SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR

BY

HIROTUGU AKAIKE

TECHNICAL REPORT NO. 40
AUGUST 1979

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited
SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR

By

Hirotugu Akaike

Technical Report No. 40
Aug 1979

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH
Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR

by

Hirotugu Akaike

1. INTRODUCTION

Shiller [4] introduced the concept of smoothness priors to develop a Bayesian estimator of the lag coefficients, or the sequence of impulse responses, of a linear system. In this approach the prior preference of the smoothness of the pattern of a set of lag coefficients is expressed by a spherical normal distribution for the fixed order differences of the lag coefficients. With an appropriate choice of the parameter of the smoothness prior it is found that Shiller's distributed lag estimator can produce results which satisfy our prior, or psychological, expectation of the smoothness of the pattern of the lag coefficients.

One practically significant problem in the application of Shiller's estimator is the choice of the variance of the spherical normal prior distribution. In this paper we propose a practical solution to this problem which is obtained by maximizing the likelihood of the Bayesian model with respect to the hyperparameter, the variance. A serious problem about Shiller's approach is whether the assumption of smoothness of the lag coefficients is a natural one. When the number of significant

1The author is grateful to Professor T. W. Anderson for the stimulus leading to the work reported in this paper. Thanks are due to Ms. E. Arahata for helping with the numerical computation. This work was partly supported by the Office of Naval Research Contract N00014-75-C-0442 in the Department of Statistics, Stanford University.

2The Institute of Statistical Mathematics, Tokyo.
coefficients is expected to be small, which is often the case with quarterly or annual economic data, this smoothness assumption may not be quite adequate.

A natural expectation for this type of situation seems to be the smoothness of the behavior of the system characteristic in the frequency domain. The integral of the squared absolute value of the derivative of the frequency response function will then provide a measure of smoothness. This measure takes a particularly simple form in terms of the lag coefficients and leads to the definition of a non-spherical normal smoothness prior.

2. SMOOTHNESS IN FREQUENCY DOMAIN

Consider the stochastic linear system

\[ y_n = \sum_{m=0}^{M} a_m x_{n-m} + w_n, \]

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 frequency response function of the output \( y_n \) to the input \( x_n \) is defined by

\[ A(f) = \sum_{m=0}^{M} a_m \exp(-i2\pi mf). \]
We propose to measure the smoothness of the frequency response function by

\[ R = \left( \left| \frac{dA(f)}{df} \right| \right)^2 \, df, \]

where \( \left| \frac{dA(f)}{df} \right| \) denotes the absolute value of \( \frac{dA(f)}{df} \) and the integral is over \(-0.5 \leq f \leq 0.5\). As is obvious from the definition, a large value of \( R \) means a rapid change of the system characteristic against the change of the frequency characteristic of the input. Thus, actually, \( R \) is a measure of unsmoothness of \( A(f) \). Since it holds that

\[ \frac{dA(f)}{df} = \sum_{m=0}^{M} (-i2\pi m)a_m \exp(-i2\pi mf), \]

we have

\[ R = (2\pi)^2 \sum_{m=1}^{M} m^2 a_m^2. \]

3. A SMOOTHNESS PRIOR

As is discussed in Akaike [2] we consider the minimization of the quantity

\[ \sum_{n=1}^{N} \left( y_n - \sum_{m=0}^{M} a_m x_{n-m} \right)^2 + \lambda^2 \sum_{m=1}^{M} m^2 a_m^2. \]
By a proper choice of the constant $\lambda$, this minimization leads to the
determination of $\alpha_m$ with small sum of squares of the residuals and
also with smooth behavior of the frequency response function.

The minimization problem is transformed into a statistical problem
by embedding it into the maximization problem of

$$
\exp \left( -\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left( y_n - \sum_{m=0}^{M} \alpha_m x_{n-m} \right)^2 \right) \exp \left( -\frac{\lambda^2}{2\sigma^2} \sum_{m=1}^{M} \alpha_m^2 \right).
$$

For a given $\sigma$, the first factor of (4) obviously represents the like-
lihood of the $\alpha_m$'s. The second factor represents our prior preference
on $\alpha_m$'s in the form of a normal distribution. Our preference on $\alpha_0$
and $\sigma$ is not specified in (4) and we assume the improper prior

$$
\frac{d\alpha_0}{\sigma} \frac{d\sigma}{\sigma}
$$

for $\alpha_0$ and $\sigma$. This choice is based on the consideration that it is
an ignorance prior of the parameters of a normal distribution $N(\alpha_0, \sigma)$,
where $c$ is a nonzero constant. For the justification of the use of
this ignorance prior distribution for its impartial performance, see
Akaike [1].

Our Bayesian model is thus defined by the data distribution

$$
p(y|\alpha, \sigma) = \left( \frac{1}{2\pi \sigma^2} \right)^{\frac{N}{2}} \exp \left( -\frac{1}{2\sigma^2} \| y - X\alpha \|^2 \right)
$$

and the (improper) prior distribution
(6) \[ p(\alpha, \sigma | \lambda) = \left( \frac{1}{2\pi \sigma^2} \right)^{\frac{M+1}{2}} |\lambda^2 D \alpha|^{\frac{1}{2}} \exp \left( - \frac{\lambda^2}{2\sigma^2} \alpha' D' D \alpha \right) \]

where \( y = (y_1, y_2, \ldots, y_N)' \), \( \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_M)' \) and \( X \) and \( D \) are respectively \( N \times (M+1) \) and \( M \times (M+1) \) matrices defined by

\[
X = \begin{bmatrix}
    x_1 & x_0 & \cdots & x_{-M+1} \\
    x_2 & x_1 & \cdots & x_{-M+2} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_N & x_{N-1} & \cdots & x_{N-M}
\end{bmatrix}, \quad D = \begin{bmatrix}
    0 & 1 & 0 & \cdots & 0 \\
    0 & 0 & 2 & \cdots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & 0 & \cdots & M
\end{bmatrix},
\]

and where \( \|y\| \) denotes the Euclidean norm of a vector \( y \) and \( |A| \) denotes the determinant of a matrix \( A \) and \( ' \) denotes transposition.

4. DETERMINATION OF THE HYPERPARAMETER

In his original work Shiller [4] demonstrated by a numerical example the insensitivity of his estimate to the choice of the hyperparameter of his smoothness prior. This hyperparameter corresponds to \( \lambda \) of (6).

We often consider this insensitivity as a proof of the robustness of the Bayesian model and take it as a justification for an arbitrary choice of the value of the hyperparameter. Nevertheless, it is only when we know the true lag coefficients that we can see that the estimates thus obtained do approximate the true coefficients. Thus it is desirable to have an objectively defined procedure for the selection of the value of the hyperparameter.
If the prior distribution \( p(\alpha, \sigma | \lambda) \) is proper, the likelihood of the hyperparameter \( \lambda \) with respect to the data \( y \) is defined by

\[
(7) \quad p(y | \lambda) = \int p(y | \alpha, \sigma) \ p(\alpha, \sigma | \lambda) \ d\alpha d\sigma .
\]

A reasonable choice of \( \lambda \) is realized by maximizing the likelihood \( p(y | \lambda) \). This procedure for the determination of a hyperparameter is called the method of Type II maximum likelihood by Good [3]. In our case the distribution \( p(\alpha, \sigma | \lambda) \) given by (6) contains an improper component. Nevertheless, for each particular situation, the improper component can be considered as an approximation to a very widely dispersed proper prior distribution. Thus we take (7) as the definition of the likelihood of the prior distribution defined by (6) and propose the use of \( \lambda \) which maximizes \( p(y | \lambda) \). Obviously it would be more desirable to develop an ignorance-type prior of \( \lambda \), but at present we leave this as a subject of further research.

For the present Bayesian model defined by (5) and (6) we have

\[
p(y | \alpha, \sigma) \ p(\alpha, \sigma | \lambda) = \left( \frac{1}{2\pi \sigma^2} \right)^{2N+M+1} \ | \lambda^2 D D' |^{\frac{1}{2}} \ \frac{1}{\sigma^2} \exp \left[ -\frac{1}{2\sigma^2} (\alpha - \alpha_0)' (X'X + \lambda^2 D'D)(\alpha - \alpha_0) \right] \exp \left[ -\frac{1}{2\sigma^2} S(\lambda) \right],
\]

where \( \alpha_0 = (X'X + \lambda^2 D'D)^{-1} X'y \) and \( S(\lambda) = \|y\|^2 - \alpha_0'(X'X + \lambda^2 D'D)\alpha_0 \). Thus we have
\[
p(y|\lambda) = \left(\frac{1}{2}\right)^{\frac{N}{2}} \frac{1}{\Gamma\left(\frac{N}{2}\right)} \left|\mathbf{X}'\mathbf{X} + \lambda^2 \mathbf{D}'\mathbf{D}\right|^{-\frac{1}{2}} \left|\lambda^2 \mathbf{D}'\mathbf{D}\right|^{-\frac{1}{2}}.
\]

By applying the formal Bayes procedure we get the posterior mean of \( \alpha \) and \( \sigma^2 \). They are given by \( \alpha_0 \) and \( S(\lambda)/(N-2) \), respectively. Computationally, \( \alpha_0 \) and \( S(\lambda) \) can be obtained directly by using the relation

\[(8) \quad S(\lambda) = \min_{\alpha} \left\| y^* - X^*\alpha \right\|^2 = \left\| y^* - X^*\alpha_0 \right\|^2,
\]

where

\[
y^* = +\begin{bmatrix} y \\ 0 \end{bmatrix} \quad \text{and} \quad X^* = +\begin{bmatrix} X \\ \lambda \mathbf{D} \end{bmatrix}.
\]

The search for the \( \lambda \) which maximizes \( p(y|\lambda) \) may practically be limited to a set of finite discrete values of \( \lambda \). Since we are familiar with the log likelihood ratio test statistic, we replace \( p(y|\lambda) \) by \((-2) \log p(y|\lambda)\) and try to minimize it. Ignoring an additive constant, \((-2) \log p(y|\lambda)\) is given by

\[(9) \quad L(\lambda) = N \log S(\lambda) + \log \left|\mathbf{X}'\mathbf{X} + \lambda^2 \mathbf{D}'\mathbf{D}\right| - \log \left|\lambda^2 \mathbf{D}'\mathbf{D}\right|.
\]

Akaike [2] reached the same procedure for the determination of \( \lambda \) by
using the maximum likelihood estimate of \( \sigma^2 \) instead of developing the improper prior distribution of \( \sigma \).

5. NUMERICAL EXAMPLES

To check the performance of our procedure it was applied to the data generated by the relation

\[
y_n = 1.2 x_n - 0.6 x_{n-1} + 0.4 x_{n-2} + \omega_n,
\]

where the input \( x_n \) was the same as that used in the second example of Shiller [4] and \( \omega_n \) was normal with mean zero and \( \sigma = 0.05 \). The length \( N \) of the time series was 40. The Bayesian model was developed with the highest lag \( M = 19 \). The search for the minimum of \( L(\lambda) \) was limited to the values \( \lambda = 10 \times 2^k \) (\( k = -9,-8,\ldots,-1 \)) and the minimum was attained with \( k = -5 \). The corresponding estimate \( \alpha_0 \) is given in Table 1 along with the true parameter and the estimate obtained by assuming the Shiller's smoothness prior for the second order differences \((\alpha_{i+2} - \alpha_{i+1}) - (\alpha_{i+1} - \alpha_i) \quad (i = 0,1,\ldots,M-2)\). The estimate denoted by Shiller was obtained by putting

\[
D = \begin{bmatrix}
1 & -2 & 1 \\
1 & -2 & 1 \\
& & \ddots & \ddots \\
& & & 1 & -2 & 1
\end{bmatrix}
\]

in (6) and then minimizing the corresponding \( L(\lambda) \) for \( \lambda = 10 \times 2^k \) (\( k = -5,-4,\ldots,10 \)). The minimum was attained at \( k = 10 \). Further increase of \( k \) would produce estimates smoother than that given in Table 1.
Comparison of estimates. $\alpha_0$ denotes the estimate obtained by the smoothness prior in the frequency domain. Shiller denotes the estimate obtained by assuming Shiller's prior for the second order differences.

<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>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1.2</td>
<td>-0.6</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>1.137</td>
<td>-0.440</td>
<td>0.230</td>
<td>0.050</td>
<td>0.017</td>
<td>0.012</td>
<td>-0.000</td>
<td>0.003</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td>Shiller</td>
<td>0.203</td>
<td>0.186</td>
<td>0.170</td>
<td>0.154</td>
<td>0.137</td>
<td>0.121</td>
<td>0.105</td>
<td>0.089</td>
<td>0.072</td>
<td>0.056</td>
<td>0.040</td>
<td>0.023</td>
<td>0.007</td>
<td>-0.009</td>
<td>-0.025</td>
</tr>
<tr>
<td>True</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.003</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.003</td>
<td>0.004</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td>Shiller</td>
<td>0.040</td>
<td>0.023</td>
<td>0.007</td>
<td>-0.009</td>
<td>-0.025</td>
<td>0.042</td>
<td>0.058</td>
<td>-0.074</td>
<td>-0.090</td>
<td>-0.048</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.000</td>
<td>-0.003</td>
</tr>
</tbody>
</table>
The result given in Table 1 demonstrates the necessity of great care in applying Bayesian models. Although the Shiller estimate is definitely in accordance with the prior preference of smoothness it is giving a definitely biased image of the true impulse response sequence. The sum of squared errors of an estimate $\alpha$ is defined by $\text{SSE} = \|\alpha - \tilde{\alpha}\|^2$, where $\alpha$ denotes the true parameter. The values of SSE are 0.06 and 1.77 for our estimate and the Shiller estimate, respectively. Another measure of inaccuracy of an estimate $\alpha$ is defined by the sum of squared errors of regression defined by

$$\text{SSER} = \|X(\alpha - \tilde{\alpha})\|^2.$$ 

The values of SSER are 0.0008 and 0.1955 for our estimate and the Shiller estimate, respectively. Thus the bias introduced by assuming Shiller's smoothness prior seems to be producing significant influence on the estimation of regression. The result shown in Table 1 is the first one obtained in a series of three experiments all of which produced similar results.

One might argue that the situation will be reversed when the present prior is applied to the case where Shiller's prior is appropriate. To check this possibility the estimates based on the two priors were obtained for a sample from the model treated in Shiller's second example. The result is shown in Table 2. For the purpose of comparison the ordinary least squares estimate, denoted by LS, is also included. The result demonstrates the extremely good performance of the Shiller estimate. This result is already reported in Akaike [2].
Table 2

Comparison of estimates. $a_0$ and Shiller are as defined in Table 1. LS denotes least squares estimate.

<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>.000</td>
<td>.000</td>
<td>.001</td>
<td>.004</td>
<td>.018</td>
</tr>
<tr>
<td>$a_0$</td>
<td>.004</td>
<td>-.005</td>
<td>-.011</td>
<td>.031</td>
<td>.037</td>
</tr>
<tr>
<td>Shiller</td>
<td>-.009</td>
<td>-.003</td>
<td>.004</td>
<td>.009</td>
<td>.017</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>$m$</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>$a_0$</td>
<td>-.006</td>
<td>.176</td>
<td>.208</td>
<td>.372</td>
<td>.378</td>
</tr>
<tr>
<td>Shiller</td>
<td>.051</td>
<td>.134</td>
<td>.242</td>
<td>.345</td>
<td>.395</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>$m$</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>$a_0$</td>
<td>.366</td>
<td>.260</td>
<td>.107</td>
<td>.055</td>
<td>.033</td>
</tr>
<tr>
<td>Shiller</td>
<td>.362</td>
<td>.257</td>
<td>.134</td>
<td>.052</td>
<td>.012</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>$m$</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>.000</td>
</tr>
<tr>
<td>$a_0$</td>
<td>-.016</td>
<td>-.020</td>
<td>.021</td>
<td>.012</td>
<td>-.003</td>
</tr>
<tr>
<td>Shiller</td>
<td>-.011</td>
<td>-.015</td>
<td>.006</td>
<td>.035</td>
<td>-.018</td>
</tr>
<tr>
<td>LS</td>
<td>-.018</td>
<td>-.050</td>
<td>.065</td>
<td>-.008</td>
<td>-.008</td>
</tr>
</tbody>
</table>
and is considered as a demonstration of the performance of our procedure for the selection of $\lambda$. The present result also shows that our estimate does not show any significant bias and that it is closer to the Shiller estimate than to the least squares estimate. The sum of squared errors of the estimate, SSE, is 0.0117 for our estimate, while it is 0.0026 for the Shiller estimate. Nevertheless, the sum of squared errors of regression, SSER, is 0.00047 for our estimate, while it takes a larger value 0.00051 for the Shiller estimate. This suggests that the use of the present prior in the situation where Shiller's prior is more appropriate may not necessarily induce significant loss in terms of the accuracy of regression.

6. DISCUSSION

The purpose of the present paper has been two-fold. The first is to introduce a new smoothness prior which might find a wider applicability than the Shiller's original smoothness prior. The second is to propose a practically useful objectively defined procedure for the choice of the hyperparameter. The numerical results given in the preceding section show that these objectives are attained fairly well.

One of the most significant results is the numerical demonstration of the possible significant bias due to the Bayesian modeling. In the Bayesian approach we specify our psychological expectation in the form of a prior distribution and proceed to get a result which satisfies this expectation. Our numerical results dramatically demonstrated that this satisfaction of psychological expectation may be quite
deceptive and misleading. Thus we see that no unique Bayesian model can be assured of a success in its practical application. The only reasonable procedure would then be to propose several possible Bayesian models and compare the likelihoods of these models, and, if necessary and feasible, develop a larger Bayesian model using these models.

Certainly the smoothness prior in the frequency domain introduced in this paper is not quite satisfactory to cover every practical situation. When some definite delay in the response is conceivable, i.e., when \( a_0, a_1, \ldots, a_d \) are expected to be close to zero for some \( d<\infty \), it may be more appropriate to assume a prior distribution over the possible delays \( d \) and, for each \( d \), replace the definition of \( x_n \) by \( x_{n-d-1} \). Practically it may be better to replace \( A(f) \) in the definition (3) of \( R \) by \( \exp(i2\pi f\delta) A(f) \) defined with some properly chosen integer \( \delta \). This replaces \( R \) of (3) by

\[
R_\delta = (2\pi)^2 \sum_{m=0}^{M} (m-\delta)^2 a_m^2 ,
\]

which clearly demonstrates the effect of the choice of \( \delta \). Since we do not know which value of \( \delta \) we should use, we will assume a prior distribution over a possible set of values of \( \delta \), such as \((-1, 0, 1, 2, 3)\), and define the "likelihood" of each model by

\[
L_\delta = \exp[-\frac{1}{2} \max L_\delta(\lambda)] ,
\]

where \( \max L_\delta(\lambda) \) denotes the maximum of \( L(\lambda) \) defined by (9) for a
specific choice of $\delta$. With a proper modification of $D$ in (6) we may even assume $\delta$ to be half integers, such as $-0.5, 0.5, 1.5, \ldots$. The practical utility of this type of approach needs further investigation.

What is practically more important than the refinement of the Bayesian model is the recognition of the limited applicability of the basic model (1) to economic time series. This is due to the fact that there is usually a feedback from the output $y_n$ to the input $x_n$. Only in the special situation where this feedback is negligible can we expect the use of the present model. Thus the result reported in this paper must be considered as only the second step following the first step of the original contribution of Shiller in making the Bayesian modeling a practical procedure. Further elaboration of the basic model is definitely necessary to make the procedure widely applicable to the analysis of complex economic time series.

REFERENCES


TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random

Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley,
July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal

6. "Estimation of Covariance Matrices with Linear Structure and Moving


8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul
Pedro Mentez, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classi-


10. "Asymptotic Evaluation of the Probabilities of Misclassification by


in the Mixed Model of the Analysis of Variance," John James Miller,


14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive
Process with Moving Average Residuals and Other Covariance Matrices

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series,
Steffen L. Lauritzen, April, 1974.


Office of Naval Research Contract N00014-75-C-0442 (NP-042-034)


**REPORT DOCUMENTATION PAGE**

1. **REPORT NUMBER**
   40

2. **GOVT ACCESSION NO.**
   ~

3. **RECIPIENT'S CATALOG NUMBER**
   ~

4. **TITLE (and Subtitle)**
   SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR

5. **AUTHOR(s)**
   HIROTUGU AKAIKE

6. **TECHNICAL REPORT NUMBER**
   ~

7. **PERFORMING ORGANIZATION NAME AND ADDRESS**
   DEPARTMENT OF STATISTICS
   STANFORD UNIVERSITY
   STANFORD, CALIFORNIA

8. **CONTRACT OR GRANT NUMBER(S)**
   N00014-75-C-0442

9. **PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS**
   (NR-042-034)

10. **DEPARTMENT OF THE NAVY RESEARCH AND DEVELOPMENT PROGRAM**
    OFFICE OF NAVAL RESEARCH
    STAT. & PROBABILITY PROGRAM CODE 436
    ARLINGTON, VA. 22217

11. **REPORT DATE**
    AUGUST 1979

12. **NUMBER OF PAGES**
    14

13. **SECURITY CLASS. (of this report)**
    UNCLASSIFIED

14. **SECURITY CLASS. (of this report)**
    UNCLASSIFIED

15. **DISTRIBUTION STATEMENT (of this report)**
    APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

16. **DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)**
    ~

17. **SUPPLEMENTARY NOTES**
    ~

18. **KEY WORDS (Continue on reverse side if necessary and identify by block number)**
    Smoothness priors, distributed lag, frequency response function, Bayesian estimation, hyperparameter, Type II maximum likelihood, Schiller's prior.

19. **ABSTRACT (Continue on reverse side if necessary and identify by block number)**
    SEE OTHER SIDE

**DD FORM**
1 JAN 73

**EIGHT OF 1 NOV 73 IS OBSOLETE**

5/16 010-014-0001

**UNCLASSIFIED**

**SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)
ABSTRACT

Shiller's distributed lag estimator based on a smoothness prior demonstrates the potential of the Bayesian approach to statistical model building. Nevertheless, when the number of significant lag coefficients is small the assumption of smoothness of the pattern of the lag coefficients may not be appropriate. In this paper, to cover such a situation, the smoothness is assumed for the behavior of the coefficients viewed in the frequency domain. This definition leads to a smoothness prior with a particularly simple form. Numerical result shows that the estimator based on this smoothness prior produces good estimates of the lag coefficients where Shiller's prior produces highly biased estimates. It is also observed that the new estimator produces reasonable results even when the Shiller's prior is more appropriate. The danger of introducing a bias by assuming a Bayesian model is stressed in the discussion.