INTRINSIC BAYES FACTORS FOR MODEL SELECTION WITH AUTOREGRESSIVE DATA

by

Julia A. Varshavsky
Purdue University

May 1995

*

Intrinsic Bayes Factors for Model Selection with Autoregressive Data *

JULIA A. VARSHAVSKY
Purdue University, USA

Abstract

One can not typically use standard noninformative priors for Bayesian model selection or for testing hypotheses of different dimensions, since such priors are only defined up to arbitrary constants which affect values of Bayes factors. A recently proposed model selection criterion, the Intrinsic Bayes Factor (IBF), overcomes this problem and performs well for a variety of situations involving sequences of IID data. The present paper suggests a modification of IBFs for model selection with dependent data structures, autoregressive models, in particular. Numerical problems arising in direct computations of the IBFs are presented and their possible solutions are discussed.

Keywords: IMPROPER PRIORS, TIME SERIES, BAYESIAN COMPUTATIONS.

1 Introduction

Suppose that we wish to choose a model M_i out of the the set of models $M_1, \ldots, M_p$ for the data Y . Let the data have the density $f_i(y|\theta_i)$ under model M_i, and $\pi(\theta_i)$ be the prior distribution of the parameter vector θ_i. The *Bayes factor of M_j to M_i* (BF) is a quantity defined by

$$B_{ji} = \frac{m_j(y)}{m_i(y)} = \frac{\int_{\Theta_j} f_j(y|\theta_j)\pi_j(\theta_j)\, d\theta_j}{\int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)\, d\theta_i} \, , \tag{1}$$

where m_i is called the marginal density of Y under M_i. If one uses noninformative priors $\pi_i^N(\theta_i)$, (1) becomes

$$B_{ji}^N = \frac{m_j^N(y)}{m_i^N(y)} = \frac{\int_{\Theta_j} f_j(y|\theta_j)\pi_j^N(\theta_j)\, d\theta_j}{\int_{\Theta_i} f_i(y|\theta_i)\pi_i^N(\theta_i)\, d\theta_i} \, . \tag{2}$$

Noninformative priors π_i^N are typically improper, and are thus defined only up to arbitrary constants. Hence the resultant Bayes factor, B_{ji}^N, is indeterminate. One way to overcome this difficulty is to consider part of the data, $y(l)$, as a so-called *training sample*, compute its

marginal, $m^N(y(l))$, with respect to the noninformative prior, and posterior $\pi_i^N(\theta_i|\mathbf{y}(l))$. One can then compute the Bayes factors with the remainder of the data, $y(-l)$, using $\pi_i^N(\theta_i|\mathbf{y}(l))$ as priors:

$$B_{ji}^\star(l) = \frac{\int_{\Theta_j} f_j(y(-l)|\theta_j, y(l))\pi_j^N(\theta_j|y(l))\,d\theta_j}{\int_{\Theta_i} f_i(y(-l)|\theta_i, y(l))\pi_i^N(\theta_i|y(l))\,d\theta_i} \; . \tag{3}$$

We can obviously do this only if the training sample marginal, $m^N(l)$, is proper. We will call a training sample $y(l)$ *minimal* (MTS) if its marginal is proper, and no subset of $y(l)$ results in a proper marginal.

These considerations led Berger and Pericchi (1993) to introduce the *Intrinsic Bayes factor* (IBF), one of the versions of which is given by

Definition 1. The (Arithmetic) Intrinsic Bayes factor is

$$B_{ji}^I = \frac{1}{L}\sum_{l=1}^{L} B_{ji}^\star(l) \, , \tag{4}$$

where L is the number of all possible minimal training samples. Other versions of the IBF are defined by using different types of averages. In this paper we will, however, concentrate on the arithmetic average version defined above. B_{ji}^I can be shown to be equal to $B_{ji}^N \cdot \frac{1}{L}\sum_{l=1}^{L} B_{ij}^N(y(l)) \stackrel{\text{def}}{=} B_{ji} \cdot \overline{B}_{ij}^{mts}$, where $B_{ij}^N(y(l)) \stackrel{\text{def}}{=} m_i^N(y(l))/m_j^N(y(l))$. We will refer to $\overline{B}_{ij}^{mts}$ as the *Correction factor* (CF) of the M_j to M_i comparison.

Numerous examples considered by Berger and Pericchi (1993) showed that the proposed model selection tool works as an "Ockham's Razor" in *iid* situations, that is, it selects the simpler model out of a set of models that describe the data equally well. A natural extension is to study the behavior of the IBF in dependent data situations. Here we consider the case of autoregressive models.

2 The Setup

Suppose that the data $Y_1, \ldots, Y_n$ comes from a stationary autoregressive process of order p ($p < n$) with a Gaussian white noise sequence:

$$Y_t - \mathbf{x}_t'\boldsymbol{\beta} - \phi_1(Y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta}) - \ldots - \phi_p(Y_{t-p} - \mathbf{x}_{t-p}'\boldsymbol{\beta}) = \epsilon_t \, , \tag{5}$$

where the ϵ_t are iid $N(0, \sigma^2)$ variables, $\mathbf{x}_t' = (x_t^1, \ldots, x_t^k)$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$. The parameters $\boldsymbol{\phi}_p = (\phi_1, \ldots, \phi_p)$, $\boldsymbol{\beta}$, and σ^2 are considered unknown, while the $\mathbf{x}_t$'s are known. The goal is to select the correct order of the autoregressive model, i.e. to select from the sequence of models:

$$M_p : \Phi^p(B)(Y_t - \mathbf{x}_{t,p}'\boldsymbol{\beta}) = \epsilon_t$$

$$\vdots \tag{6}$$

2

$$M_1 : \Phi^1(B)(Y_t - \mathbf{x}'_{t,1}\boldsymbol{\beta}) = \epsilon_t \, ,$$

where B is the backshift operator: $B^m Y_t = Y_{t-m}$, and $\Phi^m(z)$ is a polynomial defined by $\Phi^m(z) = 1 - \phi_1 z - \ldots - \phi_m z^m$. The second subscript in the regression design vectors $\mathbf{x}'_{t,i}$ reflects dependence on a particular model.

The Data Density Function. Writing down the design matrix for the regression component under model M_m as

$$\mathbf{X}_n^m = \begin{pmatrix} x_{1,m}^1 & \cdots & x_{1,m}^k \\ & \ddots & \\ x_{n,m}^1 & \cdots & x_{n,m}^k \end{pmatrix} = \begin{pmatrix} x'_{1,m} \\ \vdots \\ x'_{n,m} \end{pmatrix}$$

we can write the density of $Y_1, \ldots, Y_n$ under M_m:

$$L_n^m(Y_1, \ldots, Y_n) = (\frac{1}{2\pi\sigma^2})^{n/2} \mid M_n^m \mid^{1/2} \exp\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}_n^m\boldsymbol{\beta})' M_n^m (\mathbf{Y} - \mathbf{X}_n^m\boldsymbol{\beta})\} \quad (7)$$

where $\Sigma_n^m = \sigma^2 (M_n^m)^{-1}$ is the covariance matrix of the stationary AR(m) process, and the entries of M_n^m do not depend on σ.

The Choice of Priors. The autoregressive process (5) is called *causal* or *future-independent* if $Z_t = Y_t - \mathbf{x}'_t\boldsymbol{\beta}$ can be expressed in terms of the present and past errors, i.e. if $Z_t = \Sigma_{j=0}^{\infty} \epsilon_{t-j}$. Since prediction is our main goal here, towards which choosing the order of the process is the first step, we will be concerned only with values of ϕ_m that make the process causal. These regions (often called stationarity regions) can be shown (see, for example, Brockwell and Davis, 1991) to be: $\Phi_m = \{\phi_m : \Phi^m(z) = 0 \Rightarrow |z| > 1\}$. For model M_m, we will put noninformative uniform prior on the stationarity region Φ_m. Priors for $\boldsymbol{\beta}$ and σ are standard noninformative choices: uniform on the whole real line, and $\frac{1}{\sigma^{q+1}}$, respectively, where the choice of $q = 0$ in the latter would correspond to a reference prior on σ and $q = k$ to Jeffreys. Thus, under model M_m, $\pi^m(\boldsymbol{\beta}, \sigma, \phi_m) = \frac{1}{Volume(\Phi_m)} \cdot \frac{1}{\sigma^{q+1}} I_{\Phi_m}(\phi_m)$. The constant $\frac{1}{Volume(\Phi_m)}$ does not affect the value of B_{ji}^I. We have chosen to use it here, because then the expression for $\overline{B}_{ji}^{mts}$ (see Section 4) is simplified.

3 Modification of the Correction Term

In the time series setup it seems natural to preserve the time structure for the minimal training samples. So instead of considering all possible subsets of the data that make the marginals proper, we will consider only successive observations with this property. Let k be the MTS size, $Z^k(l) = (Z_l, \ldots, Z_{l+k-1})$, and $T_l^k = \frac{m_i(Z^k(l))}{m_j(Z^k(l))} \stackrel{\text{def}}{=} g^k(Z^k(l))$. Then the CF for the M_j to M_i comparison becomes:

$$\overline{B}_{ij}^{mts} = \frac{1}{n-k+1} \cdot \sum_{l=1}^{n-k+1} \frac{m_i(Z^k(l))}{m_j(Z^k(l))} = \frac{1}{n-k+1} \cdot \sum_{l=1}^{n-k+1} T_l^k \, . \quad (8)$$

3

An additional advantage of this modification is that the terms in the CF sum (under the stationarity assumption on the series) become identically distributed, and the strong law of large numbers holds. More precisely, the following is true:

Theorem 3.1 *Suppose* $Y_1, \ldots, Y_n$ *arises from one of the models* M_s *in (6),* $Z_t = Y_t - \mathbf{x}'_{t,s}\boldsymbol{\beta}$, $g^k : \Re^k \to \Re$ *is measurable, and that* $E\frac{m_i(Z^k(l))}{m_j(Z^k(l))}$ *exists (under* M_s*) . Then the Correction Factor* $\overline{B}_{ij}^{mts}$ *converges a.s. and in* L^1 *to* $E\frac{m_i(Z^k(l))}{m_j(Z^k(l))}$ *(under* M_s*) .*

Proof. Since the Z_l's come from a Gaussian AR process, the sequence $\{Z_n\}$ is ergodic (Hannan, 1970). Since it is also, by assumption, stationary, the sequence $\{T_l^k\}$ is ergodic and stationary (Durrett, 1991). Thus we can apply Birkhoff's Ergodic Theorem for the T_l^k's, which gives the desired result.

MTS size. A segment of data consisting of consecutive observations constitutes an MTS, if it is of length $k+1$ and the corresponding submatrix of the design matrix is of full rank. The first part of the statement follows from the fact that the prior on the autoregressive parameters is a finite measure and hence, the MTS size for the problem is equal to that for a regression problem with regression vector $\boldsymbol{\beta}$, normally distributed errors, and unknown σ (considered by Berger and Pericchi, 1993). If the submatrix of the design matrix is not of full rank, as we will see in the next section, the marginal is improper.

Example. Assume that the data $Y_1, \ldots, Y_n$ comes from an AR(m) with $\Phi^m(Y_t - \mu_1) = \epsilon_t$ for $t = 1, 2, \ldots, 10$ and $\Phi^m(Y_t - \mu_2) = \epsilon_t$ for $t = 11, \ldots, n$, where the ϵ_t are iid $N(0, \sigma^2)$. Then the MTS size is 3, and the only MTS's are: (Y_9, Y_{10}, Y_{11}), and (Y_{10}, Y_{11}, Y_{12}).

4   Expressions for Bayes Factors

Since each of the terms in the CF is by itself a Bayes factor, the derivation below applies for both the uncorrected BF (2) and a BF from the Correction factor term. Let $Y = (Y_1, \ldots, Y_r)$ be the data from a stationary AR(m) of type (5). Let Σ_r^m be the corresponding $r \times r$ covariance matrix, $M_r^m = \sigma^2 \cdot (\Sigma_r^m)^{-1}$, and let $\mathbf{X}_r^m$ be the corresponding $r \times k$ design matrix (k is the dimension of the regression vector $\boldsymbol{\beta}$). Also let $\hat{\boldsymbol{\beta}}_r^m = (\mathbf{X}_r' M_r^m \mathbf{X}_r^m)^{-1}(\mathbf{X}_r^m)' M_r^m \mathbf{Y}$, $R_r^m = (\mathbf{Y} - \mathbf{X}_r^m \hat{\boldsymbol{\beta}}_r^m)' M_r^m (\mathbf{Y} - \mathbf{X}_r^m \hat{\boldsymbol{\beta}}_r^m)$. The integration over $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ and σ can be carried out analytically, yielding

$$\frac{m_i}{m_j} = \frac{\int_{\Phi_i} \frac{\det(M_r^i)^{1/2}}{\det((X_r^i)'M_r^i X_r^i)^{1/2}} \cdot \left(\frac{1}{R_r^i}\right)^{(r-k+q)/2} d\phi_i}{\int_{\Phi_j} \frac{\det(M_r^j)^{1/2}}{\det((X_r^j)'M_r^j X_r^j)^{1/2}} \cdot \left(\frac{1}{R_r^j}\right)^{(r-k+q)/2} d\phi_j} . \tag{9}$$

Suppressing the subscripts m and r above, let $P_{\mathcal{X}}Y = X(X'X)^{-1}X'Y$ be the projection of Y onto the space spanned by the columns of X, and let (X, Y) be the $r \times (k+1)$ matrix, whose first k columns are the columns of X and the last column is Y.

Lemma 4.1

$$\det[(X, Y)'(X, Y)]^{1/2} = \det(X'X)^{1/2} \cdot |(I - P_{\mathcal{X}})Y| ,$$

where $|a|$ *denotes the Euclidian length of the vector* a.

Proof. Noting that $\det[(X,Y)'(X,Y)]^{1/2}$ is nothing but the volume of a hyperparallelepiped built on the columns of X and the vector Y, and that $\det(X'X)^{1/2}$ is the volume built on the columns of X only (Shilov, 1961), the lemma is simply saying that the volume of the hyperparallelepiped equals the product of the volume of its base and the height to the base.

Theorem 4.1 *The marginal of Y equals*

$$\frac{1}{Volume(\Phi_m)} \cdot \frac{1}{\det[(X,Y)'(X,Y)]^{(r-k+q)/2}} \int_{\Phi_m} \frac{\det(X'\Sigma_r^{-1}X)^{(r-k+q-1)/2}}{\det(\Sigma_r^{-1})^{(r-k+q-1)/2}} \, d\phi_m \; .$$

Proof. The residual sum of squares, R, can be written as:
$R = |\Sigma^{-1/2}Y - \Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y|^2 = |(I - P_{\Sigma^{-1/2}X})\Sigma^{-1/2}Y|^2$. Use of Lemma 4.1 yields

$$
\begin{aligned}
\det(\Sigma^{-1/2}) \cdot \det[(X,Y)'(X,Y)]^{1/2} &= \det[(\Sigma^{-1/2}X, \Sigma^{-1/2}Y)'(\Sigma^{-1/2}X, \Sigma^{-1/2}Y)]^{1/2} \\
&= \det(X'\Sigma^{-1}X)^{1/2} \cdot R^{1/2} \; .
\end{aligned}
$$

Combining the above and substituting into the expression for the marginal in (9), we get the result.

Corollary 4.1 *For the reference prior on σ, the correction factor is equal to*

$$\overline{B}_{ij}^{mts} = \frac{\det(X^i, Y)}{\det(X^j, Y)} \; . \tag{10}$$

If, in addition, the design matrices are equal, the CF equals 1.

Proof: For a minimal training sample $r = k + 1$, and hence the matrix (X, Y) becomes quadratic, so $\det[(X,Y)'(X,Y)]^{1/2} = \det(X,Y)$. Use of Theorem 4.1 with $q = 0$ and $r = k+1$ gives the result.

5 Computational Aspects

5.1 Integration over Φ_m

It was shown by Monahan (1984) that $\phi_m \in \Phi_m \iff \mathbf{r_m} \in D_m^{(-1,1)}$, where $\mathbf{r_m} = (r_1, \ldots, r_m)$ is the set of partial autocorrelations and $D_m^{(-1,1)}$ is the hypercube: $(-1,1)^m$. The transformation $\mathbf{r_m} \mapsto \phi_m$ is given recursively by $y_1^{(1)} = r_1$, $y_i^{(k)} = y_i^{(k-1)} - r_k \cdot y_{k-i}^{(k-1)}$, $(i = 1, \ldots, k - 1)$, $y_k^{(k)} = r_k$, $(k = 2, \ldots, m)$, and $\phi_m = (\phi_1, \ldots, \phi_m) = (y_1^{(m)}, \ldots, y_m^{(m)})$. The Jacobian of the transformation is $J(\mathbf{r_m}) = \Pi_2^m(1 - r_k)^{[k/2]}(1 + r_k)^{[(k-1)/2]}$ ($[x]$ denoting the integer part of x). It was also noticed (Jones, 1987) that $J(\mathbf{r_m})$ is nothing but a product of Beta ($[(k + 1)/2], [k/2] + 1$) densities on (-1,1) and thus, if we are seeking to evaluate integrals in (9) by generating uniform random variables on Φ_m, we can generate independent B(-1,1) instead. Using this idea, indeed, proved to be efficient for evaluating the marginals for minimal

5

training samples for $q \neq 0$, since if the MTS size is small, the integrand tends to be fairly spread out over the region of stationarity, and generating uniformly from it gives a quickly converging sum. However, when dealing with the whole sample, the integrand is likely to be concentrated around the MLE $\hat{\phi}_m$, and thus alternative methods such as importance sampling, should be used for the evaluation. It is worth noting that if $q = 0$, the calculation of the correction factor does not require any numerical computation (Corollary 4.1), which makes the use of the reference prior very appealing.

5.2   Computation of the Integrand

The entries m_{uv} of the matrix M_r^m in (6) for $r \geq m$ can be explicitly expressed in terms of the autoregressive coefficients $(\phi_1, \phi_2, \ldots, \phi_m)$ (Galbraith and Galbraith, 1974), so that the costly inversion of the covariance matrix Σ_r^m can be avoided. (Recall that the matrix M_r^m has to be computed for every point ϕ_m that is sampled). For $r \geq m$ $(1 \leq u \leq v \leq r)$, $m_{uv} = \Sigma_{j=0}^c \phi_j \phi_{j+v-u} - \Sigma_{j=d}^{m+u-v} \phi_j \phi_{j+v-u}$, where $c = \min\{u-1, m+u-v, r-v\}$, $d = \max\{u-1, r-v\}$. For $r < m$ one cannot bypass the inversion: $M_r^m \doteq A_r' A_r - (H'A_r)'(M_m^m + H'H)^{-1}(H'A_r)$, where the (i,j) elements of A_r and H are $-\phi_{i-j}$ $(i \geq j)$ and $-\phi_{m+i-j}$, respectively, with $\phi_0 = -1$ and the elements equal zero for $(i < j)$.

Computation of the determinant can also be reduced for $r \geq m$, since for an AR(m) process $\det(M_r^m) = \det(M_m^m)$.

5.3   Variance Reduction for the Correction Factor Computations

An estimate of the CF for comparing M_{k_2} to M_{k_1} $(k_1 < k_2)$ is

$$\hat{\bar{B}}_{k_1,k_2}^{mts} = \frac{1}{L} \cdot \sum_{l=1}^{L} \frac{\hat{m}_{k_1}(l)}{\hat{m}_{k_2}(l)} , \tag{11}$$

where $\hat{m}_{k_1}(l) = (1/m_1) \cdot \sum_{i=1}^{m_1} f_{k_1}(Y(l)|\theta_{k_1}^{(i)})$, $\hat{m}_{k_2}(l) = (1/m_2) \sum_{i=1}^{m_2} f_{k_2}(Y(l)|\eta_{k_2}^{(i)})$, $Y(l)$ is an MTS, f_{k_1} and f_{k_2} are the density functions under M_{k_1} and M_{k_2}, integrated over the β's and σ, and $\theta_{k_1}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_{k_1}^{(i)})$ and $\eta_{k_2}^{(i)} = (\eta_1^{(i)}, \ldots, \eta_{k_2}^{(i)})$ are the points sampled from Φ_{k_1} and Φ_{k_2}, respectively, at the ith iteration.

Variance reduction for each ratio in the sum (12) can be achieved if we make the numerator and denominator positively correlated, i.e. if we, for example, let $\theta_j^{(i)} = \eta_j^{(i)}$, $j = 1, \ldots, k_1$.

Additional variance reduction can be achieved if we monitor convergence of the entire sum (12), instead of the convergence of each of the individual terms $\frac{\hat{m}_{k_1}(l)}{\hat{m}_{k_2}(l)}$. Also, since for each of the points sampled, we have to compute the inverse of the autocovariance matrix and its determinant, one should use the *same* sample of points for all MTSs.

6 Example

In the example below we have evaluated the IBFs under reference and Jeffreys priors for the 100 Wölfer Sunspot Numbers (Brockwell and Davis, 1991). Twelve models were considered: AR(1) through AR(4), each with a constant mean, a linear drift and a quadratic trend. The AR(4) model with quadratic trend is a natural "encompassing" model for the rest of the models, and thus we can use the encompassing model approach (Berger, Pericchi, 1993) to compare the models. The idea of this approach is to choose a model M_0, within which the rest of the models are contained. Compute B_{0i}^I using definition (1) for the rest of the models. Then the *encompassing* IBF of M_j to M_i is defined as $B_{ji}^0 = B_{0i}^I / B_{0j}^I = B_{ji} \cdot (\overline{B}_{i0}^{mts} / \overline{B}_{j0}^{mts})$.

Minimal training samples have to be defined relative to *all* models, and thus consist here of 4 observations. The design matrices for the constant, linear, and quadratic versions are

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 100 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} 1 & 1 & 1 \\ & \vdots & \\ 1 & 100 & 100^2 \end{pmatrix}.$$

For the Jeffreys prior the computational approach described in Sections 5.1-5.3 was used. One thousand iterations were performed to evaluate the marginals. For reference priors no elaborate computations for MTSs were needed, since the CFs in this case depend only on the design matrices (Corollary 4.1).

Prior probabilities were assigned to the models according to the following scheme. The models were considered to belong to 3 classes: constant (1), linear (2), and quadratic (3). The prior probability $p_i^\star = i^{-1} / \Sigma_{j=1}^3 j^{-1}$, $i = 1, 2, 3$ was assigned to each class. Then within each class (i), a model AR(s) was assigned $p_s = p_i^\star * s^{-1} / \Sigma_{r=1}^4 r^{-1}$.

The posterior probabilities of the models were then computed as

$$P(M_i|Y) = (\Sigma_{j=1}^p \frac{p_j}{p_i} \cdot B_{ji}^0)^{-1} . \tag{12}$$

The table below contains the values of the prior and posterior probabilities for all the models. For comparison, the values of the standard Bayesian and non-Bayesian model selection criteria, the BIC and the AICC, a bias corrected version of the AIC (Hurvich and Tsai, 1989), are also presented. The star sign indicates the two best models chosen according to each criterion.

The AICC performs very poorly, selecting the most complex model. The BIC does significantly better, selecting an autoregressive model of order 4 with a constant trend, but still worse than the "IBF" criterion. According to the latter, we see that for both priors the AR(3) model with the constant mean has the highest posterior probability, the AR(2) model with the constant mean is the next best, while the probabilities for the rest of the models are smaller by at least a factor of 10. Diagnostic plots of residuals and autocorrelation function for these two models (not shown) indicate that both models provide a good fit, and hence the simpler should be preffered, in a complete accordance with the Ockham's razor principle.

Model	Prior	"IBF Posterior"		BIC	AICC
		Reference	Jeffreys		
AR(1), constant	.262	$\approx 10^{-10}$	$\approx 10^{-10}$	773.7	890.8
AR(2), constant	.131	.298*	.221*	753.9*	816.2
AR(3), constant	.087	.620*	.673*	755.5	806.6
AR(4), constant	.065	.047	.052	752.0*	798.0*
AR(1), linear	.131	$\approx 10^{-13}$	$\approx 10^{-13}$	789.7	892.5
AR(2), linear	.065	.007	$\approx 10^{-13}$	783.0	818.4
AR(3), linear	.044	.027	.050	785.3	808.1
AR(4), linear	.033	$\approx 10^{-3}$	$\approx 10^{-3}$	784.2	800.3
AR(1),quadratic	.087	$\approx 10^{-17}$	$\approx 10^{-16}$	830.4	894.5
AR(2),quadratic	.044	$\approx 10^{-5}$	$\approx 10^{-4}$	816.6	814.6
AR(3),quadratic	.029	$\approx 10^{-4}$	.002	814.3	807.5
AR(4),quadratic	.022	$\approx 10^{-5}$	$\approx 10^{-4}$	807.8	795.6*

Acknowledgments

References

Berger, J. O. and Pericchi, L.R. (1993). The Intrinsic Bayes Factor for Model Selection and Prediction. Technical report #93-43C, Department of Statistics, Purdue University, West Lafayette.

Berger, J. O. and Pericchi, L.R. (1994). The Intrinsic Bayes Factor for Linear Models. *Proceedings of the 5th Valencia International Conference on Bayesian Statistics*, to appear.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Spinger-Verlag, New York.

Durrett, R. (1991). *Probability: Theory and Examples*. Wadsworth, Belmont, Ca.

Galbraith, R. F. and Galbraith, J. I. (1974). On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series. *J. Appl. Prob.* **11**, 63-71.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.

Jones, M. C. (1987). Randomly choosing parameters from the stationarity and invertibility region of autoregressive moving average models. *Applied Statistics*, **36**, 134-138.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Monahan, J. F. (1984). A note on Enforcing Stationarity in Autoregressive-Moving Average Models. *Biometrika*, **71**, 2, 403-404.

Shilov, G. E. (1961). *An Introduction to the Theory of Linear Spaces*. Prentice-Hall, Englewood Cliffs, N.J.