ESTIMATION OF PARAMETERS IN DISTRIBUTED LAG ECONOMETRIC MODELS

William S. Cleveland
Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 792

December, 1971

ESTIMATION OF PARAMETERS IN DISTRIBUTED LAG ECONOMETRIC MODELS

William S. Cleveland
University of North Carolina at Chapel Hill

ABSTRACT

This paper presents a method for estimating parameters in distributed lag econometric models. A subset of the parameter space is studied which includes the maximum likelihood estimate (which incorporates no smoothing assumption on the parameters), an extreme smoothness estimate, and other values which represent an amount of smoothing ranging between that of the two extremes. From the subset an estimate is chosen which is not judged unlikely by the data but yet has an amount of smoothness in the coefficients which makes economic sense.

1. INTRODUCTION

One very common occurrence in fitting distributed lag models to economic data is that the least squares estimates of the parameters do not make economic sense, since in general the exogenous variables and their lagged values are so highly correlated and the number of observations so small that there is little information in the data about the true values of the parameters. One remedy is to eliminate variables from the model. This is an adequate procedure if prediction is the only goal, but it does not help the person interested in understanding the economic system generating the data or the person who wants to make policy decisions.

In this paper, a method of estimating parameters in distributed lag models is presented. In Section 2 notation will be established and one reason for the lack of precision in the least squares estimates given. In Section 3 a general approach to estimating parameters in distributed lag models is presented and two specific implementations are discussed. Section 4 gives a Bayesian interpretation of the two implementations. In Section 5 the parameters of two distributed lag equations from the St. Louis Federal Reserve Bank econometric model of the United States economy are estimated using the methods suggested in Section 3. In Section 6 several comments are made about the methods and a comparison made with the Almon lag technique.

2. THE MODEL AND THE PROBLEM

To keep notation from getting out of hand, it will be supposed that there are two exogenous variables a_t and b_t. The discussion that follows holds with obvious changes in the details if there are more than two. The distributed lag model relating a_t and b_t to the endogenous variable y_t is

$$y_t = \gamma + \sum_{j=0}^{a} \alpha_j a_{t-j} + \sum_{j=0}^{b} \beta_j b_{t-j} + z_t,$$

for $t = 1,\ldots,N$, where z_t is a sequence of independent normal random variables with mean 0 and variance $e^{-\phi}$ (this parameterization of the variance rather than the usual σ^2 will prove convenient later) and γ is a constant term. These N equations will now be written in matrix form. Let Y, Z, and θ be the column vectors

$$Y = (y_1,\ldots,y_N)'$$

$$Z = (z_1,\ldots,z_N)'$$

and

$$\theta = (\alpha_0,\ldots,\alpha_a, \beta_0,\ldots,\beta_b)'.$$

For $k = 1,\ldots,a+1$ and $t = 1,\ldots,N$, let $d_{tk} = a_{t-(k-1)}$. For $k = a+2,\ldots,a+b+2$ and $t = 1,\ldots,N$, let $d_{tk} = b_{t-(a+2-k)}$. Finally, let D be the $N \times (a+b+2)$ matrix

$$D = [d_{tk}]_{t=1,\ldots,N;k=1,\ldots,a+b+2}$$

and let 0 be a column vector of N ones. Then the distributed lag model is

$$Y = 0\gamma + D\theta + Z.$$

One way of reducing the degree of correlation between the variables in the model is to reparameterize in a way that does not change θ and consists of substracting from each column of D its mean. Let $d_{\cdot k} = N^{-1} \sum_{t=1}^{N} d_{tk}$ and $x_{tk} = d_{tk} - d_{\cdot k}$. Let X be the $N \times (a+b+2)$ matrix

$$X = [x_{tk}]_{t=1,\ldots,N;k=1,\ldots,a+b+2}.$$

Then the distributed lag model may be written as

$$Y = O\mu + X\theta + Z$$

where

$$\mu = \gamma + \sum_{k=1}^{a+b+2} \theta_k d_{\cdot k}.$$

The correlation between O and each column of X is zero.

The maximum likelihood estimate of θ is

$$\hat{\theta}_{MLE} = (X'X)^{-1} X'Y.$$

Let

$$\bar{Y} = N^{-1} \sum_{t=1}^{N} y_t$$

and

$$SS(\dot{\theta}) = (Y-O\bar{Y}-X\dot{\theta})' (Y-O\bar{Y}-X\dot{\theta}).$$

The likelihood function given $\theta = \dot{\theta}$, $\mu = \dot{\mu}$, and $\phi = \dot{\phi}$ is

$$L(\dot{\theta},\dot{\mu},\dot{\phi}) \propto e^{\frac{N\dot{\phi}}{2}} \exp -\tfrac{1}{2}e^{-\dot{\phi}} [SS(\hat{\theta}_{MLE}) + N(\mu-\bar{Y})^2 + (\dot{\theta}-\hat{\theta}_{MLE})' X'X(\dot{\theta}-\hat{\theta}_{MLE})].$$

Since we want to focus on the problem of estimating θ we will study the marginal likelihood function of $\dot{\theta}$

$$L(\dot{\theta}) \propto \int L(\dot{\theta},\dot{\mu},\dot{\phi})\, d\dot{\mu}\, d\dot{\phi}$$

$$\propto [SS(\dot{\theta})]^{-\frac{N-1}{2}}$$

$$= [SS(\hat{\theta}_{MLE}) + (\dot{\theta}-\hat{\theta}_{MLE})' X'X(\dot{\theta}-\hat{\theta}_{MLE})]^{-\frac{N-1}{2}}.$$

The marginal likelihood L is proportional to a multivariate Student density function [6, p. 256]. The maximum occurs at $\dot{\theta} = \hat{\theta}_{MLE}$ and the function is constant on each ellipsoid of the form

$$(\dot{\theta}-\hat{\theta}_{MLE})' X'X(\dot{\theta}-\hat{\theta}_{MLE}) = \text{constant}.$$

That is, this family of homothetic ellipsoids is the set of contours of L.

The maximum likelihood estimate $\hat{\theta}_{MLE}$ is, on the basis of the data alone, a good candidate as a point estimate of θ since it is in some sense the most likely value. But a value $\dot{\theta}$ with high relative likelihood (i.e., for which $L(\dot{\theta}) \div L(\hat{\theta}_{MLE})$ is not far from 1) is nearly as good a candidate as $\hat{\theta}_{MLE}$. Economic data very often have the two properties that the columns of X are highly correlated and the value of N, the number of observations, is small. Both properties tend to produce a likelihood function which cannot distinguish with much precision between values of $\dot{\theta}$ which have very different economic meanings. That is, they tend to produce a large region of good candidates containing values of $\dot{\theta}$ which have very different economic meanings. The smaller the value of N, the slower the relative likelihood

$$RL(\dot{\theta}) = L(\dot{\theta}) \div L(\hat{\theta}_{MLE})$$

tends to decrease as $\dot{\theta}$ moves in any direction away from $\hat{\theta}_{MLE}$. The more the amount of correlation between the columns of X the more elongated the ellipsoidal contours of L tend to be. Thus as $\dot{\theta}$ moves away from $\hat{\theta}_{MLE}$ in the direction of the major axis of these contours, $RL(\dot{\theta})$ tends to decrease slowly. Thus the data alone do not provide enough information to enable one to make a precise specification of the parameters in the model.

3. A GENERAL REMEDY FOR THE PROBLEM AND TWO SPECIFIC IMPLEMENTATIONS

The one thing that is clear from the last section is that the problem cannot be resolved by statistics alone. There is not sufficient precision in the data to be able to do this. A remedy that works must involve bringing in additional knowledge about the economic system generating the data. One way of doing this would be to simply look over the region of $\overset{\circ}{\theta}$ values with high relative likelihood $RL(\overset{\circ}{\theta})$ and pick out one that made sense with regard to all the prior knowledge about the economic system generating the data. But this procedure generally would not be feasible to carry out since RL is typically a function of a large number of variables and difficult to understand. Thus it would be difficult to codify the prior knowledge about the economic system in this way.

The solution is to settle for something a bit less. Instead of studying RL over the whole parameter space, the solution is to study it over a subset that makes sense for the distributed lag model. In the typical situation it is sensible to suppose there is some smoothness in the coefficients $\alpha_0,\ldots,\alpha_a$ and the coefficients $\beta_0,\ldots,\beta_b$. That is, there is a tendency to believe that α_j does not differ radically from α_{j-1} or α_{j+1} nor β_j from β_{j-1} or β_{j+1}. Now the extreme form of this belief would be an absolute certainty that $\alpha_0 = \alpha_1 = \ldots = \alpha_b = \alpha$ and $\beta_0 = \beta_1 = \ldots = \beta_b = \beta$. Then the model, which will be named the extreme smoothness model, would be

$$y_t = \mu + \alpha \sum_{j=1}^{a+1} x_{tj} + \beta \sum_{j=a+2}^{a+b+2} x_{tj} + z_t.$$

Let V be the $N \times 2$ matrix whose first column is $\sum_{j=1}^{a+1} x_{tj}$ for $t = 1,\ldots,N$ and whose second column is $\sum_{j=a+2}^{a+b+2} x_{tj}$ for $t = 1,\ldots,N$. Then the maximum likelihood estimates of α and β would be

$$\begin{bmatrix} \hat{\alpha}_{ESM} \\ \hat{\beta}_{ESM} \end{bmatrix} = (V'V)^{-1} V' Y.$$

Thus under this extreme smoothness condition the maximum likelihood estimate of θ would be a column vector $\hat{\theta}_{ESM}$ whose first $a + 1$ elements are $\hat{\alpha}_{ESM}$ and whose next $b + 1$ elements are $\hat{\beta}_{ESM}$. At the other extreme is the estimate $\hat{\theta}_{MLE}$ which involves absolutely no smoothing assumption on the parameters.

The general idea of the remedy is this. A subset of the parameter space will be studied which contains $\hat{\theta}_{MLE}$, $\hat{\theta}_{ESM}$, and other values $\dot{\theta}$ which represent an amount of smoothing ranging between that of the two extremes. RL will be studied for a number of values in the subset and an estimate $\hat{\theta}$ chosen from the subset which is not judged unlikely by the data (i.e., for which $RL(\hat{\theta})$ is not too far from 1) but yet has an amount of smoothness in the coefficients which makes economic sense.

In the remainder of this section, two possible choices for the subset of study will be presented.

3.1. A One-Parameter Subset

Let W be an $(a+b+2) \times (a+b+2)$ diagonal matrix. The first $a+1$ elements on the diagonal have the same value

$$w_1 = \left\{\sum_{t=1}^{N} x_{t,1}^2\right\}^{\frac{1}{2}},$$

which is the square root of the sum of squares about the mean of the first exogenous variable a_t. The next $b+1$ elements on the diagonal have the same value

$$w_2 = \left\{\sum_{t=1}^{N} x_{t,a+2}^2\right\}^{\frac{1}{2}},$$

which is the square root of the sum of squares about the mean of the second exogenous variable b_t.

The one parameter subset $\dot{\theta}(p)$ for $p \geq 0$ is defined by

$$\dot{\theta}(p) = (pW^2 + X'X)^{-1} (pW^2\hat{\theta}_{ESM} + X'X\hat{\theta}_{MLE}).$$

$\dot{\theta}(0) = \hat{\theta}_{MLE}$, and as $p \to \infty$, $\dot{\theta}(p) \to \hat{\theta}_{ESM}$. As p increases the values of $\dot{\theta}(p)$ get smoother. For a particular value of p, $\dot{\theta}(p)$ has the property that it minimizes $(\dot{\theta}-\hat{\theta}_{ESM})' W^2(\dot{\theta}-\hat{\theta}_{ESM})$ over all values $\dot{\theta}$ such that $RL(\dot{\theta}) =$

$RL(\dot{\theta}(p))$. Another way of saying this is that over all points $\dot{\theta}$ such that $(\dot{\theta}-\hat{\theta}_{ESM})' W^2(\dot{\theta}-\hat{\theta}_{ESM}) = (\dot{\theta}(p)-\hat{\theta}_{ESM})' W^2(\dot{\theta}(p)-\hat{\theta}_{ESM})$, $\dot{\theta}(p)$ has the largest likelihood. Thus, for a given amount of smoothness, $\dot{\theta}(p)$ has the most likelihood. These statements may be proved by an argument entirely analogous to that in [4, p. 59].

In practice, an estimate of θ can be chosen from the one-parameter subset $\dot{\theta}(p)$ by calculating and studying $\dot{\theta}(p)$ and $RL(\dot{\theta}(p))$ for various values of p. A particular estimate $\hat{\theta}$ would be chosen from the subset which has an amount of smoothness that makes economic sense but for which $R(\hat{\theta})$ is not so small as to render $\hat{\theta}$ unlikely in light of the data. The choice of the estimate is also aided by noting by how much $\hat{\theta}$ inflates the residual sum of squares; thus calculating $SS(\dot{\theta}(p))$ and comparing with $SS(\hat{\theta}_{MLE})$ is useful.

3.2. A Two-Parameter Subset

Let r be a number in the interval $0 \le r < 1$ and g a positive integer. $G_g(r)$ will be the $(g+1) \times (g+1)$ matrix

$$G_g(r) = [r^{|j-k|}]_{j=0,\ldots,g;k=0,\ldots,g}.$$

For $p > 0$ let $H(r,p)$ be the $(a+b+2) \times (a+b+2)$ block diagonal matrix

$$H(r,p) = \frac{1}{p}\begin{bmatrix} G_a(r) & 0 \\ 0 & G_b(r) \end{bmatrix}.$$

The northwest block of $H(r,p)$ is an $(a+1) \times (a+1)$ matrix, the southeast block is a $(b+1) \times (b+1)$ matrix, and the other elements of the matrix have the value zero. For example, if $a = 1$ and $b = 2$, then

$$H(r,p) = \frac{1}{p}\begin{bmatrix} 1 & r & 0 & 0 & 0 \\ r & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & r & r^2 \\ 0 & 0 & r & 1 & r \\ 0 & 0 & r^2 & r & 1 \end{bmatrix}.$$

Let $GI_g(r)$ be the inverse of $G_g(r)$. $GI_g(r)$ is a tri-diagonal matrix with the following elements: the first and last elements of the main diagonal are $(1-r^2)^{-1}$; all other diagonal elements are $(1+r^2)(1-r^2)^{-1}$; all elements on the two diagonals parallel and adjacent to the main diagonal are $-r(1-r^2)^{-1}$; and all other elements are zero. For example

$$GI_2(r) = (1-r^2)^{-1} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}$$

and

$$GI_3(r) = (1-r^2)^{-1} \begin{bmatrix} 1 & -r & 0 \\ -r & 1+r^2 & -r \\ 0 & -r & 1 \end{bmatrix}.$$

Let $HI(r,p)$ be the inverse of $H(r,p)$. Thus

$$H(r,p) = p \begin{bmatrix} GI_a(r) & 0 \\ 0 & GI_b(r) \end{bmatrix}.$$

For $p = 0$, $HI(r,p)$ is defined to be an $(a+b+2) \times (a+b+2)$ matrix of zeros.

The two parameter subset is

$$\dot{\theta}(r,p) = [WH(r,p)W + X'X]^{-1} X'Y$$

for $0 \le r < 1$ and $p \ge 0$. $\dot{\theta}(r,p)$ contains the maximum likelihood estimate of θ since

$$\dot{\theta}(r,0) = \hat{\theta}_{MLE}.$$

As p increases $\dot{\theta}(r,p)$ moves toward the origin.

What is most interesting is the behavior of $\dot{\theta}(r,p)$ as $r \to 1$. Let

$$\begin{bmatrix} \dot{\alpha}_{ESM}(p) \\ \dot{\beta}_{ESM}(p) \end{bmatrix} = \left[p \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} + V'V \right]^{-1} V'Y.$$

Thus $\dot{\alpha}_{ESM}(0) = \hat{\alpha}_{ESM}$, $\dot{\beta}_{ESM}(0) = \hat{\beta}_{ESM}$, and $(\dot{\alpha}_{ESM}(p), \dot{\beta}_{ESM}(p))$ is the ridge trace [4,5] of α and β under the extreme smoothness condition. Let $\dot{\theta}_{ESM}(p)$ be an $(a+b+2) \times 1$ column vector whose first $a+1$ elements are $\dot{\alpha}_{ESM}(p)$ and whose next $b+1$ elements are $\dot{\beta}_{ESM}(p)$. Let $\dot{\alpha}(r,p)$ be the first $a+1$ values of $\dot{\theta}(r,p)$ and $\dot{\beta}(r,p)$ the next $b+1$ values. If $p > 0$ then as $r \to 1$

$$\dot{\theta}(r,p) \to \dot{\theta}_{ESM}(p).$$

That is, as r increases from 0 toward 1, the values of $\dot{\alpha}(r,p)$ get closer (i.e., smoother). When r is nearly one all the $a+1$ values of $\dot{\alpha}(r,p)$ are nearly $\dot{\alpha}_{ESM}(p)$. Similarly as r increases from 0 to 1 the values of $\dot{\beta}(r,p)$ get closer together, all approaching $\dot{\beta}_{ESM}(p)$.

In practice, an estimate of θ can be chosen from the two-parameter subset by calculating $\dot{\theta}(r,p)$ for various values of r and p and selecting a particular $\hat{\theta}$ from the subset which has an amount of smoothness that makes sense with respect to the economic mechanism, but for which $RL(\hat{\theta})$ is not so small as to render $\hat{\theta}$ unlikely in light of the data. Just as in the one parameter subset, it will be of interest to also study $SS(\dot{\theta}(r,p))$, the residual sum of squares of $\dot{\theta}(r,p)$.

In many practical situations studying values of p close to zero, for example, between .01 and .1 will be quite sufficient. In this situation, as r tends to 1, $\dot{\theta}(r,p)$ tends toward a vector which is nearly $\hat{\theta}_{ESM}$. However, if the columns of V are highly correlated, larger values of p will often be needed. For, in this situation, there will not be much precision in the estimate $\hat{\theta}_{ESM} = \hat{\theta}_{ESM}(0)$ despite the strong extreme smoothness assumption, and a $\dot{\theta}_{ESM}(p)$ with a larger value of p will be needed in order to get a sensible estimate.

4. A BAYESIAN INTERPRETATION OF THE METHOD

Suppose the prior distribution of θ and μ given ϕ has the properties that θ and μ are independent and θ is multivariate normal with mean M and covariance matrix $e^{\phi}C$. Then the posterior mean of θ, which is a very reasonable point estimate for the parameter, is

$$(C + X'X)^{-1} (CM + X'X\hat{\theta}_{MLE}).$$

4.1. The One-Parameter Family

Suppose $C = p^{-1} W^{-2}$ and $M = \hat{\theta}_{ESM}$. Then the posterior mean of θ is $\dot{\theta}(p)$. This represents forming a prior by, as Dickey [3, p. 1481] puts it, "peeking at the data", since $\hat{\theta}_{ESM}$ depends on the values of Y. The prior variance of α_j is $p^{-1}w_1^{-1}$. If p is small, there is little prior information about the value of θ, so the data determine the value of the posterior mean, which is nearly $\hat{\theta}_{MLE}$. As p gets larger the information in the prior about the value of θ gets large compared with the information in the sample, the prior belief becomes stronger that θ is near $\hat{\theta}_{ESM}$, and the posterior mean of θ tends to $\hat{\theta}_{ESM}$.

4.2. The Two-Parameter Family

Suppose $C = W^{-1}H(r,p)W^{-1}$ and M is a vector of zeros. Then the posterior mean of θ is $\dot{\theta}(r,p)$. The prior variance of each α_j is $w_1^{-2}p^{-1}$ and the prior mean is 0. The prior variance of each β_j is $w_2^{-2}p^{-1}$ and the prior mean is 0.

For $p > 0$ the prior correlation between α_j and β_k is $\underline{0}$ and the prior correlation between α_j and α_k and between β_j and β_k is $r^{|j-k|}$ which depends only on the time lag between the variables corresponding to the parameters. As the time lag increases the correlation decreases which would seem to be a desirable property. As r tends to 1 there is increasingly more prior information that α_j and α_k have similar values (i.e. that there is more smoothness in the coefficients $\alpha_0,\ldots,\alpha_a$) and hence the posterior means of $\alpha_0,\ldots,\alpha_a$, the $a+1$ elements of $\mathring{\alpha}(r,p)$, tend toward the same value, $\mathring{\alpha}_{ESM}(p)$. A similar statement holds for $\mathring{\beta}(r,p)$.

If p is near 0 and r is near 1, there is little prior information about the value of an individual α_j or β_j but a strong prior belief that the values of the various α_j are nearly equal. This prior assumption is nearly the same as the extreme smoothness assumption and so $\mathring{\theta}(r,p)$ is nearly $\hat{\theta}_{ESM}$.

As p and r both get close to zero the information in the prior about the true value of θ gets small compared to the information in the sample and $\mathring{\theta}(r,p)$ is nearly $\hat{\theta}_{MLE}$.

5. EXAMPLE

The use of the one-parameter subset $\mathring{\theta}(p)$ and the two-parameter subset $\mathring{\theta}(r,p)$ will be illustrated by their application to two equations from the econometric model in [2].

The first is the price equation

$$y_t = \gamma + \sum_{j=0}^{5} \alpha_j a_{t-j} + \beta_0 b_t,$$

where y_t is the change in the price level from quarter $t-1$ to t, a_t is the demand pressure in quarter t, and b_t is the change from quarter $t-1$ to t in the anticipated price level. $t = 1$ corresponds to the first quarter of 1955 and $N = 60$. Display 1 shows the correlations between the columns of the matrix X; there is very clearly a substantial amount of correlation.

Display 2 shows the results of calculating $\dot{\theta}(p)$ for 15 values of p. The individual elements of the vector $\dot{\theta}(p)$ are defined by

$$\dot{\theta}(p) = (\dot{\alpha}_0(p),\ldots,\dot{\alpha}_5(p),\dot{\beta}_0(p))'.$$

The maximum likelihood (least squares) estimate of θ, $\hat{\theta}_{MLE} = \dot{\theta}(0)$, does not appear to make very much sense for the estimates of α_2 and α_5 are both negative, which would imply prices increase by a smaller amount during a given quarter if the demand was large 3 quarters earlier or 5 quarters earlier.

$\dot{\theta}(.01)$ is quite different from $\hat{\theta}_{MLE}$ even though $RL(\dot{\theta}(.01))$ is as high as .91 and $SS(\hat{\theta}_{MLE})$ is 99.7% of $SS(\dot{\theta}(.01))$. This is exactly a feature of the lack-of-precision phenomenon resulting from high correlations and a small sample size that was described in Section 2. $\dot{\alpha}_2(p)$ and $\dot{\alpha}_4(p)$ increase as p increases, going from negative to positive for a value of p between .01 and .05. As p increases, the values $\dot{\alpha}_0(p),\ldots,\dot{\alpha}_5(p)$ get smoother.

The choice of a particular $\dot{\theta}(p)$ as an estimate of θ would, of course, depend upon one's knowledge about the economic mechanism. p in the range .4 to .6 might make sense to many. For these values of p the $\dot{\alpha}_j(p)$ for $j = 0,\ldots,p$ are reasonably smooth and decrease as j increases, and $RL(\dot{\theta}(p))$ and $SS(\hat{\theta}_{MLE}) \div SS(\dot{\theta}(p))$ are not so small as to discredit $\dot{\theta}(p)$ in light of the data.

The second equation is the total spending equation

$$y_t = \gamma + \sum_{j=0}^{4} \alpha_j a_{t-j} + \sum_{j=0}^{4} \beta_j b_{t-j} + z_t,$$

where y_t, a_t, and b_t are the changes from quarter $t-1$ to t of, respectively, the nominal GNP, the money stock, and high-employment Federal expenditures. $t = 1$ corresponds to the first quarter of 1953 and $N = 68$. Display 3 shows the correlations between the columns of the matrix X.

Display 4 shows the results of calculating $\dot{\theta}(p,r)$ for various values of p and r. The individual elements of the vector $\theta(r,p)$ are defined by

$$\dot{\theta}(r,p) = (\dot{\alpha}_0(r,p),\ldots,\dot{\alpha}_5(r,p),\dot{\beta}_0(r,p),\ldots,\dot{\beta}_5(r,p))'.$$

The maximum likelihood (least squares) estimates $\hat{\theta}_{MLE} = \dot{\theta}(0,0)$ do not appear grossly inadequate, but the small value of the estimate of α_1 and the negative value of the estimate of α_4 might be regarded by some as unrealistic. $\dot{\theta}(.15,.6)$ might be a more realistic estimate for many. $\dot{\alpha}_1(.15,.6) = 1.35$ is more nearly in line with $\dot{\alpha}_0(.15,.6) = 1.53$ and $\dot{\alpha}_2(.15,.6) = 1.59$, and $\dot{\alpha}_5(.15,.6) = .06$ is positive. $SS(\hat{\theta}_{MLE}) \div SS(\dot{\theta}(.15,.6)) = .95$ and $R(\dot{\theta}(.15,.6)) = .23$ are not so low as to completely discredit $\dot{\theta}(.15,.6)$ in light of the data. In should be emphasized that others may find another $\dot{\theta}(p,r)$ more sensible, for the choice will (and must since there is not enough information in the data to specify the model with great precision) depend upon one's prior knowledge about the economic mechanism.

6. GENERAL COMMENTS

The procedures described in this paper may be applied to the general distributed lag model

$$y_t = \gamma + \sum_{k=1}^{n} \sum_{j=0}^{a_k} \alpha_{kj}\, a_{k,t-j} + z_t.$$

For this general model $H(r,p)$ is a block diagonal matrix with n blocks, and $\hat{\theta}_{ESM}$ and W each have n different values. Each of the n variables $a_{k,t}$ for $t = 1,\ldots,N$ might be an exogenous variable, a dummy variable, or even a lagged value of the endogenous variable.

More practical experience is needed to judge the relative merits of $\dot{\theta}(r,p)$ and $\dot{\theta}(p)$. My own tentative opinion at the time of writing of this paper is that the easier to use $\dot{\theta}(p)$ will suffice provided the columns of V are not too highly correlated but that otherwise the more delicate, and more complicated, $\dot{\theta}(r,p)$ is needed.

A very reasonable competitor to the method presented in this paper is the Almon lag technique [1]. Both employ an element of subjectivity; in the method of this paper a value of p or values of r and p must be chosen and in the Almon lag technique the degree of the smoothing polynomial must be chosen. However, the method here is, I believe, simpler, more flexible, and provides an easier to understand and more delicate way of codifying prior information about the smoothness of the coefficients. In addition, $\dot{\theta}(r,p)$ can cope with the problem of high correlation between the different exogenous variables, which the Almon lag technique cannot.

The subsets $\dot{\theta}(r,p)$ and $\dot{\theta}(p)$ while different in detail from the ridge trace of Hoerl and Kennard [4,5] are very much in the same spirit. These two excellent papers on the use of the ridge trace are very worthwhile reading.

Display 1. **Correlations Between the Columns of X for the Price Equation (values are rounded to 2 places and multiplied by 100).**

Columns	1	2	3	4	5	6
2	96					
3	89	96				
4	81	90	96			
5	74	82	89	96		
6	68	74	81	88	96	
7	45	52	56	60	63	65

Display 2. $\dot{\theta}(p)$ for the Price Equation

p	$RL(\dot{\theta}(p))$	$\frac{SS(\hat{\theta}_{MLE})}{SS(\dot{\theta}(p))}$	$\dot{\theta}(p)$						
			$\dot{\alpha}_0(p)$	$\dot{\alpha}_1(p)$	$\dot{\alpha}_2(p)$	$\dot{\alpha}_3(p)$	$\dot{\alpha}_4(p)$	$\dot{\alpha}_5(p)$	$\dot{\beta}_0(p)$
0	1	1	.029	.038	−.026	.021	.043	−.017	.869
.01	.91	1.00	.029	.028	−.010	.017	.032	−.009	.864
.05	.65	.99	.027	.022	.005	.015	.019	.002	.852
.1	.55	.98	.025	.021	.010	.014	.016	.005	.845
.2	.46	.97	.023	.019	.013	.014	.014	.008	.837
.3	.41	.97	.021	.019	.014	.014	.014	.009	.831
.4	.37	.97	.020	.018	.014	.014	.014	.010	.828
.5	.34	.96	.020	.018	.014	.014	.014	.011	.826
.6	.32	.96	.019	.018	.015	.014	.014	.011	.824
.7	.31	.96	.019	.018	.015	.014	.014	.012	.822
.8	.29	.96	.019	.017	.015	.014	.014	.012	.821
.9	.28	.96	.018	.017	.015	.015	.014	.012	.821
1	.27	.96	.018	.017	.015	.015	.014	.013	.820
2	.22	.95	.017	.016	.015	.015	.015	.014	.817
5	.19	.94	.016	.016	.015	.015	.015	.015	.816

Display 3. Correlations Between the Columns of X for the Total Spending Equation (values are rounded to 2 places and multiplied by 100).

Columns	1	2	3	4	5	6	7	8	9
2	75								
3	45	75							
4	27	47	76						
5	20	30	49	76					
6	33	28	31	33	30				
7	38	33	28	30	31	52			
8	39	39	34	27	29	53	50		
9	37	39	39	34	29	36	52	51	
10	30	36	38	40	34	30	38	50	48

Display 4. $\dot{\theta}(r,p)$ for the Total Spending Equation

r	$RL(\dot{\theta}(r,p))$	$\frac{SS(\hat{\theta}_{MLE})}{SS(\dot{\theta}(r,p))}$	$\dot{\theta}(r,p)$: $\alpha_0(r,p)$	$\alpha_1(r,p)$	$\alpha_2(r,p)$	$\alpha_3(r,p)$	$\alpha_4(r,p)$	$\beta_0(r,p)$	$\beta_1(r,p)$	$\beta_2(r,p)$	$\beta_3(r,p)$	$\beta_4(r,p)$
					p = .05							
0.0	.81	.99	1.91	.66	2.09	1.11	-.28	.44	.45	.14	-.54	-.43
0.2	.78	.99	1.86	.77	2.03	1.13	-.27	.44	.45	.14	-.54	-.44
0.4	.71	.99	1.80	.89	1.96	1.13	-.24	.45	.45	.13	-.53	-.45
0.6	.58	.98	1.73	1.05	1.85	1.11	-.17	.46	.44	.11	-.51	-.46
0.8	.35	.97	1.64	1.26	1.67	1.03	-.01	.47	.41	.08	-.46	-.46
					p = .1							
0.0	.54	.98	1.76	.82	1.90	1.07	-.19	.43	.42	.13	-.48	-.40
0.2	.53	.98	1.71	.96	1.85	1.08	-.17	.44	.43	.12	-.49	-.42
0.4	.46	.98	1.65	1.09	1.78	1.07	-.12	.45	.43	.11	-.48	-.43
0.6	.35	.97	1.60	1.24	1.69	1.03	-.03	.46	.41	.09	-.45	-.44
0.8	.18	.95	1.57	1.38	1.53	.96	.15	.46	.36	.06	-.39	-.44
					p = .15							
0.0	.33	.97	1.66	.91	1.76	1.02	-.12	.43	.39	.12	-.44	-.38
0.2	.35	.97	1.61	1.06	1.73	1.04	-.10	.43	.40	.11	-.45	-.40
0.4	.31	.97	1.56	1.19	1.68	1.02	-.04	.44	.40	.10	-.44	-.41
0.6	.23	.96	1.53	1.31	1.59	.97	.06	.45	.38	.08	-.41	-.42
0.8	.11	.94	1.53	1.41	1.45	.93	.24	.43	.34	.05	-.35	-.41
					p = .2							
0.0	.20	.95	1.58	.97	1.66	.99	-.07	.42	.37	.12	-.40	-.35
0.2	.23	.96	1.53	1.11	1.65	1.00	-.04	.42	.39	.11	-.41	-.38
0.4	.21	.95	1.50	1.24	1.60	.99	.02	.43	.38	.09	-.40	-.39
0.6	.16	.95	1.49	1.35	1.53	.95	.12	.43	.37	.07	-.38	-.40
0.8	.06	.92	1.50	1.41	1.39	.91	.31	.41	.31	.05	-.31	-.39

REFERENCES

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

[2] Andersen, Leonall C. and Carlson, Keith M., "A Monetarist Model for Economic Stabilization," *Review - Federal Reserve Bank of St. Louis*, April 1970, 7-25.

[3] Dickey, James M., "Smoothing by Cheating," *The Annals of Mathematical Statistics 40*, 1969, 1477-1482.

[4] Hoerl, Arthur E. and Kennard, Robert W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics 12*, 1970, 55-67.

[5] Hoerl, Arthur E. and Kennard, Robert W., "Ridge Regression: Application to Nonorthogonal Problems," *Technometrics 12*, 1970, 69-78.

[6] Raiffa, Howard and Schlaifer, Robert, *Applied Statistical Decision Theory*, Division of Research, Harvard Business School, 1961.