigation of this kind, generalizations can be dangerous — but hopefully no more so than those that may come from Amemiya and Morimune (1974) and Cargill and Meyer (1974).

---

[21] Recall footnote 5.

5. Empirical Illustration

To illustrate the practical application of spline lags, the following empirical illustration is presented. Shiller (1973, p. 783) discusses a model in which the corporate bond yield average is treated as a distributed lag in the commercial paper rate. There exists substantial literature (see e.g., Modigliani and Shiller (1973)) in which the theoretical model relating the long rate to the short rate is developed. Typically the lag coefficients are supposed to reflect an expectational mechanism, and hence, for the most part are assumed to lie on a smooth curve - except for β_0 (see footnote 5).

Along these lines, model (2.1) was postulated with $k = 18$ and a first-order autoregressive disturbance term. The dependent variable was the Federal Reserve Board Corporate Aaa yield series and the explanatory variable was the four to six month prime commercial paper rate, with $t = 1$ corresponding to the third quarter, 1955 and $T = 48$ corresponding to the fourth quarter, 1969.[22]

Based on many similar models contained on Modigliani and Shiller (1973), it was expected a priori that the lag shape would have the following characteristics. Beginning at a lag of one period the weights would increase rapidly from negative to positive reaching a maximum near a lag of six or seven quarters, and then declining with an

---

[22] While it was initially the intention to replicate Shiller's example, a comparison of unconstrained ordinary least squares results for identical models indicated that this was not possible. Since at exact replication of Shiller's results was not possible, we have (unlike Shiller) used the Prais-Winsten two-step procedure to correct for first-order autoregression in the disturbance term. The data used in this study were taken from the Federal Reserve Board Bulletin over the period 1953-1970.

inflection point near a lag of nine quarters and possibly a positive asymptotic runout beginning at about a lag of twelve quarters.

These a priori beliefs were utilized in picking the knots of the following two spline lags - each with four unknown parameters. First, following the general suggestions of Wold (1974), a cubic spline lag with knots at 1, 5, 9, and 18 was chosen. This knot selection reflects Wold's suggestions to have extreme points near the center of intervals and inflection points near knots. Second, a linear spline lag with knots at 1, 6, 12, and 18 was also chosen. This knot selection reflects the expected maximum near 6 and the possible asymptote-like behavior beyond 12.

The undesirability of unconstrained ordinary least squares is clearly reflected in Figure 2 which shows the estimated lag coefficients from such a procedure. Not only is the lag shape highly irregular, but the standard errors of all lag coefficients are quite large. In contrast the spline lags yield quite acceptable results. Letting the coefficient of the lag at zero be "free" and constraining all other coefficients to lie on the previously mentioned spline lags resulted in the estimated coefficients and statistics in Table 4. (γ_1, γ_2, γ_3 and γ_4 are the lag weights at the respective knots.) The constrained lag coefficients are shown in Figure 3 for the cubic spline lag and in Figure 4 for the linear spline lag. As can be seen from these figures, both spline lags give intuitively pleasing and surprisingly similiar results.[23]

---

[23]The expected asymptotic behavior did not materialize. For the linear spline lag the null hypothesis H_0: $\gamma_3 = \gamma_4$ was rejected in favor of H_1: $\gamma_3 \neq \gamma_4$ at the five percent level - indicating that the final segment is indeed significantly idfferent than a horizontal line.

Figure 2

Unconstrained Ordinary Least Squares

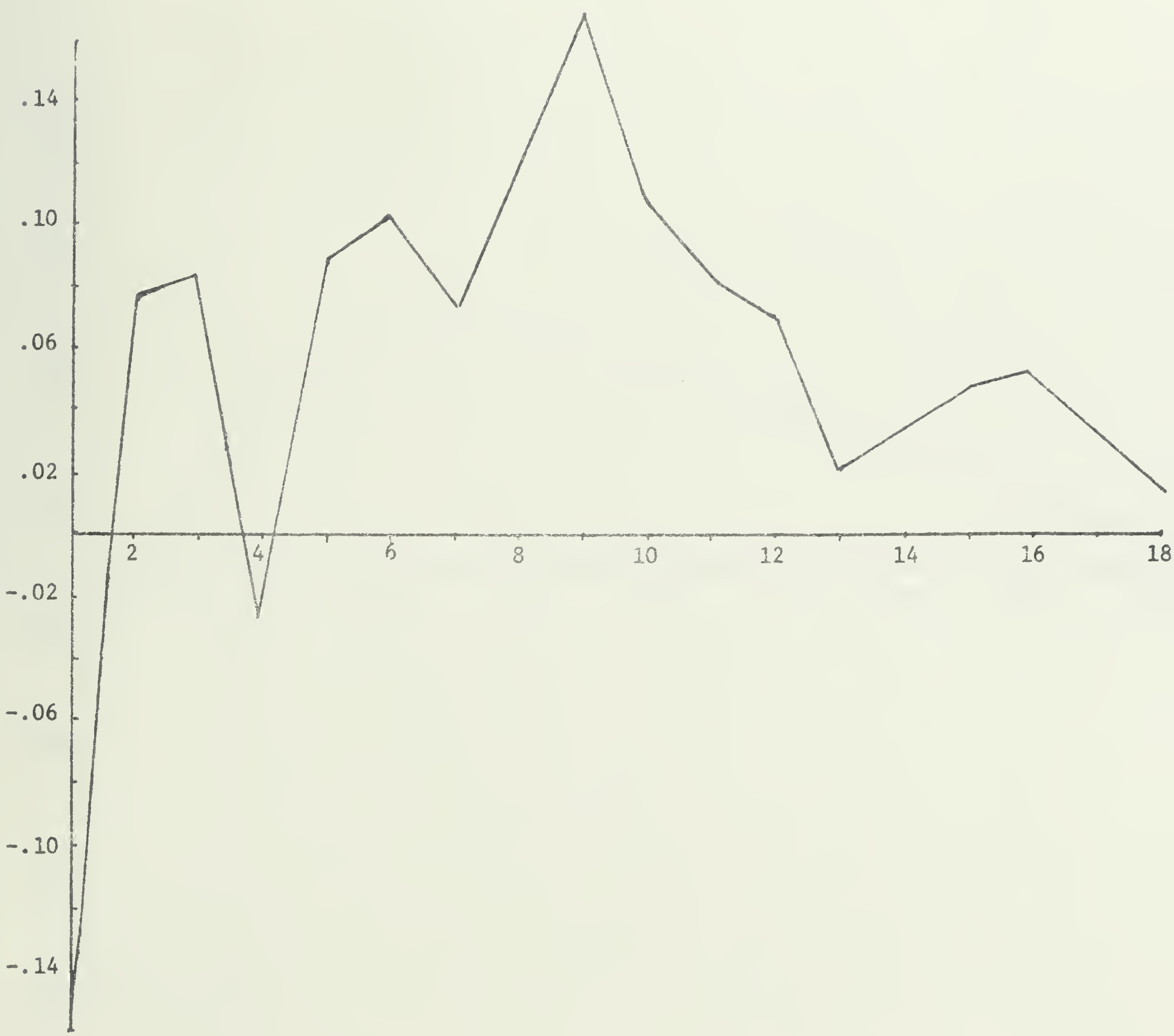


Figure 3

Cubic Spline Lag

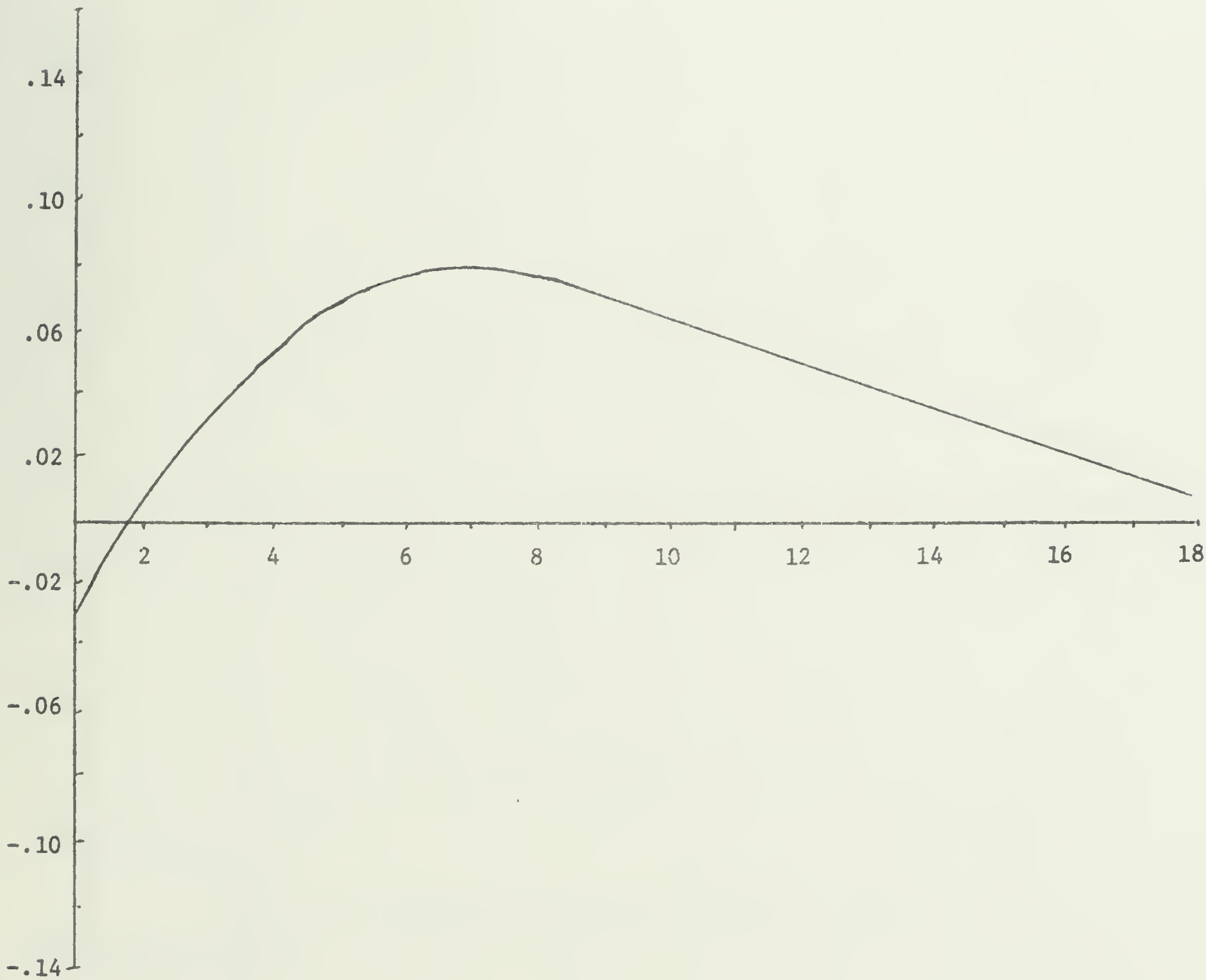
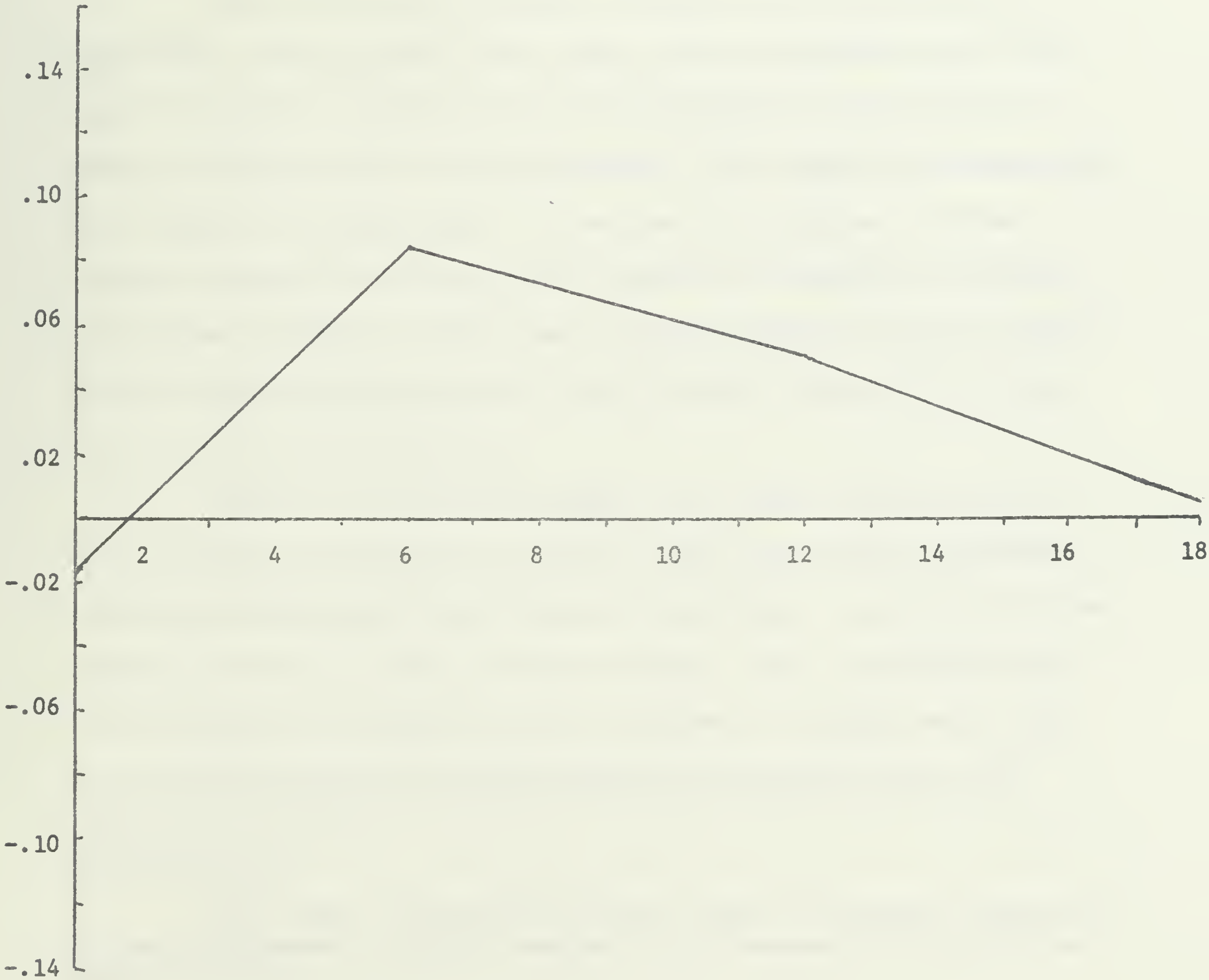


Figure 4

Linear Spline Lag



6. Conclusion

The intention of this study has been to address itself to only one aspect of the distributed lag model - the parameterization of the lag weights. While the discussion has neglected complications in the error term, the presence of other explanatory variables, and simultaneous equation problems, their introduction into the model would not appear to negate the following three conclusions.

First, if a smooth lag shape is desired over some interval of lagged time, then the natural cubic spline gives the smoothest lag in the sense of Theorem 1. Depending on the methodological beliefs of the individual researcher, such a parameterization can be imposed nonstochastically (i.e., exact) or stochastically. In the case of the latter, a "mixed estimator" analogue of the cubic spline lag can be formulated in much the same fashion as Taylor's (1974) mixed estimator analogue of Shiller's lag. The full-fledged Bayesian approach (see for example Zellner (1971)) is an even more attractive alternative.

Second, if the exact lag length is not known, but rather only a set of possible lag ending points, then a linear spline lag (or possibly a cubic spline lag) may be used to test for lag length - assuming of course that the true lag is a linear (or cubic) spline. Such a procedure allows for the testing of lag length within the context of a single model, which is more desirable than the "fishing expedition" typically used in the applied literature.[24]

---

[24] See Cohen, Gillingham, and Heien (1973) for evidence relating to the poor performance of typical goodness of fit measures for detecting lag length and shape.

Third, based on the limited investigation of section 4, there is an indication that spline lags are more robust than Almon lags. At the very least the results must be viewed as encouraging for the spline lags.

While the preceding results are encouraging, no pretense is made that they are beyond criticism. Indeed it seems that much of the distributed lag literature is quick to criticize the techniques of others, but slow to suggest positive solutions. Probably, the aspect of spline lags most susceptible to criticism is the question of knot selection. In his previous writings on spline functions, the author has emphasized selecting knot locations based on theoretical considerations. Indeed, selection based on possible lag lengths does precisely that.[25] However, this is not intended to rule out other approaches. The use of Wold's "rules of thumb" in section 5 is one such alternative. Similarly, the numerous variable knot techniques to be discussed in Poirier (1975) are other candidates. Of these, Bayesian approaches again seem most attractive,[26] if for no other reason than that informative priors force researchers into potentially fruitful thought concerning the model being investigated. Spline lags are not immune from the abuse of users, and potential fishing expeditions for knot locations can only be discouraged - not prevented.

---

[25] The selection of a knot reflecting the ending of the lag is an extreme example in which two adjacent pieces are thought to be different. Past works of the author have emphasized the testing of such "structural changes," and it remains to be seen whether such tests are of interest in the context of distributed lags.

[26] See also Halpern (1973).

In closing it may be appropriate to recall the remarks of Leamer (1972, p. 1080): "In the final analysis, choice among methods will come down to 'which bed is more comfortable,' rather than 'which bed has clean linen.'" Paraphrasing Leamer, the author finds bed 1 (Almon lags) terribly uncomfortable, but sleeps blissfully on bed 2 (spline lags), perhaps only because it is not crowded.

6. Appendix

The transformation matrix for the natural cubic spline with knots at 0, 5, 10, and 15, which was used in sections 4, is given by:

$$W = \begin{bmatrix} 1.000 & .0000 & .0000 & .0000 \\ .7488 & .3152 & -.07680 & .01280 \\ .5104 & .6016 & -.1344 & .02240 \\ .2976 & .8304 & -.1536 & .02560 \\ .1232 & .9728 & -.1152 & .01920 \\ .0000 & 1.000 & .0000 & .0000 \\ -.06400 & .8960 & .2000 & -.03200 \\ -.08000 & .6960 & .4480 & -.06400 \\ -.06400 & .4480 & .6960 & -.08000 \\ -.03200 & .2000 & .8960 & -.06400 \\ .0000 & .0000 & 1.000 & .0000 \\ .01920 & -.1152 & .9728 & .1232 \\ .02560 & -.1536 & .8304 & .2976 \\ .02240 & -.1344 & .6016 & .5104 \\ .01280 & -.07680 & .3152 & .7488 \\ .0000 & .0000 & .0000 & 1.000 \end{bmatrix}$$

The transformation matrix for the natural cubic spline with knots at 1, 5, 9, and 18, which was used in section 5, is given by:

$$W = \begin{bmatrix} 1.000 & .0000 & .0000 & .0000 \\ .6891 & .3813 & -.07448 & .004167 \\ .4025 & .7100 & -.1192 & .006667 \\ .1647 & .9338 & -.1043 & .005833 \\ .0000 & 1.000 & .0000 & .0000 \\ -.07594 & .8775 & .2093 & -.01083 \\ -.08250 & .6200 & .4825 & -.02000 \\ -.04781 & .3025 & .7645 & -.01917 \\ .0000 & .0000 & 1.000 & .0000 \\ .03778 & -.2267 & 1.145 & .04395 \\ .06222 & -.3733 & 1.200 & .1116 \\ .07500 & -.4500 & 1.175 & .2000 \\ .07778 & -.4667 & 1.083 & .3062 \\ .07222 & -.4333 & .9340 & .4272 \\ .06000 & -.3600 & .7400 & .5600 \\ .04278 & -.2567 & .5122 & .7017 \\ .02222 & -.1333 & .2617 & .8494 \\ .0000 & .0000 & .0000 & 1.000 \end{bmatrix}$$

References

Ahlberg, J. H., E. N. Nilson and J. L. Walsh, 1967, The theory of splines and their applications (Academic Press, Inc., New York).

Almon, S., 1965, The distributed lag between capital appropriations and expenditures, Econometrica 33, 178-196.

Amemiya, T. and K. Morimune, 1974, Selecting the optimal order of polynomial in the Almon distributed lag, Review of Economics and Statistics 56, 378-386.

Cargill, T. F. and R. A. Meyer, 1974, Some time and frequency domain distributed lag estimators: A comparative Monte Carlo study, Econometrica 42, 1031-1044.

Chetty, V. K., 1971, Estimation of Solow's distributed lag models, Econometrica 39, 99-117.

Clark, P. K., 1974, Operational time and seasonality in distributed lag estimation, manuscript 32 (National Bureau of Economic Research, Cambridge).

Cohen, M., R. Gillingham and D. Heien, 1973, A Monte Carlo study of complex finite distributed lag structures, Annals of Economic and Social Measurement 2, 53-63.

Cooper, J. P., 1972, Two approaches to polynomial distributed lags estimation: An expository note and comment, American Statistician, 32-35.

de Boor, C. E. and J. R. Rice, 1968, Least squares cubic spline approximation I - fixed knots, manuscript CSD TR 20 (Computer Science Department Division of Mathematical Sciences, Purdue University).

Dhrymes, P. J., 1971, Distributed lags: Problems of estimation and formulation (Holden-Day, San Francisco).

Fisher, I., 1937, Note on a short-cut method for calculating distributed lags, Bulletin de l'Institute International de Statistique 29, 323-327.

Griliches, Z., 1967, Distributed lags: A survey, Econometrica 35, 16-49.

Halpern, E. F., 1973, Bayesian spline regression when the number of knots is unknown, Journal of the Royal Statistical Society, Ser. B XXXV, no. 2, 347-360.

Handscomb, D. C., 1966, Spline functions, in: D. C. Handscomb, ed., Methods of Numerical Approximation (Pergamon Press, Oxford).

Kimeldorf, G. S. and G. Wahba, 1970, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, Annals of Mathematical Statistics 55, 371-380.

Leamer, E. E., 1972, A class of informative priors and distributed lag analysis, Econometrica 40, 1059-1081.

Leamer, E. E., 1973, Multicollinearity: A Bayesian interpretation, Review of economics and statistics 55, 371-380.

Mehta, J. S. and P. A. V. B. Swamy, 1974, The exact finite sample distribution of Theil's compatability test statistic and its application, Journal of the American Statistical Association 69, 154-158.

Modigliani, F. and R. J. Shiller, 1973, Inflation, rational expectations and the term structure of interest of rates, Economica , 12-43.

Mouchart, M. and R. Orsi, 1974, Polynomial approximation of distributed lags and linear restrictions: A Bayesian approach, manuscript 7411 (CORE Discussion Paper, Louvain).

Nerlove, M., 1971, On lags in economic behavior, Econometrica 40, 221-252.

Poirier, D. J., 1973a, Piecewise regression using cubic splines, Journal of the American Statistical Association 68, 515-524.

Poirier, D. J., 1973b, Applications of spline functions in economics, Ph.D. Dissertation (Department of Economics, University of Wisconsin at Madison).

Poirier, D. J., 1973c, Asymmetrical partial adjustment models, presented at the Winter Meetings of the Econometric Society (New York).

Poirier, D. J., 1973, An optimal growth path for the money supply subject to target constraints, manuscript (College of Commerce and Business Administration Faculty Working Paper, University of Illinois at Urbana-Champaign).

Poirier, D. J., 1975, The econometrics of structural change with special emphasis on spline functions (North-Holland Publishing Co., Amsterdam), in final preparation.

Schmidt, P. and R. N. Waud, 1973, Almon lag Technique and the monetary versus fiscal policy debate, Journal of the American Statistical Association 68, 11-19.

Shiller, R. J., 1973, A distributed lag estimator derived from smoothness priors, Econometrica 41, 775-788.

Sims, C. A., 1971, Discrete approximations to continuous time distributed lags in econometrics, Econometrica 39, 545-563.

Sims, C. A., 1973, Distributed lags, manuscript 28 (Department of Economics, University of Minnesota).

Taylor, W. E., 1974, Smoothness priors and stochastic prior restrictions in distributed lag estimation, International Economic Review 15, 803-804.

Theil, H., 1963, On the use of incomplete prior information in regression analysis, Journal of the American Statistical Association 58, 401-414.

Theil, H. and A. S. Goldberger, 1961, On pure and mixed statistical estimation in economics, International Economic Review 2, 65-78.

Wold, S., 1974, Spline functions in data analysis, Technometrics XVI, 1-11.

Yancey, T. A., M. E. Bock, and G. G. Judge, 1972, Some finite sample results for Theil's mixed regression estimator, Journal of the American Statistical Association 67, 176-179.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (John Wiley and Sons, Inc., New York).