AN ABSTRACT OF THE THESIS OF

LORETTA J. ROBB for the degree of DOCTOR OF PHILOSOPHY

in STATISTICS presented on June 4, 1979

Title: ESTIMATION OF THE ORDER OF AN AUTOREGRESSIVE TIME SERIES--A BAYESIAN APPROACH

Abstract approved: Redacted for Privacy

Fred L. Ramsey

Finite order autoregressive models for time series are often used for prediction and other inferences. Given the order of the model, the parameters of the models can be estimated by least squares, maximum likelihood, or the Yule-Walker method. The basic problem is estimating the order of the model. A number of statisticians have examined this problem. The most recent and widely accepted method was proposed by Akaike (1969, 1970, 1974), which has been shown to give quite accurate estimates for simulated data.

In this dissertation, the problem of autoregressive order estimation is placed in a Bayesian framework. This is done with the intent of illustrating how the Bayesian approach brings the numerous aspects of the problem together into a coherent structure which is both complementary to presently used methods and intuitively satisfying. A joint prior probability density is proposed for the order, the

partial autocorrelation coefficients and the variance, and the marginal posterior probability distribution for the order, given the data, is obtained. It is noted that the value with maximum posterior probability is the Bayes estimate of the order with respect to a particular loss function. The asymptotic posterior distribution of the order is also given.

In conclusion, Wolfer's sunspot data as well as simulated data corresponding to several autoregressive models are analyzed according to Akaike's method and the Bayesian method proposed in this dissertation. Both methods are observed to perform quite well, although the Bayesian method was clearly superior in most cases.

Estimation of the Order of an Autoregressive Time Series--A Bayesian Approach

by

Loretta J. Robb

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1980

APPROVED:

Redacted for Privacy

Associate Professor of Statistics

in charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented June 4, 1979

Typed by Clover Redfern for Loretta J. Robb

ACKNOWLEDGMENT

The completion of this thesis gives me great joy. It is a joy shared, I know, by my family, my professors and my close friends because a person living under the cloud of an "almost finished" dissertation often hides the sunnier side of his disposition. Thankfully, we have all endured.

> Now let us thank the Eternal Power: convinced
> That Heaven but tries our virtue by affliction, --
> That oft the cloud which wraps the present hour
> Serves but to brighten all our future days.
>
> John Brown, 1715-1766
> in Barbarossa. Act v. Sc. 3.

I especially wish to thank Dr. Fred Ramsey, my major professor, for not giving up on me, Dr. Dan Brunk and Dr. David Birkes for their extra help and encouragement at critical times, my roommate, Karen Swenson, who not only showed unreasonable patience in putting up with me for years without complaint, but who also took on the formidable job of typing the final rough draft, and Peter Thielman, who helped so generously with the computer simulations and proofreading.

There are many more people I wish to thank: the professors, fellow students, and acquaintances who have become dear friends and made me feel so at home.

I have made him my friend and now he is
unique in all the world.

Antoine de Saint Exupery
The Little Prince

Finally I wish to thank my parents for the love, encouragement, and patience they have always shown me. I dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ESTIMATION OF THE ORDER OF AN AUTOREGRESSIVE TIME SERIES--A BAYESIAN APPROACH

I. INTRODUCTION

1. The Autoregressive Model--Preliminaries

An autoregressive process $\{X_t\}$ is a linear stochastic process generated by a weighted sum of a finite number of the previous X's plus a random shock ε_t. The model for such a process may be written:

$$X_t = \mu + \alpha_1(X_{t-1}-\mu) + \alpha_2(X_{t-2}-\mu) + \ldots + \alpha_p(X_{t-p}-\mu) + \varepsilon_t, \quad (1.1.1)$$

$$\text{for} \quad t = 0, \pm 1, \pm 2, \ldots,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_p$ are the weight parameters, μ is either $E(X_t)$ if the process is stationary or merely a reference point for the level of the process if non-stationary, and $\{\varepsilon_t\}$ is a white noise sequence. That is, the ε_t are uncorrelated random variables, each with mean zero and finite variance σ^2. The process defined by (1.1.1) is called an autoregressive process of order p. Such a process will henceforth be called an $AR(p)$ process. It will be assumed that $\mu = 0$ in the following discussions.

A stochastic process is said to be stationary if its probability structure is unaffected by a shift in the time origin. Stationary

AR(p) processes satisfy the difference equations

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \ldots + \alpha_p\gamma_{k-p}, \quad k > 0 \tag{1.1.2}$$

where $\gamma_k = E[X_t X_{t+k}]$ is the k-lag autocovariance. By dividing (1.1.2) by $\gamma_0 = E[X_t^2]$, one has

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \ldots + \alpha_p\rho_{k-p}, \quad k > 0 \tag{1.1.3}$$

where $\rho_k = \gamma_k/\gamma_0$ is the k-lag autocorrelation. By substituting $k = 1, 2, \ldots, p$ into (1.1.3), one obtains a set of linear equations for $\alpha_1, \alpha_2, \ldots, \alpha_p$ in terms of $\rho_1, \rho_2, \ldots, \rho_p$. These p equations are usually referred to as the Yule-Walker equations:

$$\begin{aligned}
\rho_1 &= \alpha_1 + \alpha_2\rho_1 + \ldots + \alpha_p\rho_{p-1} \\
\rho_2 &= \alpha_1\rho_1 + \alpha_2 + \ldots + \alpha_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \alpha_1\rho_{p-1} + \alpha_2\rho_{p-2} + \ldots + \alpha_p .
\end{aligned} \tag{1.1.4}$$

An AR(p) process will be stationary if certain conditions on the α's are satisfied. Let the model (1.1.1) be written as

$$(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^p)(X_t) = \alpha(B)(X_t) = \varepsilon_t \tag{1.1.5}$$

where B is the backward shift operator, defined by $B^k X_t = X_{t-k}$ for integer k. Equation (1.1.1) has a stationary solution $\{X_t\}$

where X_t involves only past and present shocks, $\{\varepsilon_s,\ s \leq t\}$, if and only if the roots of the polynomial equation $\alpha(x) = 0$ all lie outside the unit circle in the complex plane (Box and Jenkins [7]). This translates into the conditions, for example,

when $p = 1$, $\quad -1 < \alpha_1 < 1,$

when $p = 2$, $\quad \alpha_2 + \alpha_1 < 1,$

$\alpha_2 - \alpha_1 < 1,$

and $\quad |\alpha_2| < 1,$

and when $p = 3$, $\quad \alpha_1 + \alpha_2 + \alpha_3 < 1,$

$-\alpha_1 + \alpha_2 - \alpha_3 < 1,$

$\alpha_3(\alpha_3 - \alpha_1) - \alpha_2 < 1,$

and $\quad |\alpha_3| < 1.$

As p increases, the stationarity conditions rapidly become even more complicated.

One way to get around the difficulty of insuring stationarity by the conditions given above is to parameterize the model in terms of the partial autocorrelations. Let φ_k be defined as the partial autocorrelation between X_s and X_{s+k} eliminating linear regressions on $X_{s+1}, \ldots, X_{s+k-1}$, where s is any integer and $k = 1, 2, \ldots$. Then an AR(p) is characterized by having $\varphi_k = 0$ for all $k > p$ (Ramsey [18]). It was shown by Barndorff-Nielsen and Schou [5] that there is a one-to-one mapping from $(\alpha_1, \ldots, \alpha_p)'$ to

$(\varphi_1, \ldots, \varphi_p)'$ which is both ways continuously differentiable so that all AR(p) models may be smoothly parameterized by $\underset{p}{\varphi} = (\varphi_1, \ldots, \varphi_p)'$. The stationarity conditions in terms of the φ's are simply that $-1 < \varphi_k < 1$, $k = 1, 2, \ldots, p$.

Durbin [10] proposed a procedure for sequentially solving the Yule-Walker equations (1.1.4) for the α's and φ's in terms of the ρ's. Let $\alpha_{p,k}$ denote the k^{th} α in an AR(p) model. The relevant equations are (see [18]):

$$
\begin{aligned}
&(1)\quad \varphi_1 = \alpha_{1,1} = \rho_1 \\
&(2)\quad \sigma_1^2 = 1 - \varphi_1^2 \\
&(3)\quad \varphi_{k+1} = \alpha_{k+1,k+1} = \left\{ \rho_{k+1} - \sum_{j=1}^{k} \alpha_{k,j}\rho_{k+1-j} \right\} / \sigma_k^2 \\
&(4)\quad \alpha_{k+1,j} = \alpha_{k,j} - \varphi_{k+1}\alpha_{k,k+1-j}, \quad j = 1, \ldots, k \\
&(5)\quad \sigma_{k+1}^2 = \sigma_k^2(1-\varphi_{k+1}^2)
\end{aligned}
\qquad (1.1.6)
$$

Then from (1.1.6), one may solve for the α's strictly in terms of the φ's. For example,

$\alpha_{1,1} = \varphi_1$ by (1),

$\alpha_{2,2} = \varphi_2$ by (3),

$\alpha_{2,1} = \alpha_{1,1} - \varphi_2\alpha_{1,1}$

$= \varphi_1 - \varphi_1\varphi_2$ by (4) and substitution,

$$\alpha_{3,3} = \varphi_3 \qquad \text{by (3),}$$

$$\alpha_{3,1} = \alpha_{2,1} - \varphi_3\alpha_{2,2}$$

$$= \varphi_1 - \varphi_1\varphi_2 - \varphi_2\varphi_3 , \qquad \text{and}$$

$$\alpha_{3,2} = \alpha_{2,2} - \varphi_3\alpha_{2,1}$$

$$= \varphi_2 - \varphi_3(\varphi_1 - \varphi_1\varphi_2) \qquad \text{by (4) and substitution.}$$

Therefore, the AR(1) model can be expressed as

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t ,$$

the AR(2) model as

$$X_t = \varphi_1(1-\varphi_2)X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t ,$$

and the AR(3) model as

$$X_t = (\varphi_1 - \varphi_1\varphi_2 - \varphi_2\varphi_3)X_{t-1} + (\varphi_2 - \varphi_1\varphi_3 + \varphi_1\varphi_2\varphi_3)X_{t-2} + \varphi_3 X_{t-3} + \varepsilon_t .$$

By parameterizing the model in terms of the φ's, one gains simplicity for the stationarity conditions but loses it in the expression of the model.

Now suppose the white noise sequence $\{\varepsilon_t\}$ of the AR(p) process $\{X_t\}$ is Gaussian as well as uncorrelated, and thus it is a sequence of normal, independent and identically distributed random

variables with mean zero and variance σ^2. The exact probability density of n consecutive observations from an AR(p) process is then (Box and Jenkins [7]):

$$f_p(\underline{x}_n | \underline{\alpha}_p, \sigma^2) = (2\pi\sigma^2)^{-n/2} |\underline{M}_p|^{1/2} \exp\{-H_p(\underline{\alpha}_p)/2\sigma^2\} \qquad (1.1.7)$$

where

$$\underline{x}_n = (x_1, x_2, \ldots, x_n)',$$

$$\underline{\alpha}_p = (\alpha_{p,1}, \alpha_{p,2}, \ldots, \alpha_{p,p})',$$

$$\underline{M}_p = \{\gamma_{|i-j|}\}^{-1}\sigma^2, \quad i, j = 1, 2, \ldots, p,$$

$$H_p(\underline{\alpha}_p) = \underline{\alpha}'_{1,p}\underline{D}_p\underline{\alpha}_{1,p},$$

$$\underline{\alpha}'_{1,p} = (1, \alpha_{p,1}, \alpha_{p,2}, \ldots, \alpha_{p,p}),$$

$$\underline{D}_p = \begin{bmatrix} D_{11} & -D_{12} & -D_{13} & \cdots & -D_{1,p+1} \\ -D_{12} & D_{22} & D_{23} & \cdots & D_{2,p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{1,p+1} & D_{2,p+1} & D_{3,p+1} & \cdots & D_{p+1,p+1} \end{bmatrix},$$

and

$$D_{ij} = D_{ji} = x_i x_j + x_{i+1} x_{j+1} + \cdots + x_{n+1-j} x_{n+1-i}.$$

In terms of the φ's, $|\underline{M}_p|^{1/2}$ has the uncomplicated form:

$$|\underline{M}_p|^{1/2} = \prod_{i=1}^{p} (1-\varphi_i^2)^{i/2} \qquad (1.1.8)$$

When parameterized with the φ's, $H_p(\underline{\alpha}_p)$ is an expression of nested quadratics in the φ's involving the D_{ij}'s. Let

$$f_p(\underline{x}_n | \underline{\varphi}_p, \sigma^2) = (2\pi\sigma^2)^{-n/2} |\underline{M}_p|^{1/2} \exp\{-K_p(\underline{\varphi}_p)/2\sigma^2\} \qquad (1.1.9)$$

denote the joint probability density of n observations given $\underline{\varphi}_p$ and σ^2. For example, letting $p = 2$,

$$H_2(\underline{\alpha}_2) = [1, \alpha_{2,1}, \alpha_{2,2}] \begin{bmatrix} D_{11} & -D_{12} & -D_{13} \\ -D_{12} & D_{22} & D_{23} \\ -D_{13} & D_{23} & D_{33} \end{bmatrix} \begin{bmatrix} 1 \\ \alpha_{2,1} \\ \alpha_{2,2} \end{bmatrix}$$

$$= K_2(\underline{\varphi}_2)$$

$$= [1, \varphi_1(1-\varphi_2), \varphi_2] \begin{bmatrix} D_{11} & -D_{12} & -D_{13} \\ -D_{12} & D_{22} & D_{23} \\ -D_{13} & D_{23} & D_{33} \end{bmatrix} \begin{bmatrix} 1 \\ \varphi_1(1-\varphi_2) \\ \varphi_2 \end{bmatrix}$$

$$= (D_{33} - 2D_{23}\varphi_1 + D_{22}\varphi_1^2)\varphi_2^2 - 2(D_{13} - (D_{12}+D_{23})\varphi_1 + D_{22}\varphi_1^2)\varphi_2$$

$$+ (D_{11} - 2D_{12}\varphi_1 + D_{22}\varphi_1^2),$$

which is a quadratic in φ_2 with coefficients which are quadratic in φ_1.

Let $\ell_p(\underline{\alpha}_p, \sigma | \underline{x}_n)$ denote the log likelihood function so that

$$\ell_p(\underline{\alpha}_p, \sigma | \underline{x}_n) = -(n/2) \ln \sigma^2 + (1/2) \ln |\underline{M}_p| - H_p(\underline{\alpha}_p)/(2\sigma^2). \qquad (1.1.10)$$

Maximum likelihood estimates (m.l.e.'s) are obtained for $\sigma^2, \alpha_{p,1}, \alpha_{p,2}, \ldots, \alpha_{p,p}$ by solving the equations

$$\partial \ell / \partial \sigma = -n/\sigma + H_p(\underline{\alpha})/\sigma^3 = 0 \qquad (1.1.11)$$

$$\partial \ell / \partial \alpha_{p,j} = M_j + \sigma^{-2}\{D_{1,j+1} - \alpha_{p,1} D_{2,j+1} - \ldots - \alpha_{p,p} D_{p+1,j+1}\} = 0,$$

$$\text{for} \quad j = 1, 2, \ldots, p, \qquad (1.1.12)$$

where $M_j = \partial\{(1/2)\ln|\underline{M}_p|\}/\partial \alpha_{p,j}$. Equation (1.1.11) yields the maximum likelihood estimator for σ^2 of

$$\hat{\sigma}^2 = H(\hat{\underline{\alpha}})/n. \qquad (1.1.13)$$

Unfortunately, the equations of (1.1.12) are not easily solved since the M_j are very complex functions of the α's.

One approximation to the exact m.l.e.'s of the α's results from ignoring the term of $\ell(\underline{\alpha}, \sigma|\underline{x}_n)$ involving $|\underline{M}_p|$ because $H_p(\underline{\alpha}_p)$ dominates $\ln|\underline{M}_p|$ for sufficiently large samples (Box and Jenkins [7]). Then the equations of (1.1.12) become

$$\underline{d}_p = \underline{D}_{p,1}\hat{\underline{\alpha}}_p \qquad (1.1.14)$$

where $\underline{d}_p = (D_{12}, D_{13}, \ldots, D_{1,p+1})'$, $\underline{D}_{p,1}$ is the same as $\underline{D}_p$ without the first row and first column. Then

$$\hat{\underline{\alpha}}_p = (\underline{D}_{p,1})^{-1}\underline{d}_p \qquad (1.1.15)$$

yields what are generally called the least squares estimates of $\underline{\alpha}_p$.

Another approximation can be obtained by first taking expectations in expression (1.1.12) and multiplying by σ^2. One obtains

$$M_j\sigma^2 + (n-j)\gamma_j - (n-j-1)\alpha_{p,1}\gamma_{j-1} - \cdots - (n-j-p)\alpha_{p,p}\gamma_{j-p} = 0. \tag{1.1.16}$$

If one then multiplies (1.1.2) by n (replacing k by j) and subtracts this from (1.1.16), one obtains

$$M_j\sigma^2 = j\gamma_j - (j+1)\alpha_{p,1}\gamma_{j-1} - \cdots - (j+p)\alpha_{p,p}\gamma_{j-p}. \tag{1.1.17}$$

Then substituting $\hat{\gamma}_{|j-i|} = D_{i+1,j+1}/(n-j-i)$ for $\gamma_{|j-i|}$ in (1.1.16) to estimate M_j, one obtains a set of p equations which can be written in the same matrix form as (1.1.14):

$$\underline{d}^*_p = \underline{D}^*_{p,1}\hat{\underline{\alpha}}_p \tag{1.1.18}$$

$$\hat{\underline{\alpha}}_p = (\underline{D}^*_{p,1})^{-1}\underline{d}^*_p \tag{1.1.19}$$

where $D^*_{i,j} = nD_{i,j}/\{n-(i-1)-(j-1)\}$. These estimates are commonly called "approximate maximum likelihood" estimates.

A third solution, which approximates the exact m.l.e.'s of the α's when the sample size n is moderate or large, is as follows. Let

$$c_k = 1/n \sum_{t=1}^{n-k} x_t x_{t+k} . \qquad (1.1.20)$$

Then for $k = |i-j|$, nc_k is a close approximation to D_{ij}. Substituting $S_{|i-j|} = nc_{|i-j|}$ for D_{ij} in the equations of (1.1.14) and dividing both sides by nc_0, one obtains the Yule-Walker equations with ρ_k replaced by $r_k = c_k/c_0$. (Note, according to Box and Jenkins [7] r_k is "the most satisfactory estimate of ρ_k" for moderate to large samples.) The Yule-Walker estimates of the α's are then obtained from

$$\underline{\hat{\alpha}}_p = \underline{R}_p^{-1} \underline{r}_p \qquad (1.1.21)$$

where $\underline{r}_p' = (r_1, r_2, \ldots, r_p)$, and

$$\underline{R}_p = \begin{bmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{bmatrix},$$

or one may use Durbin's procedure (1.1.6) to solve for the $\hat{\alpha}$'s sequentially with r_k replacing ρ_k.

To illustrate the differences in the three estimates, let $p = 1$. The least squares estimate is

$$\tilde{\alpha}_{1,1} = D_{12}/D_{22} .$$

The "approximate maximum likelihood" estimate is

$$\tilde{\tilde{\alpha}}_{1,1} = (n-2)D_{12}/[(n-1)D_{22}] .$$

The Yule-Walker estimate is

$$\hat{\alpha}_{1,1} = c_1/c_0 = r_1 = D_{12}/D_{11} .$$

When n is moderate or large, the differences in the estimates will be small. This is generally true for AR processes (Box and Jenkins [7]).

To estimate φ_j, $j = 1, 2, \ldots, p$ one would use one of the above three approximation methods to obtain estimates of $\alpha_{j,j}$, $j = 1, 2, \ldots, p$ (see equations (1.1.6)).

These three approximations to the exact maximum likelihood estimates, given the order p, are widely accepted. It is the estimation of the order p which is of primary interest here. The following section gives a historical review of the problem.

2. Review of Order Estimation of the Autoregressive Model

Some of the earliest procedures for estimating the order of an autoregressive process were proposed by Quenouille [17] and Bartlett

and Diananda [21]. Both procedures involve large sample goodness-of-fit tests of the null hypothesis that a time series can be represented by an AR(p) model with independent residuals against alternatives of AR(p+q) models. The tests are based on a sequence of forms which are linear functions of the sample autocorrelations, r_j. Under the null hypothesis, the forms are asymptotically independent, Normal (0, 1) random variables. For the Quenouille test, the sequence of forms may be expressed as

$$Q_t = (n-t)^{1/2}(\mathrm{var}(X_t)/\mathrm{var}(\varepsilon_t))H_t^2 r_{t+p}, \quad t = 1, 2, 3, \ldots, \qquad (1.2.1)$$

and for the Bartlett-Diananda test as

$$BD_t = (n-t)^{1/2}(\mathrm{var}(X_t)/\mathrm{var}(\varepsilon_t))H_{-t}H_t r_t, \quad t = 1, 2, 3, \ldots, \qquad (1.2.2)$$

where H_t is the linear operator $\alpha(B)$ defined in (1.1.5), $H_{-t} = \alpha(B^{-1})$.

Bartlett and Diananda pointed out that the Quenouille test might be more useful because whereas the asymptotic distribution of Q is unaltered by replacing the α_j's with efficient estimates $\hat{\alpha}_j = \alpha_j + O_p(n^{-1/2})$, $j = 1, 2, \ldots, p$, that of BD is not. Walker [21] derives the asymptotic power functions for the two test statistics above. He shows that although neither procedure dominates the other, a modification of the Q test results in a test which is always at

least as powerful as the BD test. The asymptotic distributions for the two tests is also given when the α_j, $j = 1, 2, \ldots, p$ are unspecified in the null hypothesis. He then generalizes these results to the case where the sequence of ε_t has a finite dependence. Walker concludes by comparing the likelihood ratio for obtaining large sample tests with the Quenouille tests.

Whittle [22] proposed using an AR(p) model for $0 \leq p \leq K$ to approximate the model of any purely non-deterministic process $\{z_t\}$, where K is prespecified. His test statistic for discriminating between an AR(p) model and an AR(p+q), where $K = p+q$, is

$$W = n[1-(1-\hat{\varphi}^2_{p+1})\ldots(1-\hat{\varphi}^2_{p+q})] \qquad (1.2.3)$$

where $\hat{\varphi}_i$ is the least squares estimate of φ_i. The statistic W has an approximate χ^2 distribution with q degrees of freedom. He notes that the value of K must be chosen with care since too small a value may not allow an adequate fit of the data and too large a value with respect to n results in greater deviation in the behavior of the test statistic from a χ^2 random variable.

Hannan [12] considers the same situation as Whittle except there is no prespecified maximum order K. He recommends the test statistic

$$F_{p,q} = \hat{R}^2(p+q|p) \cdot (n-p-q)/(q[1-\hat{R}^2(p+q|p)]), \qquad (1.2.4)$$

where $1 - \hat{R}^2(p+q|p) = \hat{\sigma}^2_{p+q}/\hat{\sigma}^2_p$ and $\hat{\sigma}^2_k$ is obtained from (1.1.13) using least squares estimates $\hat{\underline{\alpha}}$. The statistic $F_{p,q}$ has an asymptotic distribution of a χ^2 random variable with q degrees of freedom.

Box and Jenkins [7] suggest an informal procedure for identifying the appropriate model for an AR process. It begins with examining the first 20 or so sample autocorrelations and partial autocorrelations. The theoretical autocorrelations of AR processes tail off to zero, whereas the theoretical partial autocorrelations, φ_k are zero for k greater than the order p. The approximate standard error for estimates $\hat{\varphi}_k$, where k is greater than the hypothesized order p, is $\hat{\sigma}(\hat{\varphi}_k) = 1/n^{1/2}$. Also $z = \hat{\varphi}_k/\hat{\sigma}(\hat{\varphi}_k)$ has, approximately, a standard normal distribution under the hypothesis that the order is some value $p < k$. One might then take, as a preliminary estimate for the order, the value j where $\hat{\varphi}_k$ for $k > j$ lies within say two standard errors of zero, i.e., within the interval $(-2/n^{1/2}, 2/n^{1/2})$. One then estimates the autoregression coefficients $\alpha_{j,1}, \alpha_{j,2}, \ldots, \alpha_{j,j}$ by one of the three asymptotically equivalent methods given in Section I.1, and a "portmanteau" lack of fit test is applied to the residuals. Suppose one has the first K estimated autocorrelations for the residuals $r_i(\hat{\underline{\varepsilon}})$, from fitting a model with

order j. Then if the model is appropriate,

$$Q = n \sum_{i=1}^{K} r_i^2(\hat{\underline{\varepsilon}}) \qquad (1.2.5)$$

is approximately distributed as $\chi^2(K-j)$. If the model is inappropriate, the average values of Q will be inflated. If the initial model is found inadequate according to the test, an examination of the residuals is recommended for an indication of how to modify the model to obtain a better fit.

Anderson [4] advocates the following multiple decision approach. The minimum order m and maximum order p are prespecified. The null hypothesis H_i, $(i = m, m+1, \ldots, p)$, specifies that i is the correct order. A significance level is assigned to each null hypothesis in such a way that the probabilities of rejection are monotonically non-decreasing so that the probability of rejecting a more restricted null hypothesis when it is true is not less than that of rejecting a less restricted one when it is true. The sample space is then partitioned into $(p-m+1)$ similar regions such that if the sample points fall in region i, H_i is accepted.

Akaike [1, 2] advocates an entirely decision theoretic approach. Models of orders $p = 1, 2, \ldots, L$ are to be fitted by least squares methods. Let R_p denote the mean square residuals for order p, $p = 1, 2, \ldots, L$. Then the future prediction error for the model of

order p is

$$FPE_p = \left(\frac{n+p+k}{n-p-k}\right)R_p \qquad p = 1, 2, \ldots, L, \tag{1.2.6}$$

where k parameters are estimated in detrending the data. The value p corresponding to the minimum FPE_p is then selected as the order for the process. Jones [13] demonstrated that Akaike's FPE criterion performs very well in the simulations of a number of models. Akaike [3] later developed a more general information criterion for detecting deviations of the estimated parameters from the true values based on maximum likelihood estimates.

This more general information criterion is referred to as the AIC, Akaike's information criterion. For fitting stationary Gaussian autoregressive time series, the order estimate based on the AIC is asymptotically identical to the estimate obtained by the minimum FPE procedure. Shibata [20] obtained the asymptotic distribution of the order of regression selected by the AIC for stationary autoregressive time series. He also evaluated the asymptotic quadratic risks of estimates of the regression parameters when one uses Akaike's method. Schwarz [19] examines the problem of selecting one of a number of models of different dimensions by finding its Bayes solution and evaluating the leading terms of its asymptotic expansion. He compares his results with Akaike's criterion and concludes that

Akaike's cannot be asymptotically optimal, if one can agree with his (Schwarz's) underlying assumption.

Parzen [16] takes the point of view that for a general time series, the autoregressive representation is infinite, ($\alpha_\infty(B)X(t) = \varepsilon_t$ in (1.1.5)), and that the problem is to find an order p such that $\hat{\alpha}_p(B)$ estimates the true linear operator $\alpha_\infty(B)$ closely enough. The criterion for choosing the optimal order is to find the value $\hat{p}$ which minimizes the criterion autoregressive function (CAT_p) which involves the estimation of the theoretical infinite order variance σ_∞^2. When the procedure was applied to low order, finite autoregressive time series, the results agreed very closely with Akaike's FPE criterion.

The most recent method of autoregressive time series order estimation is one proposed by McClave [15]. He redefines the order of an autoregressive model as the number of non-zero α's in the model (1.1.1). For example, the model

$$X_t = \alpha_1 X_{t-1} + \alpha_{10} X_{t-10} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

would be identified as having order two, maximum lag ten. Using his subset regression technique (McClave [14]), one is to find the best fitting model of "order" $p = 1, 2, \ldots, K$, with maximum lag up to and including K, and calculate the statistics

$$\hat{M}_{p,1} = (n-p-1)(\hat{\sigma}_p^2 - \hat{\sigma}_{p+1}^2)/(\hat{\sigma}_{p+1}^2), \quad p = 0, 1, \ldots, K-1$$

where $\hat{\sigma}_p^2$ is the usual estimated residual variance from the fitted model of "order" p. Then for a given α, one determines d_p such that $Pr[M_{p,1} > d_p] = \alpha$ where $M_{p,1}$ is the maximum order statistic in a sequence of (K-p) independent $\chi^2(1)$ random variables. Finally, $\hat{p}$ is chosen as

$$\hat{p} = \min\{p: \hat{M}_{p,1} \leq d_p, \ 0 \leq p \leq K\}.$$

3. Organization of the Thesis

In Chapter II, a Bayesian approach to the estimation of the order of autoregressive time series is presented in a format similar to Halpern's Bayesian approach to order estimation in polynomial regression (Halpern [11]). In Section 2, a joint prior density for the order, the partial autocorrelations, and the white noise series variance is given. In Section 3, the marginal posterior probability density for the order p is derived. Section 3 also has a discussion of the asymptotic behavior of the marginal posterior probability density of p. And finally, a loss function is defined and the corresponding Bayes estimate for p is given.

Chapter III presents a comparison of the Bayesian method of Chapter II with Akaike's FPE criterion for estimating the orders of several autoregressive processes using computer simulations. Wolfer's sunspot data are also analyzed by these methods.

II. A BAYESIAN APPROACH TO ORDER ESTIMATION OF AUTOREGRESSIVE TIME SERIES

1. Introduction

In the Bayesian approach to estimation, the parameters of a model, $\underline{\theta}$, are regarded as random variables with marginal probability density function $f_{\underline{\Theta}}(\underline{\theta})$, called the prior probability density, with respect to some dominating measure ν. Cox and Hinkley [9] point out that this prior density can represent some prior information about the frequency of occurrence of the parameter values, or it can be a representation of "what is rational to believe" about the parameter values, given a situation of ignorance, or it may simply reflect one's subjective opinion of the relative likelihood of the various parameter values. The joint density of the parameters $\underline{\theta}$ is combined with the joint density of the data $\underline{x}$ to yield a joint conditional density for $\underline{\theta}$ given $\underline{x}$, called the joint posterior density for $\underline{\theta}$ given $\underline{x}$. If a loss function is defined, the Bayes estimate of $\underline{\theta}$ is then derived to be the value $\hat{\underline{\theta}}$ which minimizes the Bayes risk (expected loss) with respect to the posterior density.

The methods of order estimation of the autoregressive time series incorporating the partial autocorrelation function and partial variance are widely recognized and accepted. In this thesis, the problem of order estimation of the AR time series is approached from

the Bayesian point of view. This is done with the intent of illustrating how the Bayesian approach brings the numerous aspects of the problem together into a coherent structure which is both complementary to presently used methods and intuitively satisfying.

2. Joint Prior Density of the Order p, the Partial Autocorrelation, φ, and the White Noise Variance, σ^2

Let π_j denote the prior probability that the order of a stationary AR time series is j $(j = 0, 1, 2, \ldots, M)$, such that $\sum_{j=0}^{M} \pi_j = 1$, where M is a finite number less than n, the number of observations. That is,

$$g_p(j) = \pi_j, \qquad j = 0, 1, \ldots, M \tag{2.2.1}$$

is the marginal prior density for the order, p, with respect to counting measure. Let $g_{p,\Phi}(j, \underline{\varphi}_M)$ denote the following joint prior density of the order p and the partial autocorrelations $\underline{\varphi}_M$ with respect to (counting measure $\times$ Lebesgue measure):

$$g_{p,\Phi}(j, \underline{\varphi}_M) = \begin{cases} \pi_j 2^{-j} \prod_{k=j+1}^{M} \delta(\varphi_k), \underline{\varphi}_M \in (-1, 1)^M, & j = 0, 1, \ldots, M \\ 0, \quad \text{elsewhere} \end{cases} \tag{2.2.2}$$

where

$$\prod_{k=M+1}^{M} \delta(\varphi_k) \equiv 1,$$

and $\delta(\cdot)$ denotes the Dirac delta function (see Churchill [8]). This generalized function satisfies the expression:

$$\int_R f(x)\delta(x-a)dx = f(a) \tag{2.2.3}$$

for any integrable function f continuous at a, where R is any open set containing zero.

The joint marginal prior density of the order p with the φ's, as given in (2.2.2), allows for a formal way of combining conditional densities for the φ's, given the order, with the marginal prior probabilities for the order. For example, when $j = 0$, one obtains the product of the degenerate density with mass 1 at $\underset{\sim}{\varphi}_M = \underset{\sim}{0}_M$ and π_0, the prior probability that $p = 0$. When $j = 1$, one obtains the product of the uniform density over $(-1, 1)$ for φ_1 (where $\varphi_2 = \ldots = \varphi_M = 0$ almost surely), and π_1, the prior probability that $p = 1$. Continuing in this manner, eventually one obtains, for $j = M$, the product of the uniform density over $(-1, 1)^M$ for $\underset{\sim}{\varphi}_M$ and π_M, the prior probability that $p = M$. That $g_{p, \Phi}$ behaves like a density with respect to (counting measure $\times$ Lebesgue measure) is easily demonstrated:

$$\sum_{j=0}^{M} \int_{-1}^{1} \cdots \int_{-1}^{1} \int_{-1}^{1} g_{p,\Phi}(j, \underline{\varphi}_M) d\varphi_M d\varphi_{M-1} \cdots d\varphi_1$$

$$= \sum_{j=0}^{M} \int_{-1}^{1} \cdots \int_{-1}^{1} \int_{-1}^{1} \pi_j 2^{-j} \left[\prod_{k=j+1}^{M} \delta(k) \right] d\varphi_M d\varphi_{M-1} \cdots d\varphi_1$$

$$= \sum_{j=0}^{M} \pi_j 2^{-j} \int_{-1}^{1} \cdots \int_{-1}^{1} 1 \, d\varphi_j \cdots d\varphi_1$$

$$= \sum_{j=0}^{M} \pi_j = 1.$$

Note the very uncomplicated domain of the φ's in $g_{p,\Phi}$; the form of the domain does not depend on the value of M. Recall from Chapter I, Section 1, how the boundaries of stationarity for the α's changed drastically with the order. For example when $p = 1$, the condition on α_1 is $-1 < \alpha_1 < 1$. However, when $p = 2$, the condition on α_1 ($\alpha_{2,1}$ in present notation) is $-2 < \alpha_1 < 2$. The joint region for $(\alpha_{2,1}, \alpha_{2,2})$ is triangular. For higher order models the stationarity regions for the α's become extremely complicated to describe which makes defining the limits of integrations over these regions extremely difficult. Furthermore, a uniform density for the α's of a higher order model does not reduce to a uniform density for the α's of lower order models and so one cannot define consistent

uniform prior densities on the α's. The Bayesian approach to order estimation of autoregressive time series is intractable using the model parameterized with the α's. It was not until the development of the partial autocorrelations (the φ's) parameterization of the AR(p) model by Barndorff-Nielsen and Schou [5] and the characterization of the partial autocorrelation function by Ramsey [18] that the Bayesian approach to the order estimation problem became tractable.

To represent prior vagueness about the variance, we choose

$$g_{\underset{\sim}{\sigma}^{-2}}(\sigma^{-2}) \propto (\sigma^{-2})^{1/2}, \quad \sigma^{-2} > 0, \quad \text{zero elsewhere} \tag{2.2.4}$$

to be the marginal formal prior density for $\underset{\sim}{\sigma}^{-2}$. Note that we take $\underset{\sim}{\sigma}^{2}$ to be a priori independent of the order and the φ's.

The joint formal prior density with respect to (counting measure $\times$ Lebesgue measure) for the order p, the vector of partial autocorrelations $\underset{\sim}{\varphi}_M$ and the inverse of the white noise variance, $\underset{\sim}{\sigma}^{-2}$ is then

$$g_{p,\Phi,\underset{\sim}{\sigma}^{-2}}(j, \underset{\sim}{\varphi}_M, \sigma^{-2}) = \begin{cases} \pi_j 2^{-j}\left[\prod_{k=j+1}^{M} \delta(\varphi_k)\right](\sigma^{-2})^{1/2}, & \underset{\sim}{\varphi}_M \in (-1,1)^M, \\ & j = 0, 1, \ldots, M, \ \sigma^{-2} > 0 \\ 0, \quad \text{elsewhere} \end{cases} \tag{2.2.5}$$

3. Marginal Posterior Probability Density of the Order p Given the Data $\underset{\sim}{X}_n$

Let $\underset{\sim}{x}_n = (x_1, \ldots, x_n)'$ be a vector of n consecutive observations from a stationary AR(p) time series with p unknown and with Gaussian white noise sequence $\underset{\sim}{\varepsilon}_n$. Recall from Chapter I, Section 1, that the exact probability density with respect to Lebesgue measure on $\underset{\sim}{X}_n$ given p, $\underset{\sim}{\Phi}_p$ and $\underset{\sim}{\sigma}^{-2}$ is $f_p(\underset{\sim}{x}_n | \underset{\sim}{\varphi}_p, \sigma^{-2})$ as given in (1.1.9). By multiplying $f_M(\underset{\sim}{x}_n | \underset{\sim}{\varphi}_M, \sigma^{-2})$ and the joint formal prior density of the parameters $g_{p, \Phi, \underset{\sim}{\sigma}^{-2}}(j, \underset{\sim}{\varphi}_M, \sigma^{-2})$ as given in (2.2.5) above, one obtains the joint probability density of the data and the parameters with respect to (counting measure $\times$ Lebesgue measure), which is denoted by $f_{\underset{\sim}{X}, p, \Phi, \underset{\sim}{\sigma}^{-2}}(\underset{\sim}{x}_n, j, \underset{\sim}{\varphi}_M, \sigma^{-2})$. [Note that if the true order is $p < M$, then $\varphi_{p+1} = \ldots = \varphi_M = 0$ and so $f_M(\underset{\sim}{x}_n | \underset{\sim}{\varphi}_M, \sigma^{-2}) = f_p(\underset{\sim}{x}_n | \underset{\sim}{\varphi}_p, \sigma^2)$.]

With respect to estimating the order p, the variance $\underset{\sim}{\sigma}^2$ is a nuisance parameter. The marginal joint density of the data $\underset{\sim}{X}_n$, the order p and the partial autocorrelations, Φ_M is obtained by the following integration with respect to $\underset{\sim}{\sigma}^{-2}$:

$$f_{\underset{\sim}{X}, p, \Phi}(\underset{\sim}{x}_n, j, \underset{\sim}{\varphi}_M) = \int_0^\infty f_{\underset{\sim}{X}, p, \underset{\sim}{\Phi}, \underset{\sim}{\sigma}^{-2}}(\underset{\sim}{x}_n, j, \underset{\sim}{\varphi}_M, \sigma^{-2}) d\sigma^{-2} \qquad (2.3.1)$$

$$= G_n \pi_j 2^{-j} \left[\prod_{k=j+1}^{M} \delta(\varphi_k) \right] \left[\prod_{i=0}^{M} (1-\varphi_i^2)^{i/2} \right] [2^{-1} K_M(\underset{\sim}{\varphi}_M)]^{-\gamma_n}$$

for $\underline{x}_n \in R^n$, $j = 0, 1, \ldots, M$, $\underline{\varphi}_M \in (-1, 1)^M$, where $G_n = (2\pi)^{-n/2}\Gamma(\gamma_n)$, $\gamma_n = (n+3)/2$, $\varphi_0 \equiv 0$, $\prod_{k=M+1}^{M} \delta(\varphi_k) \equiv 1$, and $K_M(\underline{\varphi}_M)$ is defined as in (1.1.9).

The marginal joint density of the data $\underline{X}_n$ and the order p with respect to (Lebesgue measure $\times$ counting measure), $f_{\underline{X}, p}(\underline{x}_n, j)$ is found by integrating $f_{\underline{X}, p, \Phi}(\underline{x}_n, j, \underline{\varphi}_M)$ with respect to Φ_M over the region $(-1, 1)^M$. The marginal density of $\underline{X}_n$ alone with respect to Lebesgue measure, denoted $f_{\underline{X}}(\underline{x}_n)$, is found by integrating $f_{\underline{X}, p}(\underline{x}_n, j)$ with respect to p over $\{0, 1, \ldots, M\}$, i.e., summing $f_{\underline{X}, p}(\underline{x}_n, j)$ over $j = 0, 1, \ldots, M$. Then the posterior density with respect to counting measure of the order p given the data $\underline{X}_n$ is (as shown in equation (2.3.2) on page 27): where $\varphi_0 \equiv 0$, $\prod_{k=M+1}^{M} \delta(\varphi_k) \equiv 1$, and $\gamma_n = (n+3)/2$. The first expression for $\pi_{p|\underline{X}}(j|\underline{x}_n)$ follows from the definition, the second from (2.3.1) and from dividing numerator and denominator by $G_n = (2\pi)^{-n/2}\Gamma(\gamma_n)$.

Let I_j denote the numerator of (2.3.2) for $j = 0, 1, \ldots, M$. Notice that the denominator of (2.3.2) is $\sum_{j=0}^{M} I_j$, so that $\pi_{p|\underline{X}}(j|\underline{x}_n)$ is clearly a density function with respect to counting measure. Furthermore, to find an explicit expression for $\pi_{p|\underline{X}}(j|\underline{x}_n)$, one need only evaluate I_j for $j = 0, \ldots, M$.

$$\pi_{p|\underline{X}}(j|\underline{x}_n) = \frac{\int_{-1}^{1}\cdots\int_{-1}^{1} f_{\underline{X},\,p,\,\Phi}(\underline{x}_n,\,j,\,\underline{\varphi}_M)d\varphi_M\cdots d\varphi_1}{\sum_{\ell=0}^{M}\int_{-1}^{1}\cdots\int_{-1}^{1} f_{\underline{X},\,p,\,\Phi}(\underline{x}_n,\,\ell,\,\underline{\varphi}_M)d\varphi_M\cdots d\varphi_1}$$

$$= \left\{\left[\int_{-1}^{1}\cdots\int_{-1}^{1}\pi_j 2^{-j}\left[\prod_{k=j+1}^{M}\delta(\varphi_k)\right]\left[\prod_{i=0}^{M}(1-\varphi_i^2)^{i/2}\right]\right.\right.$$

$$\left.\times\,[2^{-1}K_M(\underline{\varphi}_M)]^{-\gamma_n}d\varphi_M\cdots d\varphi_1\right]$$

$$\div\left[\sum_{\ell=0}^{M}\int_{-1}^{1}\cdots\int_{-1}^{1}\pi_\ell 2^{-\ell}\left[\prod_{k=\ell+1}^{M}\delta(\varphi_k)\right]\left[\prod_{i=0}^{M}(1-\varphi_i^2)^{i/2}\right]\right.$$

$$\left.\left.\times\,[2^{-1}K_M(\underline{\varphi}_M)]^{-\gamma_n}d\varphi_M\cdots d\varphi_1\right]\right\}, \qquad (2.3.2)$$

for $j = 0, 1, \ldots, M$.

Recall that to obtain the Yule-Walker estimates of the α's and φ's for the AR(p) model the D_{ij}'s in equations (1.1.14) were replaced by

$$S_{|i-j|} = \sum_{k=1}^{n-|i-j|} x_k x_{k+|i-j|}$$

To simplify the evaluation of the I_j's, $S_{|i-j|}$ will replace D_{ij} in the function $K_j(\underline{\varphi}_j)$. For moderate to large n, the difference

between D_{ij} and $S_{|i-j|}$ is slight. Then, according to (1.1.7) and (1.1.9),

$$K_j(\underline{\varphi}_j) = H_j(\underline{\alpha}_j) = \underline{\alpha}'_{1,j}\underline{S}_j\underline{\alpha}_{1,j}(1+O_p(n^{-1})), \quad \text{(Anderson [4])},$$

where

$$\underline{\alpha}_{1,j} = (1, \alpha_{j,1}, \ldots, \alpha_{j,j})'$$

and

$$\underline{S}_j = \begin{bmatrix} S_0 & -S_1 & -S_2 & \cdots & -S_j \\ -S_1 & S_0 & S_1 & \cdots & S_{j-1} \\ \vdots & \vdots & \vdots & & \vdots \\ -S_j & S_{j-1} & S_{j-2} & \cdots & S_0 \end{bmatrix} \tag{2.3.3}$$

It then follows that the I_j's can be expressed as follows:

$$I_0 = \int_{-1}^{1} \cdots \int_{-1}^{1} \pi_0 \left[\prod_{k=1}^{M} \delta(\varphi_k)\right]\left[\prod_{i=0}^{M} (1-\varphi_i^2)^{i/2}\right]$$

$$\times [2^{-1}K_M(\underline{\varphi}_M)]^{-\gamma_n} d\varphi_M \cdots d\varphi_1$$

$$= \pi_0\left[\prod_{i=0}^{M} (1-0)^{i/2}\right][2^{-1}K_M(\underline{0}_M)]^{-\gamma_n}$$

by definition of the Dirac delta function

$$= \pi_0[2^{-1}D_{11}]^{-\gamma_n} = \pi_0[2^{-1}S_0]^{-\gamma_n}$$

and

$$I_j = \int_{-1}^{1} \cdots \int_{-1}^{1} \pi_j 2^{-j} \left[\prod_{k=j+1}^{M} \delta(\varphi_k) \right] \left[\prod_{i=0}^{M} (1-\varphi_i^2)^{i/2} \right]$$

$$\times \left[2^{-1} K_M(\underline{\varphi}_M) \right]^{-\gamma_n} d\varphi_M \cdots d\varphi_1$$

$$= \int_{-1}^{1} \cdots \int_{-1}^{1} \pi_j 2^{-j} \left[\prod_{i=0}^{j} (1-\varphi_i^2)^{i/2} \right] \left[2^{-1} K_j(\underline{\varphi}_k) \right]^{-\gamma_n} d\varphi_j \cdots d\varphi_1$$

Let $\tilde{I}_j$ be defined by

$$\tilde{I}_j = \int_{-1}^{1} \cdots \int_{-1}^{1} \pi_j 2^{-j} \left[\prod_{i=0}^{j} (1-\varphi_i^2)^{i/2} \right] \left[2^{-1} \tilde{K}_j(\underline{\varphi}_j) \right]^{-\gamma_n} d\varphi_j \cdots d\varphi_1 \qquad (2.3.4)$$

where $\tilde{K}_j(\underline{\varphi}_j) = \tilde{H}_j(\underline{\alpha}_j) = \underline{\alpha}'_{1,j} \underline{S}_j \underline{\alpha}_{1,j}$. Then $\tilde{I}_j = I_j(1+O_p(n^{-1}))$.

In order to evaluate $\tilde{I}_j$, induction will be used, and so, it is necessary to express $\tilde{K}_{j+1}(\underline{\varphi}_{j+1})$ in terms of $\tilde{K}_j(\underline{\varphi}_j)$. The following lemma does this.

Lemma 2.1.

$$\tilde{K}_{j+1}(\underline{\varphi}_{j+1}) = \tilde{K}_j(\underline{\varphi}_j)(1+\varphi_{j+1}^2) - 2L_j(\underline{\varphi}_j)\varphi_{j+1}, \quad j = 1, 2, \ldots, M-1,$$

where

$$\tilde{K}_j(\underline{\varphi}_j) = \tilde{H}_j(\underline{\alpha}_j) = \underline{\alpha}'_{1,j} \underline{S}_j \underline{\alpha}_{1,j}$$

and

$$L_j(\underline{\varphi}_j) = G_j(\underline{\alpha}_j) = S_{i+1} + \sum_{i=1}^{j}\sum_{k=1}^{j} \alpha_{j,i}\alpha_{j,j+1-k}S_{k-i} - 2\sum_{i=1}^{j}\alpha_{j,i}S_{j+1-i}.$$

Proof: For $j = 0, 1, 2, \ldots, M-1$

$$\tilde{K}_{j+1}(\underline{\varphi}_{j+1}) = \tilde{H}_{j+1}(\underline{\alpha}_{j+1}) = \underline{\alpha}'_{1,\,j+1}\underline{S}_{j+1}\underline{\alpha}_{1,\,j+1}$$

Recall from (1.1.6) that $\varphi_{j+1} = \alpha_{j+1,\,j+1}$ and $\alpha_{j+1,\,i} = \alpha_{j,\,i} - \varphi_{j+1}\alpha_{j,\,j+1-i}$ for $i = 1, 2, \ldots, j$. Then $\underline{\alpha}'_{1,\,j+1}$ can be expressed as follows:

$$\underline{\alpha}'_{1,\,j+1} = [1, \varphi_{j+1}]\Lambda'_j$$

where

$$\Lambda'_j = \begin{bmatrix} \underline{\alpha}'_{1,\,j} & 0 \\ -\underline{\check{\alpha}}'_{0,\,j} & 1 \end{bmatrix},$$

$$\underline{\alpha}_{1,\,j} = (1, \alpha_{j,\,1}, \ldots, \alpha_{j,\,j})',$$

and

$$-\underline{\check{\alpha}}_{0,\,j} = (0, -\alpha_{j,\,j}, \ldots, -\alpha_{j,\,1})'.$$

Let Λ_j and $\underline{S}_{j+1}$ be partitioned in the following manner:

$$\Lambda'_j = \left[\begin{array}{c|c} \underline{\alpha}'_{1,\,j} & 0 \\ \hline -\underline{\check{\alpha}}'_{0,\,j} & 1 \end{array}\right], \qquad \underline{S}_{j+1} = \left[\begin{array}{c|c} \underline{S}_j & \underline{\check{s}}_{j+1} \\ \hline \underline{\check{s}}'_{j+1} & S_0 \end{array}\right],$$

where

$$\underset{\sim}{s}'_{j+1} = (-S_{j+1}, S_j, \ldots, S_1).$$

Then

$$\tilde{K}_{j+1}(\underset{\sim}{\varphi}_{j+1})$$

$$= [1, \underset{\sim}{\varphi}_{j+1}]\Lambda'_j \underset{\sim}{S}_{j+1} \Lambda_j \begin{bmatrix} 1 \\ \underset{\sim}{\varphi}_{j+1} \end{bmatrix}$$

$$= [1, \underset{\sim}{\varphi}_{j+1}] \left[\begin{array}{c|c} \underset{\sim}{\alpha}'_{1,j} \underset{\sim}{S}_j \underset{\sim}{\alpha}_{1,j} & \underset{\sim}{\alpha}'_{1,j} \underset{\sim}{S}_j (-\underset{\sim}{\breve{\alpha}}_{0,j}) + \underset{\sim}{\alpha}'_{1,j} \underset{\sim}{\breve{s}}_{j+1} \\ \hline -\underset{\sim}{\breve{\alpha}}'_{0,j} \underset{\sim}{S}_j \underset{\sim}{\alpha}_{1,j} + \underset{\sim}{\breve{s}}'_{j+1} \underset{\sim}{\alpha}_{1,j} & S_0 + 2\underset{\sim}{\breve{s}}'_{j+1} \underset{\sim}{\alpha}_{1,j} + (-\underset{\sim}{\breve{\alpha}}'_{0,j} \underset{\sim}{S}_j (-\underset{\sim}{\breve{\alpha}}_{0,j})) \end{array}\right] \begin{bmatrix} 1 \\ \underset{\sim}{\varphi}_{j+1} \end{bmatrix}$$

$$= [1, \underset{\sim}{\varphi}_{j+1}] \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} 1 \\ \underset{\sim}{\varphi}_{j+1} \end{bmatrix}$$

Now

$$A_{11} = \underset{\sim}{\alpha}'_{1,j} \underset{\sim}{S}_j \underset{\sim}{\alpha}_{1,j} = \tilde{H}_j(\underset{\sim}{\alpha}_j) = \tilde{K}_j(\underset{\sim}{\varphi}_j),$$

$$A_{22} = (-\underset{\sim}{\breve{\alpha}}'_{0,j}) \underset{\sim}{S}_j (-\underset{\sim}{\breve{\alpha}}_{0,j}) + \underset{\sim}{\breve{s}}'_{j+1} (-\underset{\sim}{\breve{\alpha}}_{0,j}) + (-\underset{\sim}{\breve{\alpha}}'_{0,j}) \underset{\sim}{\breve{s}}_{j+1} + S_0$$

$$= S_0 - 2 \sum_{k=1}^{j} \alpha_{j,k} S_k + \sum_{i=1}^{j} \sum_{k=1}^{j} \alpha_{j,i} \alpha_{j,k} S_{|i-k|}$$

$$= \tilde{H}_j(\underset{\sim}{\alpha}_j) = \tilde{K}_j(\underset{\sim}{\varphi}_j),$$

and

$$A_{12} = A_{21} = -\underset{\sim}{\alpha}'_{0,j}\underset{\sim}{S}_j\underset{\sim}{\alpha}_{1,j} + \underset{\sim}{s}'_{j+1}\underset{\sim}{\alpha}_{1,j}$$

$$= -S_{j+1} + 2\sum_{k=1}^{j}\alpha_{j,k}S_{j+1-k} + \sum_{k=1}^{j}\sum_{i=1}^{j}\alpha_{j,i}\alpha_{j,j+1-k}S_{|k-i|}$$

$$= -G_j(\underset{\sim}{\alpha}_j) = -L_j(\underset{\sim}{\varphi}_j).$$

Therefore,

$$\tilde{K}_{j+1}(\underset{\sim}{\varphi}_{j+1}) = [1, \varphi_{j+1}]\begin{bmatrix} \tilde{K}_j(\underset{\sim}{\varphi}_j) & -L_j(\underset{\sim}{\varphi}_j) \\ -L_j(\underset{\sim}{\varphi}_j) & \tilde{K}_j(\underset{\sim}{\varphi}_j) \end{bmatrix}\begin{bmatrix} 1 \\ \varphi_{j+1} \end{bmatrix}$$

$$= \tilde{K}_j(\underset{\sim}{\varphi}_j)(1+\varphi_{j+1}^2) - 2L_j(\underset{\sim}{\varphi}_j)\varphi_{j+1} \cdot \quad \square$$

In the proof for the general form of $\tilde{I}_j$, $\tilde{K}_{j+1}(\underset{\sim}{\varphi}_{j+1})$ is expressed as

$$\tilde{K}_{j+1}(\underset{\sim}{\varphi}_{j+1}) = \tilde{K}_j(\underset{\sim}{\varphi}_j)[(\varphi_{j+1} - L_j(\underset{\sim}{\varphi}_j)/\tilde{K}_j(\underset{\sim}{\varphi}_j))^2 + 1 - (L_j(\underset{\sim}{\varphi}_j)/\tilde{K}_j(\underset{\sim}{\varphi}_j))^2]$$

by completing the square in the expression in Lemma 2.1. In the following lemma one finds that the Yule-Walker estimate of φ_{j+1} is obtained by evaluating $L_j(\underset{\sim}{\varphi}_j)/\tilde{K}_j(\underset{\sim}{\varphi}_j)$ at $\hat{\underset{\sim}{\varphi}}_j$, the Yule-Walker estimates of $\varphi_1, \ldots, \varphi_j$.

Lemma 2.2. If $\tilde{K}_j(\underset{\sim}{\varphi}_j)$ and $L_j(\underset{\sim}{\varphi}_j)$ are as defined in Lemma 2.1, and $\hat{\varphi}_1, \ldots, \hat{\varphi}_{j+1}$ are the Yule-Walker estimates of $\varphi_1, \ldots, \varphi_{j+1}$, then

$$L_j(\hat{\underline{\varphi}}_j) = S_0\left[\hat{\rho}_{j+1} - \sum_{k=1}^{j} \hat{\alpha}_{j,k}\hat{\rho}_{j+1-k}\right],$$

$$\tilde{K}_j(\hat{\underline{\varphi}}_j) = S_0 \prod_{i=1}^{j}(1-\hat{\varphi}_j^2)$$

and so

$$\hat{\varphi}_{j+1} = L_j(\hat{\underline{\varphi}}_j)/\tilde{K}_j(\hat{\underline{\varphi}}_j), \quad \text{for} \quad j = 0, 1, \ldots .$$

Proof: Let $j = 0$. Then

$$L_0(\hat{\varphi}_0) = G_0(\hat{\alpha}_0) = S_1$$

and

$$\tilde{K}_0(\hat{\varphi}_0) = \tilde{H}_0(\hat{\alpha}_0) = S_0$$

so

$$L_0(\hat{\varphi}_0)/\tilde{K}_0(\hat{\varphi}_0) = S_1/S_0 = \hat{\varphi}_1 ,$$

where $\hat{\varphi}_0 \equiv \hat{\alpha}_0 \equiv 0$. Therefore, the lemma is true for $j = 0$.

Let $j = 1$.

$$L_1(\hat{\varphi}_1) = G_1(\hat{\alpha}_1) = S_2 + \hat{\alpha}_{1,1}^2 S_0 - 2\hat{\alpha}_{1,1}S_1$$

$$= S_0[\hat{\rho}_2 + \hat{\alpha}_{1,1}^2 - 2\hat{\alpha}_{1,1}\hat{\rho}_1],$$

since $\hat{\rho}_k = S_k/S_0$,

$$= S_0[\hat{\rho}_2 - \hat{\alpha}_{1,1}\hat{\rho}_1] ,$$

since $\hat{\alpha}_{1,1} = \hat{\rho}_1$.

Also

$$\tilde{K}_1(\hat{\varphi}_1) = \tilde{H}_1(\hat{\alpha}_1) = S_0 - 2S_1\hat{\varphi}_1 + S_0\hat{\varphi}_1^2$$

$$= S_0[1-\hat{\varphi}_1^2] ,$$

since $\hat{\varphi}_1 = \hat{\alpha}_{1,1} = S_1/S_0$.

Then

$$L_1(\hat{\varphi}_1)/\tilde{K}_1(\hat{\varphi}_1) = [\hat{\rho}_2 - \hat{\alpha}_{1,1}\hat{\rho}_1]/[1-\hat{\varphi}_1^2]$$

$$= \hat{\varphi}_2 \qquad \text{according to (1.1.6).}$$

Therefore, the lemma is true for $j = 1$.

Assume that the lemma is true for an arbitrary order j:

$$L_j(\underline{\hat{\varphi}}_j) = S_0\left[\hat{\rho}_{j+1} - \sum_{k=1}^{j} \hat{\alpha}_{j,k}\hat{\rho}_{j+1-k}\right] ,$$

$$\tilde{K}_j(\underline{\hat{\varphi}}_j) = S_0 \prod_{i=1}^{j} (1-\hat{\varphi}_i^2) ,$$

and so

$$\hat{\varphi}_{j+1} = L_j(\underline{\hat{\varphi}}_j)/\tilde{K}_j(\underline{\hat{\varphi}}_j) .$$

Now consider the $(j+1)$ case:

$$L_{j+1}(\hat{\underline{\varphi}}_{j+1}) = G_{j+1}(\hat{\underline{\alpha}}_{j+1})$$

$$= S_{j+2} + \sum_{i=1}^{j+1} \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,i} \hat{\alpha}_{j+1,j+2-k} S_{|k-i|}$$

$$- 2 \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,k} S_{j+2-k}$$

$$= S_0 \left[\hat{\rho}_{j+2} + \sum_{i=1}^{j+1} \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,i} \hat{\alpha}_{j+1,j+2-k} \hat{\rho}_{|k-i|} \right.$$

$$\left. - 2 \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,k} \hat{\rho}_{j+2-k} \right]$$

$$= S_0 \left[\hat{\rho}_{j+2} - \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,k} \hat{\rho}_{j+2-k} \right] ,$$

since $\hat{\rho}_k = S_k / S_0$ and $\sum_{i=1}^{p} \hat{\alpha}_{p,i} S_{|k-i|} = S_k$. Also

$$\tilde{K}_{j+1}(\hat{\underline{\varphi}}_{j+1}) = \tilde{K}_j(\hat{\underline{\varphi}}_j)(1+\hat{\varphi}_{j+1}^2) - 2L_j(\hat{\underline{\varphi}}_j)\hat{\varphi}_{j+1}$$

$$= \tilde{K}_j(\hat{\underline{\varphi}}_j)[1-\hat{\varphi}_{j+1}^2] , \qquad \text{since} \quad \hat{\varphi}_{j+1} = L_j(\hat{\underline{\varphi}}_j)/\tilde{K}_j(\hat{\underline{\varphi}}_j)$$

$$= S_0 \prod_{i=1}^{j+1} (1-\hat{\varphi}_i^2) \qquad \text{by substitution.}$$

Then

$$L_{j+1}(\hat{\underline{\varphi}}_{j+1})/\tilde{K}_{j+1}(\hat{\underline{\varphi}}_{j+1}) = \left[\hat{\rho}_{j+2} - \sum_{k=1}^{j+1} \hat{\alpha}_{j+1,k}\hat{\rho}_{j+2-k}\right] \Big/ \left[\prod_{i=1}^{j+1}(1-\hat{\varphi}_i^2)\right]$$

$$= \hat{\varphi}_{j+2} \qquad \text{by (1.1.6).}$$

Therefore, the lemma is true for all j, by induction. □

The evaluation of each $\tilde{I}_j$ involves a j-fold integral. At each step, one has an integral of the form found in the following lemma. Throughout, when we have sequences of variables $\{x_n\}$ and $\{y_n\}$ and we write $x_n \doteq y_n$, we mean that $x_n = y_n(1+O(n^{-1}))$ as $n \to \infty$ and that the order is "in probability" when the sequences are random sequences.

Lemma 2.3: If $K(\varphi) = A(1+\varphi^2) - 2B\varphi$, where $|B| < A$, $A > 0$, and if $g(\varphi)$, continuous on $[-1, 1]$, has a Taylor series expansion which converges on $(-1, 1)$, then

$$\int_{-1}^{1} g(\varphi)(K(\varphi))^{-\gamma_n} d\varphi = [2\pi/(2\gamma_n+1)]^{1/2}(1-\varphi_0^2)^{-\gamma_n+1/2} A^{-\gamma_n} g(\varphi_0)$$

$$\times [1+O(n^{-1})]$$

where $\gamma_n = (n+3)/2$ and $\varphi_0 = B/A$.

Proof: Let $h(\varphi) = [K(\varphi)]^{-\gamma_n}$. Then

$$h'(\varphi) = -\gamma_n[K'(\varphi)]^{-(\gamma_n+1)} K'(\varphi),$$

and

$$h''(\varphi) = \gamma_n(\gamma_n+1)[K'(\varphi)]^{-(\gamma_n+2)}[K'(\varphi)]^2 - \gamma_n[K(\varphi)]^{-(\gamma_n+1)} K''(\varphi).$$

Note that $h(\varphi)$ has a maximum where $h'(\varphi) = 0$, which is where $\varphi = B/A = \varphi_0$. Then let $h(\varphi)$ be written as

$$h(\varphi) = A^{-\gamma_n}\{(\varphi-\varphi_0)^2+1-\varphi_0^2\}^{-\gamma_n}.$$

Note also that $h(\varphi)$ has inflection points where $h''(\varphi) = 0$, which is where

$$(\varphi-\varphi_0)^2 = (1-\varphi_0^2)/(2\gamma_n+1).$$

Now let z be defined by:

$$(\varphi-\varphi_0) = [(1-\varphi_0^2)/(2\gamma_n+1)]^{1/2} z.$$

Then

$$K(\varphi) = A(1-\varphi_0^2)[1+z^2/(2\gamma_n+1)]$$

and so

$$h(\varphi) = A^{-\gamma_n}(1-\varphi_0^2)^{-\gamma_n}[1+z^2/(2\gamma_n+1)]^{-\gamma_n}$$

$$\doteq A^{-\gamma_n}(1-\varphi_0^2)^{-\gamma_n}([1+z^2/n]^n)^{-1/2},$$

since $2\gamma_n + 1 = n + 4 \doteq n$ for large n. Then

$$h(\varphi) \doteq A^{-\gamma_n}(1-\varphi_0)^{-\gamma_n} e^{-z^2/2},$$

since

$$\lim_{n \to \infty} [1+x/n]^n = e^x .$$

So

$$h(\varphi) \doteq (2\pi)^{1/2} A^{-\gamma_n}(1-\varphi_0^2)^{-\gamma_n} \eta(z),$$

where $\eta(z) = (2\pi)^{-1}\exp\{-z^2/2\}$ is the standard normal probability density function. Then

$$\int_{-1}^{1} g(\varphi)h(\varphi)d\varphi \doteq Q_n \int_{a_n}^{b_n} g(\varphi_0+\lambda_n z)\eta(z)dz,$$

where

$$Q_n = (2\pi)^{1/2} A^{-\gamma_n}(1-\varphi_0^2)^{-\gamma_n}\lambda_n ,$$

$$\lambda_n = [(1-\varphi_0^2)/(2\gamma_n+1)]^{1/2}$$

$$a_n = -(1+\varphi_0)/\lambda_n ,$$

$$b_n = (1-\varphi_0)/\lambda_n .$$

Now using a Taylor expansion with remainder term,

$$g(\varphi_0+\lambda_n z) = g(\varphi_0) + g'(\varphi_0)\lambda_n z + g''(\zeta)\lambda_n^2 z^2/2!$$

where $\varphi_0-\lambda_n z < \zeta < \varphi_0+\lambda_n z$, and so

$$\int_{a_n}^{b_n} g(\varphi_0+\lambda_n z)\eta(z)dz = \int_{a_n}^{b_n} [g(\varphi_0)+g'(\varphi_0)\lambda_n z+g''(\zeta)\lambda_n^2 z^2]\eta(z)dz$$

$$= g(\varphi_0)[\Phi(b_n)-\Phi(a_n)] + \int_{a_n}^{b_n} g'(\varphi_0)\lambda_n z\eta(z)dz$$

$$+ \int_{a_n}^{b_n} g''(\zeta)\lambda_n^2 z^2\eta(z)dz,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. As n increases, the first term becomes

$$\lim_{n\to\infty} g(\varphi_0)[\Phi(b_n)-\Phi(a_n)] = \lim_{n\to\infty} g(\varphi_0)[\Phi([1-\varphi_0][1-\varphi_0^2]^{1/2}[n+4]^{1/2})$$

$$-\Phi(-[1+\varphi_0][1-\varphi_0^2]^{1/2}[n+4]^{1/2})]$$

$$= g(\varphi_0)[\Phi(\infty)-\Phi(-\infty)]$$

$$= g(\varphi_0).$$

If one multiplies the second term by n and allows n to increase, one obtains

$$\lim_{n\to\infty}\int_{a_n}^{b_n} ng'(\varphi_0)\lambda_n z\eta(z)dz$$

$$= \lim_{n\to\infty} g'(\varphi_0)(1-\varphi_0^2)^{1/2}\int_{a_n}^{b_n} n(n+4)^{-1/2} z\eta(z)dz$$

$$\doteq g'(\varphi_0)(1-\varphi_0^2)^{1/2}c^{-1}\lim_{n\to\infty}\int_{a_n}^{b_n} |z|z\eta(z)dz,$$

for some c, since

$$a_n = -(1+\varphi_0)(1-\varphi_0^2)^{-1/2}(n+4)^{1/2} \doteq d\cdot n^{1/2}$$

and

$$b_n = (1-\varphi_0)(1-\varphi_0^2)^{-1/2}(n+4)^{1/2} \doteq e\cdot n^{1/2}$$

and so

$$|z| \doteq c\cdot n^{1/2}$$

$$= g'(\varphi_0)(1-\varphi_0^2)^{1/2}c^{-1}\lim_{n\to\infty}\left[\int_{a_n}^{0}|z|z\eta(z)dz+\int_{0}^{b_n}|z|z\eta(z)dz\right]$$

$$= g'(\varphi_0)(1-\varphi_0^2)^{1/2}c^{-1}\left[\int_{-\infty}^{0}|z|z\eta(z)dz+\int_{0}^{\infty}|z|z\eta(z)dz\right]$$

$$= g'(\varphi_0)(1-\varphi_0^2)^{1/2}c^{-1}[-1/2+1/2],$$

since $\int_{-\infty}^{\infty} z^2\eta(z)dz = 1,$

$$= 0$$

and so the second term is $o(n^{-1})$. If one also multiplies the third term by n and allows n to increase, one obtains

$$\lim_{n\to\infty}\left[n\lambda_n^2/2\right]\int_{a_n}^{b_n} g''(\zeta)z^2\eta(z)dz$$

$$= \left[(1-\varphi_0^2)/2\right]\cdot\left[g''(\zeta)\right]\cdot\left[\lim_{n\to\infty} n/(n+4)\right]\cdot\left[\lim_{n\to\infty}\int_{a_n}^{b_n} z^2\eta(z)dz\right]$$

$$= \left[(1-\varphi_0^2)g''(\zeta)/2\right]\cdot[1]\cdot\left[\int_{-\infty}^{\infty} z^2\eta(z)dz\right]$$

$$= (1-\varphi_0^2)g''(\zeta)/2,$$

a constant, and so the third term is $O(n^{-1})$. Therefore

$$\int_{-1}^{1} g(\varphi)(K(\varphi))^{-\gamma_n}d\varphi \doteq Q_n g(\varphi_0)[1+O(n^{-1})]$$

$$= [2\pi/(2\gamma_n+1)]^{1/2}(1-\varphi_0^2)^{-\gamma_n+1/2}A^{-\gamma_n}g(\varphi_0)$$

$$\times [1+O(n^{-1})]. \quad \square$$

Now with the use of these lemmas, one can evaluate the $\tilde{I}_j$'s of (2.3.4).

Theorem 2.4. Let

$$\tilde{I}_j = \int_{-1}^{1}\cdots\int_{-1}^{1}\pi_j 2^{-j}\left[\prod_{k=j+1}^{M}\delta(\varphi_k)\right]\left[\prod_{i=0}^{M}(1-\varphi_i^2)^{i/2}\right][2^{-1}K_M(\underline{\varphi}_M)]^{-\gamma_n}d\varphi_M\cdots d\varphi_1,$$

for $j = 0, 1, \ldots, M$.

Then

$$\tilde{I}_j = (S_0/2)^{-\gamma_n}\pi_j 2^{-j}[2\pi/(2\gamma_n+1)]^{j/2}\left[\prod_{i=0}^{j}(1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}\right][1+O(n^{-1})],$$

where $\hat{\varphi}_0 \equiv 0$, $\hat{\varphi}_i$ is the Yule-Walker estimate for φ_i, $i = 1, \ldots, M$, $\gamma_n = (n+3)/2$ and $S_0 = \sum_{i=1}^{n} x_i^2$.

Proof: Let $j = 0$. Then

$$\tilde{I}_0 = \int_{-1}^{1}\cdots\int_{-1}^{1}\pi_0 2^{-0}\left[\prod_{i=1}^{M}\delta(\varphi_i)\right]\left[\prod_{i=0}^{M}(1-\varphi_i^2)^{1/2}\right]$$

$$\times [2^{-1}\tilde{K}_M(\underline{\varphi}_M)]^{-\gamma_n} d\varphi_M \cdots d\varphi_1$$

$$= \pi_0 \cdot 1 \cdot \left[\prod_{i=0}^{M}(1-0^2)^{i/2}\right][2^{-1}\tilde{K}_M(\underline{0})]^{-\gamma_n}, \quad \text{by (2.2.3)}$$

$$= \pi_0[S_0/2]^{-\gamma_n},$$

since $\tilde{K}_M(\underline{0}_M) = \tilde{H}_M(\underline{0}_M) = S_0$ by substitution into (2.3.4). Therefore, the theorem is true for $j = 0$.

Let $j = 1$. Then

$$\tilde{I}_1 = \int_{-1}^{1}\cdots\int_{-1}^{1}\pi_1 2^{-1}\left[\prod_{i=2}^{M}\delta(\varphi_i)\right]\left[\prod_{i=0}^{M}(1-\varphi_i^2)^{i/2}\right]$$

$$\times [2^{-1}\tilde{K}_M(\underline{\varphi}_M)]^{-\gamma_n} d\varphi_M \cdots d\varphi_1 =$$

$$= \int_{-1}^{1} \pi_1 2^{-1+\gamma_n} (1-\varphi_1^2)^{1/2} [\tilde{K}_1(\varphi_1)]^{-\gamma_n} d\varphi_1 \quad \text{by (2.2.3)}$$

$$= \pi_1 2^{-1+\gamma_n} \int_{-1}^{1} (1-\varphi_1^2)^{1/2} [S_0(1+\varphi_1^2) - 2S_1\varphi_1]^{-\gamma_n} d\varphi_1 \ , \quad \text{by substitution}$$

$$= \pi_1 2^{-1+\gamma_n} (2\pi)^{1/2} (S_0)^{-\gamma_n}$$

$$\times \ [(1-(S_1/S_0)^2]^{-\gamma_n+1/2} (2\gamma_n+1)^{-1/2} [(1-[S_1/S_0]^2)^{1/2} + O(n^{-1})]$$

by Lemma 2.3

$$= (S_0/2)^{-\gamma_n} \pi_1 2^{-1} \ [2\pi/(2\gamma_n+1)]^{1/2} (1-\hat{\varphi}_1^2)^{-\gamma_n+1} [1+O(n^{-1})] \ ,$$

since S_1/S_0 is the Yule-Walker estimate for φ_1. Therefore, the theorem is true for $j = 1$.

Assume the theorem is true for any arbitrary j, and then consider $\tilde{I}_{j+1}$:

$$\tilde{I}_{j+1} = \int_{-1}^{1} \cdots \int_{-1}^{1} \pi_{j+1} 2^{-(j+1)} \left[\prod_{i=0}^{j+1} (1-\varphi_1^2)^{i/2} \right]$$

$$\times \ [2^{-1} \tilde{K}_{j+1}(\underline{\varphi}_{j+1})]^{-\gamma_n} d\varphi_{j+1} d\varphi_j \cdots d\varphi_1$$

$$= \pi_{j+1} 2^{-(j+1)+\gamma_n} \int_{-1}^{1} \cdots \int_{-1}^{1} \left[\prod_{i=0}^{j} (1-\varphi_i^2)^{i/2} \right] [\tilde{K}_j(\underline{\varphi}_j)]^{-\gamma_n} \quad \times$$

$$\times\left[\int_{-1}^{1}(1-\varphi_{j+1}^2)^{(j+1)/2}[1+\varphi_{j+1}^2-2\tilde{\varphi}_{j+1}(\underline{\varphi}_j)]^{-\gamma_n}d\varphi_{j+1}\right]\ d\varphi_j\ldots d\varphi_1\ ,$$

where

$$\tilde{\varphi}_{j+1}(\underline{\varphi}_j)\ =\ L_j(\underline{\varphi}_j)/\tilde{K}_j(\underline{\varphi}_j)\ .$$

Then

$$\tilde{I}_{j+1}\ =\pi_{j+1}2^{-(j+1)+\gamma_n}[2\pi/(2\gamma_n+1)]^{1/2}\int_{-1}^{1}\ldots\int_{-1}^{1}\left[\prod_{i=0}^{j}(1-\varphi_i^2)^{i/2}\right]$$

$$\times\ [\tilde{K}_j(\underline{\varphi}_j)]^{-\gamma_n}[1-\tilde{\varphi}_{j+1}^2(\underline{\varphi}_j)]^{-\gamma_n+(j+2)/2}\ [1+O(n^{-1})]d\varphi_j\ldots d\varphi_1\ ,$$

by Lemma 2.3,

$$=\pi_{j+1}2^{-(j+1)+\gamma_n}[2\pi/(2\gamma_n+1)]^{(j+1)/2}\left[\prod_{i=0}^{j}(1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}\right]$$

$$\times\ [1-\tilde{\varphi}_{j+1}^2(\hat{\underline{\varphi}}_j)]^{-\gamma_n+(j+2)/2}\ [1+O(n^{-1})]\ ,$$

by Lemma 2.3 and the inductive assumption,

$$=\pi_{j+1}2^{-(j+1)}(S_0/2)^{-\gamma_n}[2\pi/(2\gamma_n+1)]^{(j+1)/2}$$

$$\times\left[\prod_{i=0}^{j+1}(1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}\right][1+O(n^{-1})]\ ,$$

since $\tilde{\varphi}_{j+1}(\hat{\underline{\varphi}}_j) = L_j(\hat{\underline{\varphi}}_j)/\tilde{K}_j(\hat{\underline{\varphi}}_j) = \hat{\varphi}_{j+1}$ by Lemma 2.2. Therefore, the theorem is true for $j = 0, 1,$ and assuming it is true for arbitrary j, it is also true for $j+1$, and so true for all j, by induction. □

Now then, according to (2.3.2), the posterior probability of the order given the data is:

$$\pi_{p|\underline{X}}(j|\underline{x}_n) = \frac{I_j}{\sum_{k=0}^{M} I_k}, \qquad j = 0, 1, \ldots, M$$

$$\doteq \frac{\pi_j 2^{-j}[2\pi/(2\gamma_n+1)]^{j/2} \prod_{i=0}^{j} (1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}}{\sum_{k=0}^{M} \pi_k 2^{-k}[2\pi/(2\gamma_n+1)]^{k/2} \prod_{i=0}^{k} (1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}}, \qquad (2.3.5)$$

The Bayes estimator of a parameter θ depends on the choice of a loss function. Once the loss function is defined, the Bayes estimator is that which minimizes the Bayes risk with respect to that loss function. Consider then the following loss function:

$$L(a, p) = 1, \quad \text{if} \quad a \neq p, \quad \text{zero otherwise} \qquad (2.3.6)$$

It is well-known that, in general, the Bayes risk, $r(\pi, \delta)$ is defined by

$$r(\pi, \delta) = \int_{\chi} F(\delta(x), x) d\mu_x,$$

where $F(a, x) = \int_{\Theta} L(a, \theta) f(x; \theta) d\pi(\theta)$, $f(x; \theta)$ is the density of X given $\underline{\theta}$, μ_x is a σ-finite measure dominating the probability

measure of the random variable X, $\pi(\theta)$ is the prior probability distribution for $\underline{\theta}$, and χ and Θ are the domains of X and $\underline{\theta}$, respectively. Then $r(\pi, \delta)$ is minimized by minimizing $F(\delta(x), x)$. In the present situation, one has:

$$F(k, \underline{x}_n) = \sum_{j=0}^{M} L(k, j) f(\underline{x}_n ; j) \pi_j$$

$$= \sum_{j=0}^{M} L(k, j) [f(\underline{x}_n , j) / \pi_j] \pi_j$$

$$= \sum_{j=0}^{M} L(k, j) I_j$$

$$= \sum_{j \neq k} I_j , \qquad \text{for some } k \in \{0, 1, \ldots, M\}.$$

If k is chosen to be the integer between zero and M such that I_k is a maximum, the $F(k, \underline{x}_n)$ is minimized, and so such a k is the Bayes estimator of the order p. Since the posterior probability that the order of the time series is j is the ratio of I_j to the sum of the I_i's, this Bayes estimator is the order with the maximum posterior probability. The following theorem shows that as n increases, the cumulative posterior probability of the order p approaches a cumulative probability distribution which assigns all the

probability to the correct order p.

Theorem 2.5. Let M be the upper bound for the order p of a stationary autoregressive time series and $\pi_{p|\underline{X}}(j|\underline{x}_n)$ be the marginal posterior probability that the order of the time series is j, where

$$\pi_{p|\underline{X}}(j|\underline{x}_n) = \frac{\pi_j 2^{-j}[2\pi/(2\gamma_n+1)]^{j/2} \prod_{i=0}^{j} (1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}}{\sum_{k=0}^{M} \pi_k 2^{-k}[2\pi/(2\gamma_n+1)]^{k/2} \prod_{i=0}^{k} (1-\hat{\varphi}_i^2)^{-\gamma_n+(i+1)/2}},$$

for $j = 0, 1, 2, \ldots, M$, and where $\hat{\varphi}_0 \equiv 0$, $\hat{\varphi}_i$ is the Yule-Walker estimate of φ_i for $i = 1, 2, \ldots, M$, $\pi_k > 0$, $k = 0, 1, \ldots, M$, is the prior probability that the order is k, and $\gamma_n = (n+3)/2$. Then if J_n is the random variable with cumulative distribution function

$$\Pi_{p|\underline{X}}(j|\underline{x}_n) = \sum_{i=0}^{j} \pi_{p|\underline{X}}(i|\underline{x}_n),$$

the sequence $\{J_n\}$ converges in probability to the random variable J with cumulative distribution function:

$$\Pi_J(j) = \begin{cases} 0, & j < p \\ 1, & j \geq p. \end{cases}$$

That is, the probability that $J = p$ is 1.

Proof: Let $i < p$, and consider the following limits in distribution:

$$\lim_{n\to\infty} \frac{(n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{(n+4)^{-p/2} \prod_{\ell=0}^{p} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}$$

$$= \lim_{n\to\infty} (n+4)^{(p-i)/2} \prod_{\ell=i+1}^{p} (1-\hat{\varphi}_\ell^2)^{(n-\ell+2)/2}$$

$$= 0 ,$$

since the Yule-Walker estimates $\hat{\varphi}_\ell$ converge in distribution to φ_ℓ, (see Anderson [4]), and $0 < |\varphi_\ell| < 1$, for $\ell = 0, 1, \ldots, p$. Then letting $T_k = \pi_k(\pi/2)^{k/2}$,

$$0 \leq \lim_{n\to\infty} \pi_{p|\underline{X}}(i|\underline{x}_n)$$

$$= \lim_{n\to\infty} \frac{T_i(n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{\sum_{k=0}^{M} T_k(n+4)^{-k/2} \prod_{\ell=0}^{k} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}$$

$$\leq \lim_{n\to\infty} \frac{T_i(n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{T_p(n+4)^{-p/2} \prod_{\ell=0}^{p} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}$$

$$= 0,$$

since T_i/T_p is a finite constant and the limit involving n is zero, as shown above. Therefore,

$$\lim_{n\to\infty} \pi_{p|\underline{X}}(i|\underline{x}_n) = 0, \quad \text{for} \quad i < p.$$

Now let $i > p$ and consider the following limit:

$$\lim_{n\to\infty} \frac{(n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{(n+4)^{-p/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}$$

$$= \lim_{n\to\infty} (n+4)^{-(i-p)/2} \prod_{\ell=p+1}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}$$

$$= 0,$$

since the Yule-Walker estimate $\hat{\varphi}_\ell$ converges to φ_ℓ in distribution and $\varphi_\ell \equiv 0$ for $\ell = p+1, p+2, \ldots,$ (Ramsey [18]). Then

$$0 \leq \lim_{n\to\infty} \pi_{p|\underline{X}}(i|\underline{x}_n)$$

$$= \lim_{n\to\infty} \frac{T_i(n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{\sum_{k=0}^{M} T_k(n+4)^{-k/2} \prod_{\ell=0}^{k} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}} =$$

$$= \lim_{n \to \infty} \frac{T_i (n+4)^{-i/2} \prod_{\ell=0}^{i} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}{T_p (n+4)^{-p/2} \prod_{\ell=0}^{p} (1-\hat{\varphi}_\ell^2)^{-(n-\ell+2)/2}}$$

$$= 0 ,$$

since T_i / T_p is a finite constant and the limit involving n is zero, as shown above. Therefore,

$$\lim_{n \to \infty} \pi_{p|\underline{X}}(i|\underline{x}_n) = 0, \quad \text{for} \quad i > p.$$

Now, for any n,

$$\Pi_{p|\underline{X}}(M|\underline{x}_n) = \sum_{i=0}^{M} \pi_{p|\underline{X}}(i|\underline{x}_n) = 1 ,$$

and so

$$\lim_{n \to \infty} \sum_{i=0}^{M} \pi_{p|\underline{X}}(i|\underline{x}_n) = 1.$$

Then

$$\lim_{n \to \infty} \pi_{p|\underline{X}}(p|\underline{x}_n) = 1 - \lim_{n \to \infty} \left[\sum_{i=0}^{p-1} \pi_{p|\underline{X}}(i|\underline{x}_n) + \sum_{i=p+1}^{M} \pi_{p|\underline{X}}(i|\underline{x}_n) \right]$$

$$= 1 - \sum_{i \neq p} \lim_{n \to \infty} \pi_{p|\underline{X}}(i|\underline{x}_n)$$

$$= 1,$$

since the limits of the probabilities, $\pi_{p|\underline{X}}(i|\underline{x}_n)$, are all zero for $i \neq p$, as shown above.

Therefore, for $j < p$,

$$\lim_{n\to\infty} \Pi_{p|\underline{X}}(j|\underline{x}_n) = \sum_{i=0}^{j} \lim_{n\to\infty} \pi_{p|\underline{X}}(i|\underline{x}_n)$$

$$= 0 = \Pi_J(j),$$

and for $j \geq p$

$$\lim_{n\to\infty} \Pi_{p|\underline{X}}(j|\underline{x}_n) = \sum_{i=0}^{j} \lim_{n\to\infty} \pi_{p|\underline{X}}(i|\underline{x}_n)$$

$$= 1 = \Pi_J(j)$$

That is,

$$\lim_{n\to\infty} \Pi_{p|\underline{X}}(j|\underline{x}_n) \to \Pi_J(j) \quad \text{for} \quad j = 0, 1, \ldots, M.$$

Therefore, by definition, $\{J_n\}$ converges in distribution to J. In addition, since $P(J=p) = 1$, $\{J_n\}$ converges to J in probability. □

Therefore, for large n, the maximum posterior probability for the order of the autoregressive time series will correspond to the correct order p.

III. COMPARISON OF THE BAYESIAN APPROACH TO ORDER ESTIMATION WITH AKAIKE'S FPE CRITERION

1. Introduction

In this chapter, Akaike's FPE criterion for choosing the order of a stationary autoregressive time series is compared to the Bayes estimate obtained in Chapter II. First Wolfer's sunspot data will be analyzed by the two methods. Then the two approaches will be applied to some computer simulated stationary autoregressive time series.

2. Wolfer's Sunspot Data

Wolfer's sunspot data consists of the average number of sunspots observed in 176 consecutive years, beginning in 1749. The data can be found in Anderson's book [4] among other places. It is a well accepted fact that the square root of this sunspot data, corrected for the mean, is best fitted by an autoregressive model of order two. A graph of the transformed sunspot data is found in Figure 3.1.

Recall that both Akaike's method and the Bayesian method require the choice of a maximum order M. Let $M = 15$. Since the square roots of the observations are corrected for the mean, which had to be estimated, Akaike's FPE_j for the model of order j is:

$$FPE_j = (n+j+1)/(n-j-1)\cdot\hat{\sigma}_j^2, \quad j = 0, 1, \ldots, M \tag{3.2.1}$$

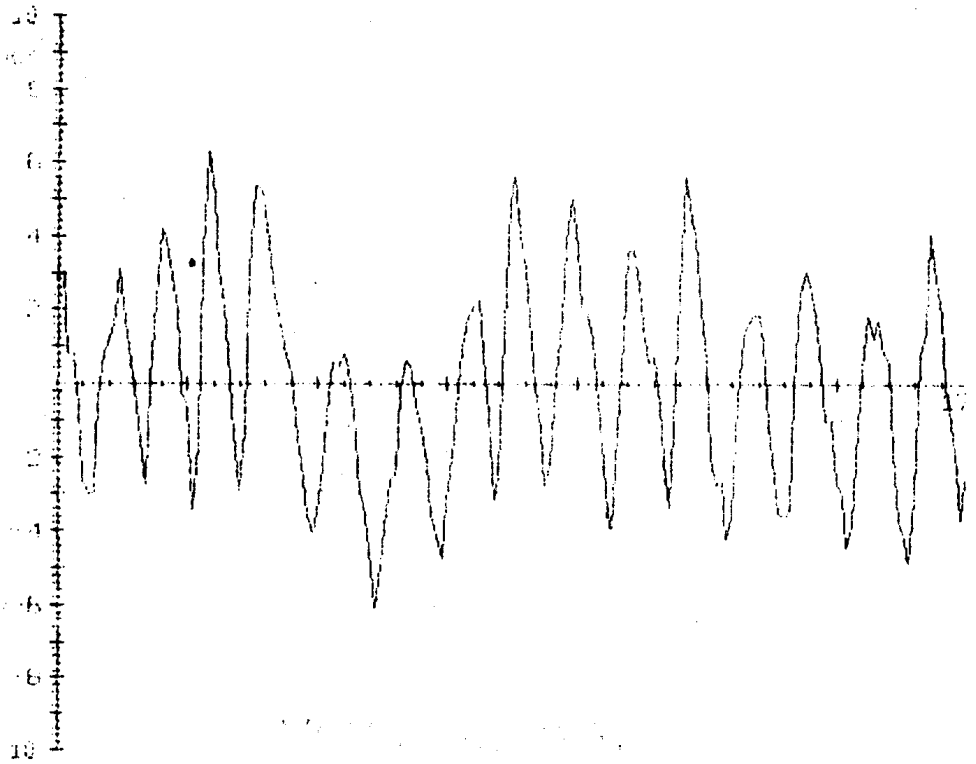

Figure 3.1. Wolfer's sunspot data--average number of sunspots from 1749 to 1924 (mean-corrected square roots).

where $n = 176$, the number of observations, the estimated prediction variance for order j is

$$\hat{\sigma}_j^2 = (S_0/n) \prod_{i=0}^{j} (1-\hat{\varphi}_i^2) ,$$

$\hat{\varphi}_0 \equiv 0$, $\hat{\varphi}_i$ is the Yule-Walker estimate for φ_i, $i = 1, 2, \ldots, M$, and

$$S_0 = \sum_{i=1}^{n} x_i^2 .$$

Akaike's estimate of the order of an autoregressive time series is the value k, where FPE_k is the minimum, $0 \leq k \leq M$.

For the Bayes estimate of the order, let $\pi_j = 1/16$, $j = 0, 1, \ldots, 15$, indicating uniform prior uncertainty about the order p. Then the posterior probability that the order is p is estimated by

$$\pi_{p|\underline{X}}(j|\underline{x}_n) \doteq \frac{T_j(n+4)^{-j/2} \prod_{i=0}^{j} (1-\hat{\varphi}_i^2)^{-(n-i+2)/2}}{\sum_{k=0}^{M} T_k(n+4)^{-k/2} \prod_{i=0}^{k} (1-\hat{\varphi}_i^2)^{-(n-i+2)/2}} , \qquad (3.2.2)$$

$$j = 0, 1, \ldots, \quad M = 15,$$

where $T_k = \pi_k(\pi/2)^{k/2}$, $\pi_k = 1/(M+1) = 1/16$ and $\hat{\varphi}_i$ is as defined above, $k = 0, 1, \ldots, M=15$. For the loss function defined in Section 2.3, the Bayes estimate is the integer k such that $\pi_{p|\underline{X}}(k|\underline{x}_n)$ is a maximum, $0 \leq k \leq M=15$.

The results of the analysis of the transformed sunspot data by the two methods is given in Table 3.1. Note that FPE_j is an absolute minimum for $j = 9$, however it dropped to 0.1961 for $j = 2$ and hovers between 0.19 and 0.20 for order 2 through 15. The posterior probability $\pi_{p|\underline{X}}(j|\underline{x}_n)$ is a distinct maximum for $j = 2$. Therefore, both methods support the widely accepted second order for the transformed sunspot data, although a strict interpretation of Akaike's method would lead to order nine.

3. Computer Simulations

The two methods are now applied to computer simulations of time series corresponding to the following three models:

Model I: $X_t = .3X_{t-1} + \varepsilon_t$,

Model II: $X_t = -.8X_{t-1} + \varepsilon_t$, and

Model III: $X_t = .5X_{t-1} - .5X_{t-2} + \varepsilon_t$,

with 100 time series of sizes $n = 50$ and $n = 100$, for each model. Note that both Models I and II have order $p = 1$ but the

Table 3.1. Wolfer's sunspot data--Akaike's FPE criterion and the Bayesian posterior probabilities for the order p.

	j: 0	1	2	3	4	5	6	7
FPE_j	1.0114	.3381	.1961	.1959	.1980	.1998	.1990	.1958
$\pi_{p\|\underline{X}}(j\|\underline{x}_n)$	.0000	.0000	.7677	.2070	.0204	.0024	.0008	.0008
	8	9	10	11	12	13	14	15
FPE_j	.1944	.1909	.1925	.1940	.1962	.1984	.1984	.2006
$\pi_{p\|\underline{X}}(j\|\underline{x}_n)$	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

coefficient of X_{t-1} is more extreme in Model II. Model III has order $p = 2$. The coefficients of X_{t-1} and X_{t-2} are both moderate. The results of the estimation of the orders of each time series by the two methods are summarized in Table 3.2.

As one might expect, the order of Model I is more often underestimated than that of Model II, since the absolute value of φ_1 is smaller in Model I than in Model II. For the smaller time series, $n = 50$, the Bayesian approach was slightly worse than Akaike's in estimating the order of Model I, in that it failed more often to identify the order as 1. Note, however, the Bayesian approach tends to underestimate in this model, whereas the Akaike method tends to overestimate the order in all the models. For the larger time series, $n = 100$, both estimation procedures are more accurate, but the Bayesian method appears superior.

Table 3.3 shows the maximum and minimum posterior probabilities that j is the order for these computer simulations. From that table, one can see the increase in number of observations from 50 to 100 has a substantial effect on the posterior probability distribution of the order. With minor exceptions, the maximum and minimum probabilities for the correct order both increase and the maximum and minimum probabilities both decrease for the incorrect orders.

Therefore, it can be seen in the analysis of both the sunspot data and the computer simulations that the Bayesian procedure

Table 3.2. Comparison of Akaike's FPE criterion and the Bayesian method.

Model	n		0	1	2	3	4	5	6	7	8	9	10
I	50	Akaike	25	55	9	5	2	1	3	0	0	0	0
		Bayes	52	44	4	0	0	0	0	0	0	0	0
	100	Akaike	5	66	15	6	3	2	0	3	0	0	0
		Bayes	28	67	5	0	0	0	0	0	0	0	0
II	50	Akaike	0	77	16	3	3	1	0	0	0	0	0
		Bayes	0	94	5	1	0	0	0	0	0	0	0
	100	Akaike	0	79	12	2	2	1	1	1	1	1	0
		Bayes	0	96	2	2	0	0	0	0	0	0	0
III	50	Akaike	0	2	81	7	6	1	1	1	0	0	1
		Bayes	2	2	93	2	1	0	0	0	0	0	0
	100	Akaike	0	0	67	12	9	7	3	0	1	0	1
		Bayes	0	0	94	6	0	0	0	0	0	0	0

Table 3.3. Minimum and maximum posterior probabilities for the order in computer simulations.

Model	n = 50 min	n = 50 max	n = 100 min	n = 100 max
I				
0	.0008	.8252	.0000	.8356
1	.0107	.8203	.1242	.8720
2	.0248	.7480	.0175	.7369
3	.0042	.2135	.0050	.1714
4	.0008	.0702	.0007	.0479
5	.0002	.1031	.0001	.1423
6	.0000	.0501	.0000	.0264
7	.0000	.0126	.0000	.0106
8	.0000	.0021	.0000	.0014
9	.0000	.0006	.0000	.0003
10	.0000	.0002	.0000	.0000
II				
0	.0000	.1004	.0000	.0000
1	.0411	.8275	.1551	.8757
2	.0830	.7812	.0211	.5954
3	.0241	.3353	.0132	.6816
4	.0043	.2959	.0017	.1236
5	.0007	.0594	.0002	.0278
6	.0001	.0160	.0000	.0082
7	.0000	.0065	.0000	.0018
8	.0000	.0012	.0000	.0005
9	.0000	.0003	.0000	.0002
10	.0000	.0001	.0000	.0000
III				
0	.0000	.5818	.0000	.0344
1	.0000	.5526	.0000	.2083
2	.0520	.8293	.0998	.8756
3	.0361	.4940	.0826	.7585
4	.0084	.5493	.0127	.2253
5	.0014	.1060	.0018	.3021
6	.0002	.0427	.0002	.0386
7	.0000	.0119	.0000	.0049
8	.0000	.0022	.0000	.0022
9	.0000	.0004	.0000	.0020
10	.0000	.0001	.0000	.0003

outlined in Chapter II compares favorably with Akaike's FPE procedure for large time series. For moderate to small time series ($n = 50$) with the autoregressive parameter close to zero, as in Model I, the Bayesian procedure tends to underestimate the order of the model. A comparison of the maximum and minimum posterior probabilities of the orders 0 through 10 for all three models when n is increased from 50 to 100 indicate that the posterior probabilities for incorrect orders are decreasing and the posterior probabilities for the correct orders are increasing as the asymptotic theory predicts.

BIBLIOGRAPHY

1. Akaike, H. (1969), "Fitting Autoregressive Models for Prediction", Annals of the Institute of Statistical Mathematics, 21, 243-247.

2. _____. (1970), "Statistical Predictor Identification", Annals of the Institute of Statistical Mathematics, 22, 203-217.

3. _____. (1974), "A New Look at the Statistical Model Identification", IEEE Transactions on Automatic Control, AC-19, 716-723.

4. Anderson, T.W. (1972), The Statistical Analysis of Time Series, John Wiley & Sons, New York.

5. Barndorff-Nielsen, O. and Schou, G. (1973), "On the Parameterization of Autoregressive Models by Partial Autocorrelations", Journal of Multivariate Analysis, 3, 408-419.

6. Bhansali, R.J. (1973), "A Monte Carlo Comparison of the Regression Method and the Spectral Method of Prediction", Journal of the American Statistical Association, 68, 621-625.

7. Box, G.E.P. and Jenkins, G.M. (1970), Time Series Analysis: Forecasting and Control, Holden Day, San Francisco.

8. Churchill, R.V. (1958), Operational Mathematics, McGraw-Hill, New York.

9. Cox, D.R. and Hinkley, D.V. (1974), Theoretical Statistics, Chapman and Hall, London.

10. Durbin, J. (1960), "Fitting of Time Series Models", Review of the Institute of International Statistics, 28, 233-244.

11. Halpern, E.F. (1973), "Polynomial Regression from a Bayesian Approach", Journal of the American Statistical Association, 68, 137-143.

12. Hannan, E.J. (1970), Multiple Time Series, John Wiley & Sons, New York.

13. Jones, R.H. (1975), "Fitting Autoregressions", Journal of the American Statistical Association, 70, 590-592.

14. McClave, J.T. (1975), "Subset Autoregression", Technometrics, 17, 213-220.

15. _____. (1978), "Estimating the Order of Autoregressive Models: The Max χ^2 Method", Journal of the American Statistical Association, 73, 122-128.

16. Parzen, E. (1974), "Some Recent Advances in Time Series Analysis", IEEE Transactions of Automatic Control, AD-19, 723-730.

17. Quenouille, M.H. (1947), "A Large-Sample Test for the Goodness of Fit of Autoregressive Schemes", Journal of the Royal Statistical Association, Ser. A, 110, 123-129.

18. Ramsey, F.L. (1974), "Characterization of the Partial Autocorrelation Function", Annals of Statistics, 2, 1296-1301.

19. Schwarz, G. (1978), "Estimating the Dimension of a Model", Annals of Statistics, 6, 461-464.

20. Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion", Biometrika, 63, 117-126.

21. Walker, A.M. (1952), "Some Properties of the Asymptotic Power Functions of Goodness-of-Fit Tests for Linear Autoregressive Schemes", Journal of the Royal Statistical Society, Ser. B, 14, 117-134.

22. Whittle, P. (1952), "Test of Fit of Time Series", Biometrika, 39, 309-318.

23. _____. (1963), Prediction and Regulation, D. Van Nostrand Co., Princeton, New Jersey.