THE SELECTION OF SMOOTHNESS PRIORS FOR DISTRIBUTED LAG ESTIMATES

JUN 82 H AKAIKE

UNCLASSIFIED MRC-TSR-2394
THE SELECTION OF SMOOTHNESS PRIORS FOR DISTRIBUTED LAG ESTIMATION

Hirotugu Akaike

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

June 1982

(Received May 16, 1982)

Approved for public release
Distribution unlimited

Sponsored by
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709
THE SELECTION OF SMOOTHNESS PRIORS FOR DISTRIBUTED LAG ESTIMATION

Hirotugu Akaike*

Technical Summary Report #2394
June 1982

ABSTRACT

In the application of Shiller's smoothness prior for distributed lag estimation the main difficulty is the selection of hyperparameters of the prior distribution. In this paper the use of a maximum likelihood procedure is proposed for this purpose and its performance is demonstrated by numerical examples.

AMS (MOS) Subject Classifications: 62F15, 62M10, 90A20

Key Words: Smoothness prior; Bayesian model; Likelihood; Hyperparameter; Distributed lag; Impulse response

Work Unit Number 4 (Statistics and Probability)

*The Institute of Statistical Mathematics, Tokyo, Japan

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
The distributed lag estimator developed by R. J. Shiller based on the concept of smoothness prior is a significant example that demonstrates the potential of Bayesian approach in statistics. However, the practical application of Shiller's estimator has been hampered by the difficulty of specifying the prior distribution.

In this paper an objective procedure for the selection of the prior distribution is proposed and its performance is checked by both artificial and real numerical examples. The result clearly demonstrates the practical utility of the procedure. It also shows the danger inherent in an arbitrary subjective choice of the prior distribution.

It is expected that the result reported in this paper will contribute to the development of practical applications of Bayesian statistics.
1. INTRODUCTION

Shiller (1973) introduced the concept of smoothness priors to define a Bayesian estimator of the lag coefficients, or the impulse response function, of a linear system. In this approach the prior preference of the smoothness of the lag function is expressed by a spherical normal distribution of the fixed order differences of the lag coefficients.

One outstanding problem in the application of Shiller's estimator is the choice of the lag length, the order of differencing and the variance of the spherical normal prior distribution. In this paper we propose a practical solution to this problem obtained by maximizing a properly defined likelihood of the Bayesian model.

The performance of the estimator is checked by numerical examples. The result shows that the estimator performs satisfactorily under widely varying condition. It also shows that an arbitrary subjective choice of the prior distribution can produce awkward estimate.

2. SMOOTHNESS PRIOR FOR DISTRIBUTED LAG ESTIMATION

Consider the stochastic linear system defined by

\[ y_n = \sum_{m=0}^{\infty} a_n x_{n-m} + w_n \]

where \( y_n \), \( x_n \) and \( w_n \) denote the output, input and error term of the system, respectively. Here \( \{w_n\} \) is assumed to be a sequence of random variables which are independent of \( \{x_n\} \) and are independently identically distributed as normal with mean zero and variance \( \sigma^2 \).

---

*The Institute of Statistical Mathematics, Tokyo, Japan

Sponsored by the United States Army under Contract No. DAAE29-80-C-0041.
Shiller introduced the smoothness prior defined by a spherical normal distribution of the \( d^{th} \) differences of the distributed lag coefficients \( \{a_n\} \). This prior distribution is given by

\[
p(a|\sigma, \lambda) = \frac{1}{(2\pi\sigma^2)^{n/2}} \lambda^{n/2} a \exp\left(-\frac{1}{2\sigma^2} a'R_1 a\right),
\]

where \( | \cdot | \) denotes the determinant, \( a = (a_0, a_1, \ldots, a_n)' \), \( \lambda \) is the ratio of the standard deviation \( \sigma \) of \( v_n \) to that of the \( d^{th} \) difference of \( a_n \) and \( R_1 = R_1^d \), where \( R_1 \) is an \((n + 1) \times (n + 1)\) matrix defined by

\[
R_1 = \begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
0 & 1 & -1 & \cdots & 0 \\
& & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}.
\]

It should be noticed here that the last \( d \) rows of \( R_1 \) define incomplete differences. However, as was mentioned by Shiller, this is equivalent to connecting the final coefficients to \( a_n = 0 \) for \( n > N \), which will be a reasonable assumption for practical application if the lag length \( N \) is taken sufficiently large.

By combining the present prior distribution with the data distribution

\[
p(y|a, \sigma) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} y - Z a\right),
\]

where \( y = (y_1, y_2, \ldots, y_N)' \), \( | \cdot | \) denotes the Euclidean norm, and

\[
Z = \begin{bmatrix}
Z_1 & Z_2 & \cdots & Z_{N-H} \\
& & \ddots & \ddots \\
& & & \ddots & \ddots \\
& & & & \ddots & Z_{N-H}
\end{bmatrix},
\]

the estimate \( a^* \) of the lag coefficients is defined as the posterior mean of \( a \) and is
given by

\[ a_n = (z'k + \lambda^2 a'_n q)^{-1} k'y.\]

Miller demonstrated by numerical examples the superiority of this type of estimator to both the ordinary least squares estimator and the Almon lag estimator. This is one of the earliest examples of successful Bayesian modeling of practical importance.

The crucial point in applying Miller's estimator is the choice of the hyperparameter \( \lambda \). Miller suggested some rule of thumb for the choice of \( \lambda \). However, this is deeply concerned with the basic problem of the selection of a Bayesian model which is vital in implementing a Bayesian procedure.

3. SELECTION OF A PRIOR DISTRIBUTION BY LIKELIHOOD

When there are finite number of Bayesian models defined by the data distributions \( f_k(*)|\theta_k \) and corresponding prior distributions \( \pi_k(\theta_k) \) \( (k = 1, 2, \ldots, K) \) the posterior probability of each model is given by

\[ p(k|y) = \frac{f(y|k)C_k}{\sum_{k=1}^{K} f(y|k)C_k}, \]

where \( y \) denotes the observation, \( C_k \) denotes the prior probability of the \( k \)th model and \( f(y|k) \) is defined by

\[ f(y|k) = \int f_k(y|\theta_k)\pi_k(\theta_k) d\theta_k. \]

The above formula of \( p(k|y) \) shows that it is natural to call \( f(y|k) \) the likelihood of the Bayesian model specified by \( f_k(*)|\theta_k \) and \( \pi_k(\theta_k) \).

Recent work by the present author suggests that when there is no further prior information available and the distributions \( f(*)|k \) are well separated, i.e., only one model attains high likelihood for one particular observation, then the equal prior probability \( C_k = 1/K \) is a natural choice that let the data speak most (Akaike, 1982).

With this choice of the prior probability distribution the maximum likelihood selection that chooses the Bayesian model with maximum \( f(y|k) \) is equivalent to the selection by maximum posterior probability. This shows that under certain circumstances the selection
of a Bayesian model by maximising the likelihood can be a reasonable procedure from the Bayesian point of view. In this paper we pursue the possibility of applying this idea to the problem of selection of the smoothness prior.

In practical applications of the smoothness prior for the distributed lag estimation we usually do not know the value of $\sigma$. Accordingly we have to specify a prior distribution of $\sigma$. To make the resulting estimator applicable to observations in arbitrary scale unit we consider the use of Jeffrey's ignorance prior $\sigma^{-1} \, d\sigma$. For each particular application we may consider a proper prior distribution $C(u, v)\sigma^{-1}$ obtained by restricting the range of $\sigma$ to a finite interval $(u, v)$ of positive numbers. However, since the integral of $p(y | a, \sigma)p(a | \sigma, \lambda)\sigma^{-1}$ with respect to $d\sigma$ is finite, we may use this integral as the likelihood of the Bayesian model specified by $p(y | a, \sigma)$ and $p(a | \sigma, \lambda)C(u, v)\sigma^{-1}$ with sufficiently small $u$ and sufficiently large $v$. Since only ratios of the likelihoods are of interest we define the likelihood of the Bayesian model, specified by the data distribution $p(y | a, \sigma)$ and the (improper) prior distribution $p(a, \sigma | \lambda) = p(a | \sigma, \lambda)\sigma^{-1}$, by

$$p(y | \lambda) = \int p(y | a, \sigma)p(a | \sigma, \lambda) \, d\sigma \, d\sigma.$$  

Since the hyperparameter $\lambda$ has a clearly defined technical meaning as the ratio of the standard deviation of an error term to that of a difference of lag coefficients, it is not difficult to find a reasonable selection of finite number of possible values of $\lambda$. The model selection is realised by maximising the likelihood over this set of $\lambda$.

For our present model we have

$$p(y | a, \sigma) = \frac{1}{2\pi \sigma^2} e^{-\frac{1}{2\sigma^2}(y - a)^T(y - a)}$$

and

$$p(a | \sigma, \lambda) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{1}{2\sigma^2}(a - \lambda)^T(a - \lambda)}.$$

where $a_0 = (X'X + \lambda^2\sigma^2I_d)^{-1}X'y$ and $s(\lambda) = 1y^2 - a_0^T(X'X + \lambda^2\sigma^2I_d)a_0$. The likelihood $p(y | \lambda)$ of this model is then given by

$$p(y | \lambda) = \int p(y | a, \sigma)p(a | \sigma, \lambda) \, d\sigma.$$
Obviously the estimator $a_*$ is given as the solution of the least squares problem that minimizes $\|y - Xa\|_2^2$, where

$$X^* = \begin{bmatrix} x \\ \lambda_r \end{bmatrix}, \quad y^* = \begin{bmatrix} y \\ 0 \end{bmatrix}. $$

Here it holds that $s(\lambda) = \|y - Xa\|_2^2$. By successively applying the Householder transformation to $[X'y^*]$, first to transform $\lambda_r$ into upper triangular form and then to transform the whole matrix into upper triangular form, the necessary quantities for the likelihood computation can easily be obtained during the process of the least squares computation. For the purpose of the comparison of models we may ignore the constant factor $\frac{N}{2 \pi}$ of the likelihood.

It should be noted here that the likelihood $p(y|\lambda)$ is also a function of $d$, the order of differencing, and $M$, the lag length. Thus our search for the best model is realised by maximising $p(y|\lambda)$ over some finite number of possible combinations of $\lambda$, $d$ and $M$. The practical utility of this procedure will be demonstrated by numerical examples in the next section. The feasibility of this type of procedure was first discussed in Akaike (1980a) and its application to seasonal adjustment was discussed by Akaike (1980b). The idea of maximising a likelihood with respect to the hyperparameter is discussed in an earlier paper by Good (1965).

4. NUMERICAL INVESTIGATION

From the point of view of data analysis a Bayesian model is simply an artificial construction that allows to generate an output from a given set of data. Only when we confirm that it often produces results superior to those obtained by other conventional procedures we can claim the usefulness of the Bayesian procedure. In this section we will
discuss the practical utility of the distributed lag estimation realised by the maximum likelihood selection of the smoothness prior.

First we discuss the application to the artificial example discussed by Shiller as his second example. In this example the input series was the series of four to six month commercial paper rate which gave a typical nearly collinear matrix \( X \). The output series \( y \) was generated by the relation \( y = Ax + w \) with \( w = 0.05e \), where \( e \) is a vector of standard normal random numbers. The lag coefficients were defined by \( a_n = \phi(t_i)(n - 9) \) with \( N = 19 \), where \( \phi(t) \) denotes the density of the standard normal distribution. The dimension, or the length, of \( y \) was given by \( N = 40 \). One data set generated by this model was used in the subsequent analysis.

The search for the smoothness prior was extended over the models defined with \( \lambda = 5 \times 2^k \ (k = -10, (1), 10) \), \( N = -1, (1), 19 \), where \( N = -1 \) denotes zero regression, and \( d = 1, 2, 3 \). Since we are accustomed to the use of log likelihood ratio test we used \((-2)\log p(y|\lambda)\) as our criterion. By ignoring the additive constant the criterion is given by

\[
\text{ABIC} = N \log S(\lambda) + \log|2^{1\times}X + \lambda^2\Sigma_{\lambda}^d| - \log|\lambda^2\Sigma_{\lambda}^d|,
\]

where \( N \) is the dimension of the vector \( y \), \( \log \) denotes natural logarithm and \( \text{ABIC} \) stands for a Bayesian information criterion (Akaike, 1980a). Our search for the model was realized by finding the minimum of \( \text{ABIC} \).

The best choice in terms of the criterion was given by \( N = 13, \ d = 2 \) and \( \lambda = 5 \times 2^{-3} \). The resulting estimate is shown in Table 1 along with the true values and the ordinary least squares estimate for \( N = 19 \), denoted by \( LS \). By Table 1 it is obvious that the present estimation procedure is producing significantly improved estimate over the ordinary least squares estimate. Although the present result is not directly comparable with Shiller's result due to the use of different realizations of the error term, the shape of the estimated distributed lag coefficients is quite similar to some of the best ones given by Shiller.

One might be concerned with the possibility of non-smooth behavior of distributed lag coefficients. To check the performance of our procedure under such circumstances we
### TABLE 1

**RESULTS OF AN EXPERIMENT OF SHILLER'S SECOND EXAMPLE.**

*LS denotes least squares estimate.*

<table>
<thead>
<tr>
<th>n</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>.000</td>
<td>.000</td>
<td>.001</td>
<td>.004</td>
<td>.018</td>
</tr>
<tr>
<td>$a^*$</td>
<td>-.011</td>
<td>-.001</td>
<td>.008</td>
<td>.014</td>
<td>.015</td>
</tr>
<tr>
<td>LS</td>
<td>-.010</td>
<td>.021</td>
<td>-.045</td>
<td>.037</td>
<td>.078</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>.054</td>
<td>.130</td>
<td>.242</td>
<td>.352</td>
<td>.399</td>
</tr>
<tr>
<td>$a^*$</td>
<td>.045</td>
<td>.136</td>
<td>.260</td>
<td>.347</td>
<td>.400</td>
</tr>
<tr>
<td>LS</td>
<td>-.074</td>
<td>.255</td>
<td>.113</td>
<td>.462</td>
<td>.334</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>.352</td>
<td>.242</td>
<td>.130</td>
<td>.054</td>
<td>.018</td>
</tr>
<tr>
<td>$a^*$</td>
<td>.370</td>
<td>.250</td>
<td>.125</td>
<td>.060</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>.359</td>
<td>.329</td>
<td>.046</td>
<td>.072</td>
<td>.042</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>.004</td>
<td>.001</td>
<td>.000</td>
<td>.000</td>
<td>0</td>
</tr>
<tr>
<td>$a^*$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>-.018</td>
<td>-.050</td>
<td>.065</td>
<td>-.008</td>
<td>-.008</td>
</tr>
</tbody>
</table>
considered a system defined by
\[ y_n = 1.2 x_n - 0.6 x_{n-1} + 0.4 x_{n-2} + w_n, \]
where the input \( x_n \) was the same as that in the preceding example and thus \( N = 40 \), and \( w_n \) was also normal with mean zero and \( \sigma = 0.05 \). The range of the search for the parameters was the same as in the preceding example. The minimum of \( \text{ABIC} \) was attained at \( d = 1, M = 2 \) and \( \lambda = 5 \times 2^{-7} \). The resulting estimate \( a_0 \) and the estimate \( a_{\ast\ast} \), which was obtained by the parameters used for the computation of \( a_0 \) of the preceding example, are given in Table 2 along with the true values and \( \text{LS} \), the ordinary least squares estimate for \( N = 19 \).

The most remarkable finding with this result is that our procedure produced extremely good result. This was made possible by the correct determination of \( M \) and the choice of a very small value as \( \lambda \). Contrary to this, \( a_{\ast\ast} \), which was obtained by the parameters of Table 1 produced very poor result, even worse than \( \text{LS} \). This clearly demonstrates the danger of applying a Bayesian model based on an arbitrary choice of the prior distribution. It is obvious that a proper procedure of adaptation is necessary for the practical application of the smoothness prior.

Having confirmed the performance of our procedure at the two extreme situations, most favorable and unfavorable, we now turn to the example of real data handled by Shiller as his first example. In this example Shiller analyzed the response of the Federal Reserve Board Aaa new issue yield series to the four to six month prime commercial paper rate. Due to the unavailability of the Federal Reserve Board Aaa new issue yield series we used the corresponding series of Moody's AAA bond yield as the output series. The similarity of our result to Shiller's confirms that the substitution did not change the essential aspect of the problem.

In this example it is already noticed by Shiller that the coefficient \( a_0 \) has different characteristic from other coefficients and should be freed from the rest. The validity of this hypothesis can easily be checked by our present approach by considering models obtained by multiplying the first row of \( R_d \) by a small positive number \( \delta \). For the present example \( \text{ABIC} \) kept decreasing when \( \delta \) was decreased from 1 to 0.0001 in
TABLE 2

TEST RESULT FOR NON-SMOOTH DISTRIBUTED LAG COEFFICIENTS.
*a** denotes the estimate obtained by the parameters of a* of Table 1.

<table>
<thead>
<tr>
<th>m</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1.2</td>
<td>-.6</td>
<td>.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*</td>
<td>1.206</td>
<td>-.601</td>
<td>.396</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a**</td>
<td>.859</td>
<td>.068</td>
<td>-.075</td>
<td>.051</td>
<td>.090</td>
</tr>
<tr>
<td>LS</td>
<td>1.193</td>
<td>-.587</td>
<td>.390</td>
<td>-.046</td>
<td>.045</td>
</tr>
<tr>
<td>m</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a**</td>
<td>.037</td>
<td>-.025</td>
<td>-.012</td>
<td>.021</td>
<td>-.010</td>
</tr>
<tr>
<td>LS</td>
<td>.018</td>
<td>-.018</td>
<td>.037</td>
<td>-.037</td>
<td>.026</td>
</tr>
<tr>
<td>m</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a**</td>
<td>-.000</td>
<td>.039</td>
<td>-.000</td>
<td>-.038</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>-.029</td>
<td>.030</td>
<td>-.027</td>
<td>.059</td>
<td>-.067</td>
</tr>
<tr>
<td>m</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a**</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>.028</td>
<td>-.015</td>
<td>-.052</td>
<td>.092</td>
<td>-.043</td>
</tr>
</tbody>
</table>
several steps. This confirmed the validity of Shiller’s argument. The following result was obtained with $\delta = 0.0001$ and the same range of the parameters as in the preceding examples.

The minimum of ABIC was attained at $d = 3$, $M = 19$ and $\lambda = 5 \times 2^5$. The amount of reduction of the minimum ABIC obtained by reducing $\delta$ from 1 to 0.001 was 20.7. The estimate $a_0$ is given in Table 3 along with the estimate $a_0^{1.0}$ obtained by putting $\delta = 1.0$. In this example the constant term $a_{-1}$ was included to represent the effect of non-zero average values of the input and output series. The estimate denoted by $a_0^d$ was obtained from the series of first order differences of the input and output series. This was to check the possible effect of the trends of both series. The estimate $a_0^d$ was obtained with $d = 3$, $M = 19$ and $\lambda = 5 \times 2^4$. The corresponding ordinary least squares estimate for the differenced series is given by $LS$.

The similarity between $a_0$ and $a_0^d$ is remarkable and confirms that the smooth behavior of $a_0$ does not represent the spurious response due to the trend components. Also the distortion caused by the inclusion of $a_0$ can be seen clearly from $a_0^{1.0}$. The usual erratic pattern of the least squares estimate persists in this example.

Summarizing our observations of numerical results, including those not reported here, we may conclude as follows:

1) the estimator is most sensitive to the choice of $\lambda$,
2) the choice of $M$, the lag length, is also fairly critical,
3) the choice of $d$, the order of differencing, is not so critical.

We also found that the selection of $M$ must be done with ABIC minimized with respect to $\lambda$. Without the adjustment of $\lambda$ the selection of $M$ by minimizing ABIC produced poor results.
### Table 3

**RESULT OF APPLICATION TO REAL ECONOMIC DATA.** $a^0$ denotes the result obtained under the assumption of smooth connection to $a_0$. $a^d$ denotes the result obtained from differenced series.

<table>
<thead>
<tr>
<th>$m$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^*$</td>
<td>.212</td>
<td>.029</td>
<td>.045</td>
<td>.058</td>
<td>.068</td>
<td>.074</td>
<td>.079</td>
<td>.078</td>
<td>.076</td>
<td>.072</td>
</tr>
<tr>
<td>$a^1.0$</td>
<td>.104</td>
<td>.088</td>
<td>.076</td>
<td>.070</td>
<td>.068</td>
<td>.069</td>
<td>.072</td>
<td>.075</td>
<td>.076</td>
<td>.075</td>
</tr>
<tr>
<td>$a^d$</td>
<td>.281</td>
<td>.005</td>
<td>.031</td>
<td>.052</td>
<td>.068</td>
<td>.078</td>
<td>.083</td>
<td>.084</td>
<td>.081</td>
<td>.076</td>
</tr>
<tr>
<td>$LS$</td>
<td>.352</td>
<td>-.132</td>
<td>.081</td>
<td>.127</td>
<td>-.009</td>
<td>.106</td>
<td>.106</td>
<td>.084</td>
<td>.054</td>
<td>.144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m$</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^*$</td>
<td>.066</td>
<td>.058</td>
<td>.049</td>
<td>.040</td>
<td>.031</td>
<td>.022</td>
<td>.014</td>
<td>.008</td>
<td>.004</td>
<td>.001</td>
</tr>
<tr>
<td>$a^1.0$</td>
<td>.072</td>
<td>.065</td>
<td>.056</td>
<td>.046</td>
<td>.035</td>
<td>.024</td>
<td>.014</td>
<td>.007</td>
<td>.002</td>
<td>0</td>
</tr>
<tr>
<td>$a^d$</td>
<td>.068</td>
<td>.059</td>
<td>.049</td>
<td>.039</td>
<td>.030</td>
<td>.022</td>
<td>.014</td>
<td>.008</td>
<td>.004</td>
<td>.001</td>
</tr>
<tr>
<td>$LS$</td>
<td>.077</td>
<td>.058</td>
<td>.057</td>
<td>.025</td>
<td>.038</td>
<td>.079</td>
<td>.048</td>
<td>.063</td>
<td>-.042</td>
<td>.037</td>
</tr>
</tbody>
</table>
5. DISCUSSION

The purpose of the present paper has been to show the feasibility of an objectively defined procedure for the selection of the smoothness prior for the estimation of distributed lag coefficients. It was confirmed that the maximum likelihood procedure proposed in this paper can produce results comparable to the results reported by Shiller in his original paper. Since Shiller's results may be considered as typical examples produced by an expert this shows that the present procedure is producing a good approximation to the judgement procedure of an expert. The next step of the Bayesian modeling will be the specification of a prior distribution for the lag length $M$.

The result of Table 2 demonstrated the robustness of the present estimation procedure. However, it also demonstrated the danger inherent in the purely subjective selection of a prior distribution.

Although the result reported in this paper has demonstrated the feasibility of the smoothness prior selection for distributed lag estimation, the practical applicability of the single input single output model to the analysis of economic data is rather limited. This is due to the common existence of feedback between the input and output in econometric applications. In such a case multivariate time series modeling is required. Whether the smoothness prior can find a useful application in this case is a subject of further study.

ACKNOWLEDGEMENTS

This work is a continuation of the former work "Smoothness priors and the distributed lag estimator", Technical Report No. 40, Stanford University, 1979, which was supported in part by the Office of Naval Research Contract N00014-75-C-0442. By the introduction of the model selection procedure discussed in this paper some of the conclusions of the original work have been completely revised.
REFERENCES


**THE SELECTION OF SMOOTHNESS PRIORS FOR DISTRIBUTED LAG ESTIMATION**

In the application of Shillor's smoothness prior for distributed lag estimation, the main difficulty is the selection of hyperparameters of the prior distribution. In this paper, the use of a maximum likelihood procedure is proposed for this purpose and its performance is demonstrated by numerical examples.