Order Number 9334939

One channel at-a-time multichannel autoregressive modeling: Applications to stationary and nonstationary covariance time series

Stone, David, Ph.D.

University of Hawaii, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

ONE CHANNEL AT-A-TIME MULTICHANNEL AUTOREGRESSIVE MODELING: APPLICATIONS TO STATIONARY AND NONSTATIONARY COVARIANCE TIME SERIES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMMUNICATION AND INFORMATION SCIENCES

AUGUST 1993

by

David Stone

Dissertation Committee:

Will Gersch, Chairperson
Wes Peterson
Bill Remus
Tony Kuh
Marvin Ortel

ABSTRACT

The research explores and develops a new strategy for the multichannel (multivariate) autoregressive (MCAR) time series modeling of multichannel stationary and nonstationary time series. The multichannel time series modeling is achieved doing things one channel at-a-time using only scalar computations on instantaneous data. Under the one channel at-a-time modeling paradigm, three long standing and important problems in multichannel time series modeling are studied. First, one channel at-a-time scalar autoregressive (AR) time series modeling in combination with subset selection and a subsequent linear transformation achieves a relatively parsimonious multichannel autoregressive model of stationary time series and reduced one-step-ahead prediction variance as compared to conventional MCAR model fitting. Second, enhanced power spectral density estimation for multichannel stationary time series may be achieved with one channel at-a-time multichannel AR modeling in combination with a smoothness priors distribution on the scalar AR model parameters. Third, estimates of the time varying power spectral density matrix for multichannel nonstationary covariance time series are achieved using the one channel at-a-time paradigm in conjunction with a Bayesian smoothness priors stochastic linear regression model of the partial correlation coefficients (PARCORS) of a scalar lattice AR model. In this case, only a small number of hyper-parameters are fitted for the multichannel time varying AR model which has many more parameters than data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 THE GENERAL CONTEXT OF THE RESEARCH

The problems of multichannel autoregressive (MCAR) modeling of stationary and nonstationary time series data are addressed here doing things "one channel at-a-time" using only scalar computations on instantaneous data. The one channel at-a-time modeling is achieved via the known but insufficiently attended to idea of an instantaneous response multichannel autoregressive model with orthogonal innovations variance. Using that model, problems of fitting multichannel autoregressive models to data are reduced to problems of fitting scalar autoregressive models to data. The fitted scalar autoregressive models are combined into a multichannel instantaneous response-orthogonal innovations model. In turn the multichannel instantaneous response-orthogonal innovations model is expressible as conventional MCAR model via a linear algebraic transformation.

Three longstanding time series modeling problems are addressed using the one channel at-a-time paradigm. The first is to achieve relatively parsimonious modeling of both nonstationary mean and second order stationary time series as compared to conventional MCAR model fitting. This is achieved here by modeling the individual scalar autoregressive (AR) models in terms of a linear regression model and subsequently using subset selection. The objective in this problem, is to decrease the one-step-ahead prediction variance as compared to more conventional MCAR modeling methods. Secondly, as compared to more conventional MCAR modeling methods, the one channel at-a-time paradigm used in conjunction with the quasi-Bayesian "smoothness priors long AR" modeling of the individual channels achieves enhanced power spectral density estimation for multichannel stationary time series. Thirdly, multichannel time varying AR modeling of nonstationary covariance time series data is achieved using the one channel at-a-time AR models in conjunction with a smoothness priors time-varying stochastic linear regression modeling of the partial correlation coefficients in each AR channel. The d-channel time varying AR model, a model with many more parameters than data, is fitted with only $2\,d$ explicitly computed hyperparameters.

1.2 PROBLEM STATEMENT

This dissertation is addressed to ascertaining the potential usefulness of the doing multichannel things "one channel at-a-time" paradigm in the modeling of multichannel stationary and nonstationary covariance time series. The algorithms developed under this paradigm can be used for fitting multichannel autoregressive models for stationary time series forecasting and for understanding the dynamic interrelations between channels of stationary and nonstationary time series. A key idea in this thesis is to exploit a known but insufficiently attended to idea. That is, to fit multichannel instantaneous response orthogonal innovations autoregressive models to multichannel time series data. By modeling multichannel time series one channel at-a-time, the problems of modeling multichannel time series are reduced to problems in the modeling of scalar time series. Our effort in this dissertation is to develop new algorithms using the one channel at-a-time paradigm and to compare their statistical performance with the statistical performance of other, known multichannel modeling methods. This research also aims at stimulating further research and development of the one channel at-a-time paradigm.

1.3 OVERVIEW OF THE DISSERTATION

The one channel at-a-time paradigm to the analysis of multichannel time series analysis is presented in Chapter 2. Under this paradigm multichannel time series modeling is achieved one channel at-a-time. In that chapter the small number of seminal concepts and methodologies in time series that we identify as the intellectual basis of this thesis are reviewed. These works include Levinson (1947), Kalman (1960), Whittle (1963), Akaike (1973) and Akaike (1980). Levinson introduced an algorithm for fitting successively increasing order autoregressive models to scalar stationary time series via a recursive computational procedure. Prior to Levinson's algorithm, fitting scalar autoregressive models to data implied solving linear equations using matrix inversion techniques. Levinson's algorithm is achieved with only scalar computations. In Whittle (1963), Levinson's algorithm is extended to the fitting of multichannel autoregressive models to stationary time series data. Increasing order multichannel autoregressive models are fitted to data recursively requiring only the inversion of two matrices of constant dimension (number of channels). Whittle (1963), required the innovation of forward and backward in time multivariate autoregressive models. That innovation subsequently was

adapted to the Levinson algorithm to permit it to be realized in a modular "lattice" form in VLSI chips for practical applications. (In Chapter 5, we exploit the Levinson type lattice structure in our fitting of time varying autoregressive models to data.) The Levinson and Whittle algorithms were contributions to the computational issues in fitting autoregressive models to data. The long standing (statistical hypothesis testing) problem of determining which particular (scalar or multivariate) autoregressive model order was best in a statistical sense was solved in Akaike (1973). Akaike's AIC is an information theoretic based statistical criterion for determining the best of alternative parametric models fitted to data. (In a survey of prominent statisticians, Akaike's AIC was identified as one of the 20 statistical breakthroughs in the twentieth century, Kotz and Johnson 1992.) The value of the AIC statistic that is computed for a contender model with p estimated parameters is: $AIC(p) = -2\times log-likelihood\ of\ the\ fiited\ model + 2(p)$. The computation of the likelihood of a model is clearly a computational issue in statistical data analysis. As a consequence of Kalman (1960), we have seen that every variety of stationary and nonstationary time series model can be embedded into a state space model form. The Kalman filter is a computationally economical state space method for computing the likelihood of models. In general, in the modeling of nonstationary time series we are confronted with the attempt to fit models with substantially more parameters than data. Those large numbers of parameters can be expressed in terms of the evolution of simple stochastic difference equations such as random walks. That representation invokes the Bayesian concept of stochastic regression in which the unknown variances of the random walks have the role of "hyperparameters". Akaike(1980) developed a breakthrough quasi-Bayesian method of analysis that yields the modeling of nonstationary times series in terms of least squares computations. The paper developed the first practical application of the Bayesian model which used the likelihood of the hyperparameters as a measure of the goodness of fit of a model to data. In the review papers, Gersch and Kitagawa (1988) and Gersch (1992), Akaike's analysis method was embedded into a state space model framework and applied to a variety of time series modeling problems not considered by Akaike.

In chapter 3 the problem of parsimonious multichannel autoregressive modeling of nonstationary mean and second order stationary time series is addressed. Our efforts in parsimonious multichannel autoregressive modeling were motivated by the inherent tendency to overparameterize that is present in

the conventional multichannel autoregressive (MCAR) modeling. The application of the one channel at-a-time paradigm and subsequent subset selection of the individual scalar autoregressive (AR) modeled channels achieves a "thinning" of the overparametrized MCAR models. Akaike's AIC is employed in selecting the best of the alternative subset models. That procedure is suitable for the modeling of both nonstationary mean and second order stationary time series. Comparative studies of published parsimonious modeling attempts showed our modeling to be more parsimonious than previous attempts. In a series of comparative Monte Carlo simulation studies we show superior performance in mean square tracking error and one-step-ahead prediction variance using our MCAR modeling as compared to more conventional MCAR modeling methods.

Chapter 4 addresses the problem of power spectral density estimation for multichannel stationary time series. Our objective is to achieve statistically more efficient (smaller variance) power spectrum estimation than is customary with conventional MCAR-AIC modeling methods. We combined the one channel at-a-time methodology and the scalar case smoothness priors long AR model for spectral estimation (Kitagawa and Gersch, 1985a), to realize a smoothness priors multichannel autoregressive SPMCAR), modeling of stationary time series for spectral estimation. A particular class of frequency domain smoothness priors is assumed for the (normal) distribution of the coefficients of scalar autoregressive model of order m, (AR_m), with m relatively large. Monte Carlo simulation results demonstrate that the SPMCAR method does achieve superior statistical performance as compared with the power spectral density estimation obtained with the more conventional MCAR-AIC modeling method.

In Chapter 5 the problem of modeling multichannel nonstationary covariance time series data with a multichannel time varying autoregressive model is achieved with a one channel at-a-time smoothness priors, (SPMCTVAR), paradigm. The lattice structure of the one channel at-a-time autoregressive model and a smoothness priors random walk model in the regression equation for the partial correlation coefficients, (PARCORS), are exploited in this development. The fitting of a multichannel time varying autoregressive model to data requires the estimation of many more parameters than there are data. Specifically if there are d channels of N data points each, or a total of $N \times d$ observations and an order P multichannel time varying autoregressive model is fitted to that data, then $N \times P \times d \times d$ autoregressive parameters and $N \times d \times (d + 1)/2$ instantaneous innovations variance parameters describe

that model. The SPMCTVAR modeling algorithm requires that only $2 \times d$ hyperparameters to be explicitly estimated for each multichannel AR model order considered. The modeling of multichannel nonstationary covariance time series is relevant to many econometric, human electroencephalogram, oceanographic, meteorological and seismic data analysis situations. We show its application to both simulated data and real physical data bivariate and trivariate nonstationary covariance time series.

The last chapter summarizes the research results in this dissertation and points out areas where additional research may be fruitfully conducted.

CHAPTER 2

BACKGROUND

In this chapter the technical methods in time series analysis modeling that are relevant to the dissertation are reviewed. First we remark on some historical points.

2.1 HISTORICAL AND OTHER REMARKS

A time series is a collection of data arranged in chronological order. The study of time series goes back to antiquity. Undoubtably early cultures noted and studied astronomical and meteorological evidence. From the old testament, Ecclesiastes, 1:9: " . . . *the thing that hath been, is that which shall be, and that which is done, is that which shall be done, and there is no new thing under the sun* .... Some statisticians interpret that quotation to suggest knowledge in biblical times of a persistent structure principle that is particular relevant for forecasting. More recently, we have the Beveridge wheat price index, the annual index of prices at which wheat was sold in European markets from 1500 to 1869, (Beveridge 1921, Anderson 1971) and Waldmeir (1961) a record of sunspot activity in the years 1610 to 1960. The earliest western world papers in the analysis of time series that are still interesting today are very likely those by Shuster (1888) on the analysis of "hidden" periodicities in data and by Einstein (1905) on Brownian motion. The subsequent development has seen important contributions by other very illustrious individuals including Norbert Wiener, John Tukey and Kolmogorov.

Our concern here is with the modeling of multivariate time series. Despite the fact that time series analysis is now widely used in many branches of engineering, the physical sciences and economics there are only a small number of books on the modeling of multivariate time series. The totality of that literature includes, Quenouille (1957), Hannan (1970) and Lutkepohl (1991). Concerning the work presented here, there are only a small number of seminal works that we identify as the intellectual basis of this thesis. Those include Levinson (1947), (a recursive computational procedure for fitting increasingly complex scalar time series to data using only scalar computations), Kalman (1960), (effectively the introduction of the state-space methodology to time series analysis), Whittle (1963), (a recursive computational procedure for fitting increasingly complex multivariate time series to data), Akaike

(1973), (an information theoretic based statistic to select the best of alternative parametric models fitted to data) and Akaike (1980), (the development of a quasi-Bayesian analysis of the linear time series Gaussian disturbances model).

2.1.1 The Organization of This Chapter

In Section 2.2, the one channel at-a-time paradigm, the idea that is the key to our approach to the multivariate autoregressive modeling of stationary and nonstationary covariance time series, is introduced. Briefly that paradigm permits us to do multichannel things, (which in general require matrix computations), one channel at-a-time. These computations are done on instantaneous data using only scalar computations.

Following that, in Section 2.3 the more conventional methods for the fitting of both scalar and multivariate autoregressive models to time series data are treated. In Section 2.4 Akaike's AIC criterion, an information theoretic method for determining the best of alternative parametric models fitted to data is discussed. In that context, both the problems of autoregressive model order selection and subset selection are treated. Autoregressive model order selection has a role in each part of this thesis where autoregressive models are fitted to data. Subset selection, (in a linear regression scalar autoregressive model fitting), is relevant in achieving the parsimonious modeling of multivariate autoregressive modeling. This parsimonious modeling permits a decrease in one-step-ahead prediction variance in our one channel at-a-time instantaneous response modeling as compared to the more conventional methods of multichannel autoregressive modeling. Finally, in Section 2.5 the topic of smoothness priors is treated. The smoothness priors methodology, a linear Gaussian disturbances stochastic regression modeling, plays a critical role in conjunction with the one channel at-a-time paradigm, in achieving enhanced multichannel spectral density estimation as compared to conventional multichannel stationary time series spectral density estimation. It also plays a critical role in the fitting of multichannel time varying autoregressive models to multichannel nonstationary covariance time series.

2.2 THE ONE CHANNEL AT-A-TIME PARADIGM

The key idea in our approach to the multichannel autoregressive modeling of stationary and nonstationary time series is to do multichannel things one channel at-a-time. The one channel at-a-time autoregressive modeling is achieved via an orthogonal innovations instantaneous response multichannel autoregressive model. An instantaneous response-orthogonal innovations MCAR parameterization was introduced by Pagano (1978). It was used in the program MULMAR by Kitagawa (Akaike et al. 1978). The instantaneous response-orthogonal innovations MCAR model permits multichannel autoregressive modeling one autoregressive channel at-a-time with computations on instantaneous data. It is related to the instantaneous response orthogonal innovations model by an invertible linear transformation.

2.2.1 Instantaneous Response-Orthogonal Innovations and Conventional MCAR Models

From Theorem 1, Pagano (1978) an instantaneous response-orthogonal innovations MCAR model of the stationary d-vector random process $\{x(n)\}$, is in the form

$$\sum_{m=0}^{M} B_m x(n-m) = \varepsilon(n) \tag{2.2.1}$$

with ε a zero mean uncorrelated process with diagonal innovations variance matrix D

$$D = \begin{bmatrix} \sigma_{11} & . & . & . & . & . \\ . & \sigma_{22} & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & \sigma_{dd} \end{bmatrix}. \tag{2.2.2}$$

and B_0 a unit (ones on the diagonal) lower triangular matrix

$$B_0 = \begin{bmatrix} 1 & 0 & 0 & . & 0 & 0 \\ b_{210} & 1 & 0 & . & . & . \\ b_{310} & b_{320} & 1 & . & . & . \\ . & . & . & . & 0 & . \\ . & . & . & . & 1 & 0 \\ b_{d10} & b_{d20} & b_{d30} & . & b_{d(d-1)0} & 1 \end{bmatrix}. \tag{2.2.3}$$

The model in (2.2.1) can be re-expressed in the conventional MCAR form (Whittle 1963, Lutkepohl 1991)

$$x(n) + \sum_{m=1}^{M} A_m x(n-m) = \varepsilon(n) \, . \tag{2.2.4}$$

with general innovations matrix V and MCAR parameter matrices, $A_m = B_0^{-1}B_m$. Also, $\varepsilon(n) = B_0^{-1}\tilde{\varepsilon}(n)$. The relation between the innovations variance matrix V of the model in (2.2.4) and D the diagonal innovations matrix in (2.2.2) is

$$V = E[\varepsilon(n)\varepsilon(n)'] = B_0^{-1}DB_0^{-T}. \tag{2.2.5}$$

The key observation here is that the instantaneous response-orthogonal innovations MCAR model permits multichannel data to be modeled one AR channel at-a-time.

Conversely, consider the conventional MCAR model in (2.2.4). The unique modified Cholesky decomposition of V is, $V = LDL^T$, where L is unit lower triangular and D is as in (2.2.2). Then $L^{-1} = B_0^{-1}$ resulting in the instantaneous response model in (2.2.1) with orthogonal innovations. This reparameterization is useful when we consider multivariate AR models from the literature and examine their realizations as instantaneous response-orthogonal innovations equivalents in simulation studies.

A useful way of interpreting the relationship between the conventional MCAR representation in (2.2.4) and the instantaneous response model in (2.2.1) is that the latter can be thought of an interlacing of the data in the former.

For completeness here and later use, the power spectral density matrix of the conventional MCAR model (2.2.4)-(2.2.5) is given by

$$S(f) = A(f)^{-1}VA^*(f)^{-T} \tag{2.2.6}$$

where $A(f)$, the "whitening" filter of the AR process is

$$A(f) = \sum_{k=0}^{M} A_k \exp[-2\pi i f k],$$

with $A_0 = I$, $A^*(f)$ is the complex conjugate matrix of A(f) and A^T is the transpose of the matrix A.

The power spectral density matrix for the system expressed in the instantaneous response orthogonal innovations form is given by

$$S(f) = B(f)^{-1} D B^*(f)^{-T} \tag{2.2.7}$$

where

$$B(f) = \sum_{k=0}^{M} B_k \exp[-2\pi i f k].$$

For example, consider a bivariate or $d = 2$ AR_4 model of the Box-Jenkins Series J input-output physical plant data as in Newton (1988), p. 342. The conventional MCAR model form (2.2.4) with parameter matrix $A_0 = I$, AR coefficient matrices $A_1, \ldots, A_4$ and innovations matrix V (to within four decimal digits), is;

$$A_1 = \begin{bmatrix} -1.9259 & 0.0012 \\ -0.0505 & -1.2998 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1.2017 & -0.0042 \\ 0.0205 & 0.3277 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -0.1169 & 0.0087 \\ 0.7118 & 0.2570 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} -0.1042 & -0.0033 \\ -0.1954 & -0.1334 \end{bmatrix}, \quad V = \begin{bmatrix} 0.3516 & -0.0731 \\ -0.0731 & 0.9724 \end{bmatrix}.$$

Correspondingly the equivalent instantaneous response-orthogonal innovations matrix representation may be computed. By simple algebraic manipulation, from (2.2.3)-(2.2.5), the matrix B_0 is

$$B_0 = \begin{bmatrix} 1.0000 & 0.0000 \\ 0.2081 & 1.0000 \end{bmatrix}$$

where $b_{210} = -V(1,2)/V(1,1)$. Then computing $B_m = B_0 A_m$, $m = 1, \ldots, 4$ and $D = B_0 V B_0^T$, we arrive at the equivalent theoretical instantaneous response orthogonal innovations model. The B_i and D matrices are then

$$B_1 = \begin{bmatrix} -1.9259 & 0.0012 \\ -0.4512 & -1.2995 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1.2016 & -0.0042 \\ 0.2705 & 0.3268 \end{bmatrix}, \quad B_3 = \begin{bmatrix} -0.1169 & 0.0087 \\ 0.6875 & 0.2588 \end{bmatrix},$$

$$B_4 = \begin{bmatrix} -0.1042 & -0.0033 \\ -0.2171 & -0.1341 \end{bmatrix}, \quad D = \begin{bmatrix} 0.3516 & 0.0000 \\ 0.0000 & 0.9572 \end{bmatrix}.$$

2.3 CLASSICAL SCALAR AND MULTICHANNEL AUTOREGRESSIVE MODELING

The earliest paper on (scalar) autoregressive modeling of time series is that by Yule (1927). The method of analysis was least squares and the paper contained an application to the modeling of

Wolfer's sunspot series. Mann and Wald (1943) demonstrated that least squares formulas were asymptotically valid for large data sample maximum likelihood estimates of the autoregressive model parameters. Following Mann and Wald, an asymptotic goodness of fit test of an autoregressive model of order p could be made within the class of autoregressive series of some higher order q using the asymptotic properties of the fitted coefficients, which include asymptotic normality. The classical likelihood theory applies. Then the sum of squares of residuals asymptotically have a chi-square distribution with $q - p$ degrees of freedom. Additional contributions to autoregressive model order determination in the context of goodness of fit test were made by Quenouille (1947) and Walker (1954).

The computational solution of fitting successively higher order autoregressive models to data required the inversion of successively higher order matrices. The inversion of a $p \times p$ matrix was required for the fitting of a pth order autoregression to data. The Levinson (1947) algorithm was a computational breakthrough. It permitted increasing order autoregressive models to be fitted using only scalar computations.

Multivariate autoregressive models were probably first introduced in the econometric literature with the monograph *Statistical Inference in Dynamic Models* edited by Koopmans (1950). They are also discussed in Whittle (1953) and Bartlett (1955). The fitting of d simultaneous time series to a d variate autoregressive time series of order p required the inversion of $p \times d \times d$ linear "Yule-Walker" equations. For $d = 10$ channels and an autoregressive model order of $p = 10$, the inversion of a 1000×1000 matrix is required. That large computational burden inhibited the application of multivariate autoregressive modeling. The Whittle (1963) algorithm was a computational breakthrough. It permitted increasing order multivariate autoregressive models to be fitted requiring only successive inversion of $d \times d$ matrices. Hannan (1970) contains a formal asymptotic distribution theory for the parameters of multivariate autoregressive models fitted to data as well as some (unwieldy) goodness of fit tests for multivariate autoregressive model order determination.

In practice, the classical goodness of fit tests for determining the orders of scalar and multivariate autoregressive models fitted to data appeared to be difficult to use. Akaike (1969) was a breakthrough paper for determining the order of scalar autoregressive models fitted to data. The "final prediction

error" concept in that paper for scalar autoregressive model order determination was extended to multivariate autoregressive model order determination in Akaike (1972). Akaike (1973) was a consolidation of Akaike's previous work on model order determination. It yielded an information theoretic statistic, now known as Akaike's AIC, for the determination in a multiple hypothesis testing situation of the best of alternative parametric models. That paper was identified in a survey of outstanding statisticians as a breakthrough paper in the last half century of work in statistics, (Kotz and Johnson 1992).

Each of the topics identified in the preceding discussion is treated in considerable detail in the remaining material in this section.

2.3.1 Scalar Autoregressive Modeling

A scalar autoregressive model of order p of a stationary stochastic process, $\{x(n)\}$, is defined by

$$\sum_{i=0}^{p} \alpha_i x(n-i) = \varepsilon(n) . \tag{2.3.1}$$

In (2.3.1), $\varepsilon(n) \sim N(0,\sigma^2)$, i.i.d., $\alpha_0 = 1.0$ and $\alpha_p \neq 0$. σ^2 is referred to as the innovations variance. Let the mean and covariance functions of $\{x(n)\}$ be $E[x(n)] = \mu$ and $E[(x(n)-\mu)(x(n-k)-\mu)] = \gamma(k)$, where $\gamma(k)$, the covariance function, is an even function of k, and $\gamma(k) = \gamma(-k)$. For convenience assume that the mean μ is zero. Then postmultiplying (2.3.1) by $x(n-k)$ and taking the expectation yields the set of $p+1$ "Yule-Walker" equations.

$$\sum_{i=0}^{p} \alpha_i \gamma(k-i) = \sigma^2 \delta_{k,0}, \quad k=0,1,2,\,,\,,\,p` . \tag{2.3.2}$$

In (2.3.2), $\delta_{k,0}$ is the Kronicker delta and $\gamma(k) = E[x(n)][x(n-k)]$ is the covariance of the $\{x(n)\}$ process evaluated at lag k. The AR coefficients α_i's for $i=1,\,,\,p$ satisfy the last p simultaneous equations of (2.3.2). Let Γ_x be the $p \times p$ matrix of covariances

$$\Gamma_x = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix} .$$

Then from (2.3.2) and the definition of Γ_x, the solution vector for the AR model coefficients satisfy

$$[\,\alpha_1\,,\ldots,\alpha_p\,]\,\Gamma_x = -[\,\gamma(1),\ldots,\gamma(p)\,]\,. \tag{2.3.3}$$

In (2.3.3), we have used the fact that the covariance is an even function. The α_i's are obtained by postmultiplying (2.3.3) by the inverse of the $p \times p$ matrix Γ_x.

To fit the scalar autoregressive model (2.3.1) to a set of observed data $\{x(1),\ldots,x(N)\}$, the covariance function $\{\gamma(k)\}$ can be estimated by the sample covariance function, $\{c(k)\}$. The sample covariances are computed from the observed data, by the formula,

$$c(k) = \frac{1}{N}\sum_{n=k+1}^{N}(x(n)-\bar{x})(x(n-k)-\bar{x})\ ,\ \ \bar{x} = \frac{1}{N}\sum_{n=1}^{N}x(n)\ . \tag{2.3.4}$$

We replace the occurrences of $\gamma(k)$ with $c(k)$ in (2.3.2) and (2.3.3). This yields a_i, the estimates of the AR coefficients α_i. The estimate of the innovations variance is given by $\hat{\sigma}^2 = \sum_{i=0}^{p} a_i\, c(-i) = \sum_{i=0}^{p} a_i\, c(i)$, where $a_0 = 1$.

2.3.1.1 The Levinson-Durbin Algorithm for Scalar AR Model Fitting

The Levinson-Durbin algorithm (Levinson 1947) is a recursive algorithm which permits increasing order scalar models to be fitted to data using only scalar computations. The Levinson 1947 article was reprinted in an appendix to Norbert Wiener's book, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," (1949). The Levinson algorithm was the only practical way to realize Wiener's theoretical results. Durbin (1960) is an apparently independent but technically equivalent development of Levinson's work. Whittle (1963), extended the Levinson-Durbin algorithm to the fitting of multivariate autoregressive time series. In so doing, it was necessary to use forward time and backward time multivariate time series models. The forward-backward time autoregressive model concept was later embedded into the Levinson-Durbin algorithm for scalar autoregressive modeling. That development resulted in what is known as a "lattice structure", a modular form which facilitates physical implementation of autoregressive modeling on semiconductor chips.

First we briefly describe the Levinson-Durbin algorithm. Following that a brief treatment of the lattice-structure is shown. The lattice structure interpretation of scalar autoregressive modeling is crucial for our own development of the smoothness priors multichannel time varying autoregressive modeling (SPMCTVAR) which is discussed in Chapter 5.

Concerning the Levinson-Durbin algorithm: Let a_{pk} be the kth coefficient in a scalar AR order p model fitted to the data $x(1), \ldots x(N)$, and let v_p be the empirical residual variance of that model. The algorithm yields the values of the coefficients of an increased order AR model, $a_{p+1,k}$ in terms of the a_{pk}. The updated order coefficient $a_{p+1,p+1}$ is computed by $a_{p+1,p+1} = -\Delta_p / v_p$, where

$$v_p = \sum_{k=0}^{p} a_{pk} c_{-k} , \qquad \Delta_p = \sum_{k=0}^{p} a_{pk} c_{p-k+1} . \tag{2.3.5}$$

and the remaining coefficients of the order updated AR model are then given by

$$a_{p+1,k} = a_{pk} + a_{p+1,p+1} a_{p,p-k+1}, \quad k = 1, 2, \ldots, p . \tag{2.3.6}$$

In (2.3.5), the $\{c_k , k = 0,1, \ldots\}$ are the empirical covariances computed from (2.3.4).

Concerning the lattice form development: The forward and backward AR models of order M are respectively

$$x(n) = \sum_{i=1}^{M} \alpha_i x(n-i) + \varepsilon(n) , \qquad x(n-M) = \sum_{i=1}^{M} \alpha_i x(n-i+1) + \varepsilon(n-M) . \tag{2.3.7}$$

Respectively these models expresses an observation at time n as a linear combination of its past values, and as a linear combination of its future values.

A set of normal equations for each model can be derived from (2.3.7)

$$\sum_{m=0}^{M} \alpha_M(m) c(m-k) = \begin{cases} P_{f,M}, & k=0 \\ 0, & 1 \le k \le M \end{cases} , \qquad \sum_{m=0}^{M} \alpha_M(M-m) c(m-k) = \begin{cases} 0, & 0 \le k \le M-1 \\ P_{b,M}, & k=M \end{cases} \tag{2.3.8}$$

In (2.3.8), $P_{f,M} = \sum_{m=0}^{M} \alpha_M(m) c(m)$ is the forward prediction error variance dictated by the Mth order AR model (2.3.7) and $P_{b,M} = \sum_{m=0}^{M} \alpha_M(M-m) c(m)$ is the prediction error variance of the Mth order backward AR model.

As in the original development of the Levinson algorithm,

$$\Delta_M = \sum_{m=0}^{M} \alpha_M(m) c(m-M-1) . \tag{2.3.9}$$

After some algebraic manipulations (Haykin 1989), we obtain the Levinson-Durbin recursion for the updated order coefficient.

$$\alpha_{M+1}(m) = \alpha_M(m) + \rho_{M+1} \alpha_M(M-m+1) , \; m = 0,1, \cdots , M+1 \tag{2.3.10}$$

In (2.3.10), ρ_{M+1} is the "reflection coefficient" as it is commonly referred to the engineering literature.

The negative of ρ_{M+1} is known as the partial correlation (PARCOR) coefficient in the statistical literature. It can be easily seen that $\rho_{M+1} = \alpha_{M+1}(M+1)$ when m is substituted by $M+1$ in equation (2.3.10). To start the Levinson-Durbin recursion, the initial conditions are derived by setting $p=0$. So, $\alpha_0(1)=0$, and $P_0=c(0)$, where $c(0)$ is the sample variance of the time series.

2.3.2 Multivariate Autoregressive Modeling

Fitting a d-channel, order M AR model requires the successive inversion of a $(d^2\times M)\times(d^2\times M)$ matrix. This computational burden inhibited the widespread application of the multivariate autoregressive models to the fitting of multichannel time series data. The seminal paper, Whittle (1963), permitted multichannel autoregressive models to be fitted to time series data requiring only the inversion of two $d\times d$ matrices. The method is a generalization of the Levinson-Durbin algorithm to the multivariate case. Among the earliest papers to exploit the Whittle algorithm in fitting MCAR models to data were Jones (1964) in an application to meteorological prediction, Gersch (1970) in the modeling of electroencephalogram data and Akaike (1972) in the modeling of a rotary cement kiln for automatic control purposes.

Let $x(n)$ be a d-dimensional stationary stochastic process. Consider the multichannel autoregressive (MCAR) model of order M

$$\sum_{m=0}^{M} A_m x(n-m) = \varepsilon(n) \tag{2.3.11}$$

with $A_0 = I_d$ the d-dimensional identity matrix and $\varepsilon(n) \sim N(0,V)$, i.i.d.. Define the matrix covariance function $\Gamma(k)$ of the stationary process $x(n)$

$$\Gamma(k) = E[(x(n)-\mu_x)(x(n-k)-\mu_x)'] = \Gamma(k) = \Gamma'(-k),\ k=0,\pm1,\pm2,\ ,\ ,\ ,$$

where $E[x(n)] = \mu_x$. The apostrophe denotes the matrix and vector transpose operation. Postmultiplying equation (2.3.11) by $(x(n-k)-\mu_x)'$ and taking expectations yields the set of Yule-Walker equations,

$$\sum_{m=0}^{M} A_m \Gamma(k-m) = V\delta_{k,0},\ \ k=0,1,2,,,M. \tag{2.3.12}$$

The AR matrix coefficients A_m's for $m=1,,,M$ satisfy the last M simultaneous equations of (2.3.12).

That is,

$$\left[A_1,\ldots,A_M\right]\begin{bmatrix} \Gamma(0) & \Gamma(1) & \ldots & \Gamma(M-1) \\ \Gamma'(1) & \Gamma(0) & \ldots & \Gamma(M-2) \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \Gamma'(M-1) & \Gamma'(M-2) & \ldots & \Gamma(0) \end{bmatrix} = -\left[\Gamma(1),\ldots,\Gamma(M)\right]. \qquad (2.3.13)$$

The solution requires the inversion of a $(d^2\times M)$ *by* $(d^2\times M)$ matrix.

In fitting the MCAR model (2.3.11) to the d-dimensional observed data $x(n)$,n=1,2,...,N, the Yule-Walker estimators of the AR matrix coefficients A_m's are obtained by replacing $\Gamma(k)$'s in (2.3.13) with their estimates $C_k=\frac{1}{N}\sum_{(n=k+1)}^{N}[x(n)-\bar{x}][x(n-k)-\bar{x}]'$, where $\bar{x}=\frac{1}{N}\sum_{n=1}^{N}x(n)$ is the sample mean vector. That is,

$$\left[\hat{A}_1,\ldots,\hat{A}_M\right]\begin{bmatrix} C_0 & C_1 & \ldots & C_{M-1} \\ C'_1 & C_0 & \ldots & C_{M-2} \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ C'_{M-1} & C'_{M-2} & \ldots & C_0 \end{bmatrix} = -\left[C_1,\ldots,C_M\right] \qquad (2.3.14)$$

Also, the estimate of the innovations matrix V is given by $\hat{V}=\sum_{m=1}^{M}A_m C'_m$.

For large values of d and M, inversion of the $(d^2\times M)$ *by* $(d^2\times M)$ matrix in (2.3.14) can be a prohibitive numerical computation. The Whittle (1963) algorithm requires only the successive inversion of two $d\times d$ matrices to achieve increasing order MCAR modeling.

In Whittle's recursive algorithm both the forward and backward multivariate autoregressive models are required. The forward AR model expresses an observation at time t of a stationary process, $x(t)$, as a linear combination of its past values. The backward AR model expresses the same observation as a linear combination of it future values. Let $A_{p,m}$ and $\bar{A}_{p,m}$ respectively be the $d\times d$ *AR* coefficient matrices in the forward and backward MCAR models.

$$X(t)=\sum_{m=1}^{p}A_{p,m}X(t-m)+\varepsilon(t), \qquad X(t-p)=\sum_{m=1}^{p}\bar{A}_{p,m}X(t-p+m)+\varepsilon(t-p). \qquad (2.3.15)$$

The algorithm requires the simultaneous solutions for $A_{p,m}$'s and $\bar{A}_{p,m}$'s from the following system of

equations,

$$\sum_{m=0}^{p} A_{p,m}\Gamma(k-m) = 0, \quad k=1,2,...,p \qquad \sum_{m=0}^{p} \bar{A}_{p,m}\Gamma(m-k) = 0, \quad k=1,2,...,p. \tag{2.3.16}$$

The equations (2.3.16) are derived from the Yule-Walker equations for the forward and backward MCAR models. Given the $\Gamma(k)$ matrices, the AR matrix coefficients for successively increasing AR model orders can be recursively obtained from the following system of equations (Whittle, 1963),

$$A_{p+1,p+1} = -\Delta_p \ \bar{V}_p^{-1}, \qquad \bar{A}_{p+1,p+1} = -\bar{\Delta}_p \ V_p^{-1} \tag{2.3.17}$$

$$A_{p+1,k} = A_{p,k} + A_{p+1,p+1}\bar{A}_{p,p-k+1}, \qquad \bar{A}_{p+1,k} = \bar{A}_{p,k} + \bar{A}_{p+1,p+1}A_{p,p-k+1},$$

where $k=1,...,p$, and

$$V_p = \sum_{m=0}^{p} A_{p,m}\Gamma(-m), \qquad \Delta_p = \sum_{m=0}^{p} A_{p,m}\Gamma(p-m+1),$$

$$\bar{V}_p = \sum_{m=0}^{p} \bar{A}_{p,m}\Gamma(m), \qquad \bar{\Delta}_p = \sum_{m=0}^{p} \bar{A}_{p,m}\Gamma(m-p-1).$$

When fitting AR models using Whittle's algorithm to a set of observed data, the $\Gamma(k)$'s are replaced by their estimates $C(k)$'s.

A multichannel autoregressive model of a given order can be fitted to the observed data with the direct Yule-Walker solution method. We may also fit successive increasing orders of the multichannel autoregressive model to data with Whittle algorithm. In practice, the true model order of the system generated the data is unknown and must be estimated. The question of what order model to use is not addressed by either of those procedures. The topic of autoregressive model order selection is described in Section 2.4.

2.3.2.1 Prediction Using The Multivariate Autoregressive Model

One of the main objectives of time series analysis is prediction. Future values of a time series may be predicted based on its own past observed values. The most widely used criterion in modeling is the minimum mean squared errors (MSE) prediction. In this section, the linear MSE prediction using the multichannel autoregressive model (MCAR) is reviewed. We are only interested in results for one-step-ahead predictions. Two results are needed; a prediction formula and the prediction error variance.

Consider the zero mean d-dimensional time series $\{x(n)\}$. Here, the multichannel autoregressive model of order P is given by $x(n) = \sum_{m=1}^{P} A_m x(n-m) + \varepsilon(n)$

Let $\hat{x}(n+1|n)$ denote the optimal linear mean square error one-step ahead predictor for the given model,

$$\hat{x}(n+1|n) = \sum_{m=1}^{P} A_{m,p}\, x(n-m+1)\ , \tag{2.3.18}$$

where $A_{m,p}$'s, $m=1\ ,\ \cdots\ ,\ P$, are the predictor AR coefficient matrices.

Let $V(n+1|n) = E[x(n+1)-\hat{x}(n+1|n)][x(n+1)-\hat{x}(n+1|n)]'$ denote the one-step ahead prediction error variance matrix. Then, using the z-transform notation, $A(z)x(t) = \varepsilon(t)$. Since the stochastic process is stationary, the inverse $A^{-1}(z)$ exists and $x(t)$ has an equivalent infinite order moving average representation. So, $x(t) = A^{-1}(z)\varepsilon(t) = h(z)\varepsilon(t) = \sum_{i=0}^{\infty} h_i\, \varepsilon(t-i)$. Then the moving average representation of $x(t+1)$ is,

$$x(t+1) = \sum_{i=0}^{\infty} h_i\, \varepsilon(t+1-i) = h_0\varepsilon(t+1) + \sum_{i=0}^{\infty} h_{i+1}\varepsilon(t-i) \tag{2.3.19}$$

Since $\varepsilon(t\ +\ 1)$ is predicted to be identically zero at time t, from (2.3.19), the one-step-ahead prediction of $x(t\ +\ 1)$ given the data up to time t is

$$\hat{x}(t+1|t) = \sum_{i=0}^{\infty} h_{i+1}\varepsilon(t-i). \tag{2.3.20}$$

The one-step ahead prediction error variance matrix is then

$$V(n+1|n) = E[x(n+1)-\hat{x}(n+1|n)][x(n+1)-\hat{x}(n+1|n)]' = h_0 V h_0' = V\ , \tag{2.3.21}$$

In (2.3.21), h_0 is the identity matrix and V is the innovations variance matrix. The optimal linear mean square error one-step ahead prediction variance is therefore identically equal to the innovations matrix of the stationary process.

Assume that we have N observations from the stationary stochastic process $\{x(n)\}$. Let the optimal prediction variance of an h-step ahead predictor be $V(n+h|n)$. Then $\hat{V}(n+h|n) = (1+\frac{Pd}{N})V(n+h|n)$, where $\hat{V}(n+h|n)$ is the prediction variance of an h-step predictor using a MCAR model of order P (Lewis and Reinsel 1985, Lutkepohl 1991). The important conse-

quence of this result is that the prediction variance of the h-step ahead predictor increases as the order of the MCAR model increases. Excessively large model orders result in unreliable predictions.

2.3.2.2 MCAR Power Spectrum Density Estimation

Power spectrum estimation has evolved through the turn of the century. The first estimator to be used extensively was the periodogram which, although it was known to be a poor estimator because the variance of the estimate exceeds its mean value, is still used today. Blackman and Tukey (1958) was an autocorrelative method for power spectrum estimation. The steps of that procedure include estimating the autocorrelation function from the observed data, windowing the autocorrelation function in an "appropriate manner" and then Fourier transforming the windowed autocorrelation function to finally obtain the estimated power spectrum. Jenkins and Watts (1968), in a highly regarded book on spectral analysis, refer to the empirically based windowing method as "window carpentry". In the last paragraph of the last chapter of that book the authors caution that the models fitted by the aforementioned method should be regarded as tentative and advocate parametric model methods for superior modeling. Power spectrum estimation of both scalar and multivariate stationary time series can be statistically more reliably estimated through the fitting of autoregressive models.

Akaike (1970) in his breakthrough paper used autoregressive models for scalar power spectrum estimation. The relationship between the parameters of an AR model of order p and its estimated power spectrum density was well known at that time, (see Jenkins and Watts 1968). Akaike's contribution was the development of a statistical criterion to determine the order of the AR model to be fitted to the data. The generalization of that statistical criterion to multivariate processes followed soon after in Akaike (1971). That development was followed by Akaike (1973), in an information theoretic criterion to select the best of alternative parametric models fitted to data, the AIC, (see Section 2.4).

In what follows, we treat only the estimation of the matrix power spectral density using the multivariate autoregressive model. Given N observations of a d-dimensional stationary time series $\{X(t), t=1, \cdots, N\}$ we fit an MCAR model to that data using the method described in Section 2.3.2. The AIC statistic used to determine the order of that AR model is treated in Section 2.4.

As before let the MCAR model fitted to the data be $\sum_{i=0}^{P} A_i x(n-i) = \varepsilon(n)$, where $\varepsilon(n) \sim N(0,V)$, i.i.d., and V is the $d \times d$ innovations matrix.

Let $S(f)$ be the $d \times d$ power spectrum density matrix function at frequency f. An estimate of $S(f)$ is computed from the fitted MCAR model given by

$$\hat{S}(f) = A^{-1}(f) V_p A^{*}(f)^{-T} , \tag{2.3.22}$$

where $A(f) = \sum_{k=0}^{p} A_k \exp(-2\pi i f k)$ is the "whitening filter" of the AR process. Also in (2.3.22), $A^{*}(f)$ is the complex conjugate of $A(f)$ and the domain over which the power spectral density is computed is normalized so that $-1/2 \leq f \leq 1/2$.

2.4 THE AKAIKE AIC: SELECTING THE BEST OF ALTERNATIVE PARAMETRIC MODELS

Akaike's AIC (Akaike 1973, 1974) is an information theoretic criterion for the selection of the best of alternative parametric models based on observed data. The AIC is extensively applicable in statistics. It is an approximately unbiased estimate of the expected Kullback-Leibler information. The Kullback-Leibler information or the K-L number is an information theoretic measure of the dissimilarity between two distributions (Kullback and Leibler 1951; Kullback 1959). The larger the measure the greater the dissimilarity between the two distributions.

Based on a survey of leading statisticians, a tutorial on the AIC was selected for publication in "*Breakthroughs in Statistics*" Volume 1, (Kotz and Johnson 1992), as one of the nineteen breakthroughs of the twentieth century in the foundations and basic theory of statistics. Some of the recent applications of the AIC in the selection of models in addition to time series analysis are, factor analysis, regression analysis, multinomial logistic regression, cluster analysis, discriminant analysis and covariance structure models. Respectively these treatments appear in Akaike (1987a, 1987b), Terasvirta (1987), Fornell and Rust (1989), Shibata (1982), Takane et al (1987), Bozdogan and Sclove (1984), Sambamoorthi (1989) and Joreskog (1992).

The AIC is relevant in this thesis in scalar and multivariate AR model order selection and in subset selection in Chapters 3, 4 and 5. With regard to scalar time series AR model order selection, Akaike's first paper on that subject was Akaike (1969). The significance of that contribution can be appreciated when we recognize that problem had been under consideration for a long time and that distinguished scientists, including John Von Neuman, (Neumann (1941), had worked on that problem. Akiake (1971) is an extension of AR model order determination to multivariate AR models. The first formal paper on the AIC, (that paper subsumed the contributions of Akaike 1969, and 1971), was given by Akaike at an international information theory meeting in Taskent, USSR in 1971. The publication of that paper did not appear until 1973, (Akaike 1973). Therefore it may be of interest to note that the first actual publication on the AIC appeared in the Proceedings of the 5th HICSS, (Hawaii International Symposium on System Sciences), Akaike (1972).

Being such a significant contribution, the AIC has provoked the intellectual curiosity and competitiveness of many statisticians. Akaike himself did a Bayesian interpretation of the AIC for scalar AR model order determination, (Akaike 1978). Other contributions related to AR model order selection and subset selection in regression among many others include, Gersch and Sharp (1973), Jones (1976), Parzen (1977), Bhansali and Downham (1977), Schwartz (1978), Sugira (1978), Shibata (1980, 1981) ,Findley (1985), Lutkepohl (1985), and Hurvich and Tsay (1989, 1992a, 1992b).

In Section 2.4.1, we exhibit some of the theory of the AIC. Section 2.4.2 treats the application of the AIC to scalar and multivariate AR model order selection. That material is relevant in Chapters 4 and 5. Section 2.4.3 treats the application of the AIC to subset selection. That material is relevant to the parsimonious MCAR modeling and one-step-ahead MCAR prediction in Chapter 3.

2.4.1 The Kullback Leibler Measure and the Akaike AIC

Consider a probability density function $g(x)$ of a continuous distribution function g, and a probability density function f(x) of f, an approximation to g. The Kullback-Leibler information for g and f

is defined as the following

$$I(g;f) = E_X(\log(\frac{g(X)}{f(X)})) \quad = \int_{-\infty}^{\infty} g(x)log(\frac{g(x)}{f(x)})\,dx \tag{2.4.1}$$

$$= \int_{-\infty}^{\infty} g(x)log(g(x))dx - \int_{-\infty}^{\infty} g(x)log(f(x))dx\ .$$

For a specific probability density function $g(x)$, $\int_{-\infty}^{\infty} g(x)log(g(x))dx$ is a constant in (2.4.1). Therefore the crucial term in (2.4.1) in selecting a minimum Kullback-Leibler information model is $E_X \log f(X) = \int_{-\infty}^{\infty} g(x)log(f(x))dx$, which is the expected log-likelihood of the model probability density function f(x) with respect to $g(x)$. In selecting a minimum Kullback-Leibler number model, one seeks to maximize $E_X \log f(X)$.

Two important properties of the K-L number are:

a) $I(g;f) \geq 0$ and

b) $I(g;f) = 0$ if and only if $g(x) = f(x)$.

Let f_a and f_b be two contending estimates of the true distribution g. Then, we say that f_a is closer to g than is f_b if $I(g;f_a) < I(g;f_b)$. Thus in choosing a K-L number best model of g from a family of candidate models $\{f_m, m=1,...,M\}$, we are required to determine $m_j, 1 \leq j \leq M$ such that $I(g;f_{m_j}) \leq I(g;f_m)$, $m=1,...,M$. Equivalently, the best K-L number model can be chosen to satisfy the following

$$E_X \log f_{m_j}(X) \geq E_X \log f_m(X)\ ,\ m=1,...,M. \tag{2.4.2}$$

Consider n observations of independent and identically distributed random variables, $(x_1, \ldots, x_n)$ with unknown probability density function $g(x)$. Assume a set of alternative or candidate models $f(*, |\Theta_m|)$, where Θ_m is the parameter vector of the $f_m(*)$-th density. We wish to estimate $I(g;f_m(*,\Theta_m))$, $m = 1, \ldots, M$ the K-L number for the dissimilarity between each of the alternative models f_m and the true density g. For a member of the alternate models $f_m(*,\Theta_m)$, the likelihood is

$$\prod_{n=1}^{N} f_m(x_n, \cdots x_N|\Theta_m) = \prod_{n=1}^{N} f_m(x_n|\Theta_m).$$

From (2.4.2), the best K-L number model is the one with the greatest average expected log likelihood, $E_X \log f(X|\Theta_m)$. Assume that the parameter vectors Θ_m's are known. Then from the law of large numbers, the expected log likelihood of the model can be computed by

$$E_X \log f_m(X|\Theta_m) = \int_{-\infty}^{\infty} g(x)\log(f_m(x|\Theta_m))dx = \frac{1}{N}\sum_{n=1}^{N} \log f_m(x_n|\Theta_m). \tag{2.4.3}$$

The parameter vectors for the model distributions $f_m(*,\Theta_m)$ are unlikely to be known in practice. They must be estimated from the data and the assumed model $f_m(*,\Theta_m)$. Let $\hat{\Theta}_m$ be the maximum likelihood estimates of the parameter vector Θ_m. In this case, the law of large numbers equality in (2.4.3) does not hold when the parameter vectors Θ_m are replaced by their maximum likelihood estimates $\hat{\Theta}_m$, (Akaike 1973, Sakamoto, Ishiguro and Kitagawa 1986). That is,

$$E_x \log f(x|\Theta_m) \neq \frac{1}{N}\sum_{n=1}^{N} \log f(x_n|\Theta_m). \tag{2.4.4}$$

Thus, from (2.4.4), the maximized log likelihood is a biased estimator of the average expected log likelihood. The bias equals to the number of parameters estimated in the model. An approximate correction of the bias is reflected in the definition of the AIC given below.

$$\begin{aligned} AIC(m) &= -2(\textit{maximized log likelihood of the model}) \\ &\quad + 2(\textit{number of estimated parameters in the model}) \\ &= -2(\sum_{n=1}^{N} \log(f_m(x_n|\hat{\Theta}_m)) + 2(|\hat{\Theta}_m|)\,. \end{aligned} \tag{2.4.5}$$

In (2.4.5), $|\hat{\Theta}_m|$ denotes the dimension of the vector $\hat{\Theta}_m$.

2.4.2 Application of the AIC to Scalar AR Model Order Selection

In this section we describe the autoregressive model order selection using the AIC statistic. In practice, both the true model and the true model order of the data generating system are generally unknown. In fact, the true model is very likely not to be a finite order autoregressive model. We consider the suggestion in Akaike (1974), Parzen (1978) and Shibata (1980) that the true model is very likely infinite-dimensional. Therefore we adopt the policy that fitting an AR model to data is only an approximation to the unknown infinite-dimensional model and recognize that model order determination procedures allow us to select a finite-dimensional approximating model. When using finite-dimensional

autoregressive models to estimate the assumed infinite-dimensional truth, the AIC provides an asymptotically efficient solution to selecting the best fit to the data.

Let the family of candidate autoregressive models be defined by

$$x(n) = \sum_{j=1}^{p} \alpha_j \, x(n-j) + \varepsilon(n), \tag{2.4.6}$$

where $\varepsilon(n)$ is independently identically distributed and $\varepsilon(n) \sim N(0,\sigma^2)$. The joint probability density function for this model is

$$f(x(1), \cdots, x(n)) = \prod_{i=1}^{N} f(x(i)|\alpha_j, j=1,\ldots,p) \tag{2.4.7}$$

$$= \prod_{i=1}^{N} (2\pi\sigma^2)^{-(1/2)} \exp\left(-\frac{1}{2\sigma^2}(x(i) - \sum_{j=1}^{p} \alpha_j \, x(n-j))^2 \right).$$

The log-likelihood function for (2.4.7) is

$$\log \prod_{i=1}^{N} f(x(i)|\alpha_j, j=1,\ldots,p) = -\frac{N}{2}(\log(2\pi) + \log(\sigma^2)) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x(i) - \sum_{j=1}^{p} \alpha_j \, x(n-j))^2 \tag{2.4.8}$$

Maximizing (2.4.8) with respect to the unknown parameters, yields the maximum likelihood estimates of the AR model parameters and of σ^2, $\hat{\sigma}^2$. The latter is given by,

$$\hat{\sigma}^2 = \frac{SSE}{N}, \qquad SSE = \sum_{i=1}^{N} (x(i) - \sum_{j=1}^{p} \hat{\alpha}_j \, x(n-j))^2. \tag{2.4.9}$$

In (2.4.9), the initial conditions, that is the data $\{x(1-p), \cdots, x(0)\}$ are assumed to be available.

Substituting the quantities in (2.4.9) into (2.4.8) and taking into account the definition of the AIC in (2.4.5) yields the AIC for the scalar AR model of orer p (2.4.6),

$$AIC(p) = N[\log(2\pi) + \log(\frac{SSE}{N}) + 1] + 2(p+1). \tag{2.4.10}$$

In selecting the AIC best model, if it happens that $|AIC(p_1) - AIC(p_2)| \leq 1$ and $p_1 < p_2$, the principle of parsimony in model building requires that p_1 be selected as the approximating model order (Akaike 1974).

2.4.2.1 The AICc for Scalar AR Model

The AIC is an approximately unbiased estimator of the expected Kullback-Leibler information. Hurvich and Tsai (1989, 1992) introduced a small sample modification of the AIC, referred to as the

AIC_C. They demonstrated that in small samples, the AIC_c estimates the expected Kullback-Leibler number with less bias than does the AIC.

The AIC_c criterion for the use of autoregressive model selection as defined by Hurvich and Tsai is,

$$AIC_c = N\log(2\pi\hat{\sigma}^2) + N\frac{1+\frac{p}{N}}{1-\frac{(p+2)}{N}}. \tag{2.4.11}$$

A drawback in using AIC is that it tends to select large model order, especially in small samples. In general, the overfitting occurs when the ratio of the model order to the data length is relatively large (significantly different from zero), Hurvich (1989,1992), Shumway (1988), Linhart and Zuccini (1986). The overfitting is the result of the tendency of the AIC to become strongly negatively biased with an increase model order p when the data length N is held at constant (Hurvich 1992). The AIC_C is a small sample correction to the AIC which ameliorates that condition.

The AIC and AIC_c are related by the following

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2}. \tag{2.4.12}$$

Taking the limit of (2.4.12) as n -> ∞ we have that $AIC_c = AIC$. That is the AIC_c and AIC are asymptotically equivalent.

2.4.3 Application of the AIC to the MCAR Model Selection

Using the AIC criterion to select the best of the alternative multichannel autoregressive models is similar to that of the scalar case. Let $X(n)$ be a d-dimensional stationary stochastic process. Consider the MCAR model of order P

$$\sum_{m=0}^{P} A_m x(n-m) = \varepsilon(n)$$

with $A_0 = I_d$ the $d\times d$ identity matrix and $\varepsilon(n) \sim N(0,V)$, i.i.d.. A_m's are $d\times d$ AR matrix coefficients and V is the $d\times d$ innovations matrix. With the independence assumption, the joint probability density function is

$$f(x(1), \cdots, x(N)) = \prod_{i=1}^{N} f(x(i)|A_j, j=1,\ldots,P) \tag{2.4.13}$$

$$= (2\pi)^{-dN/2}|V|^{-1/2}\exp(-1/2\sum_{i=1}^{N} x(i)V^{-1}x'(i)) .$$

The log-likelihood function for (2.4.13) is (Sakamoto, Ishiguro and Kitagawa 1986)

$$\log\prod_{i=1}^{N} f(x(i)|A_j, j=1,\ldots, P) = -\frac{dN}{2}\log(2\pi) - \frac{N}{2}\log(|V|) - \frac{N}{2} . \tag{2.4.14}$$

Correspondingly, the AIC is

$$AIC(P) = N[d\log(2\pi) + \log|\hat{V}| + 1] + 2(Pd^2 + \frac{d(d+1)}{2}) , \tag{2.4.15}$$

where $\hat{V}$ is an estimate of innovations matrix V.

In determining the AIC best model order, a sequence of AR models are fitted to the observed data by Whittle's algorithm or the Yule-Walker solution method. The model which gives the smallest AIC is defined to be the AIC best model. As in the scalar case, if for two integers P_1 and P_2 such that $|AIC(P_1) - AIC(P_2)| \leq 1$, the principle of parsimonious modeling requires that the smaller of P_1 and P_2 be selected as the best model order.

2.4.3.1 The AICc for Multivariate AR Model

An analogous corrected AIC criterion for multichannel autoregressive model selection was developed in Hurvich and Tsai, (1991). For MCAR models, the potential of overfitting is much greater because the MCAR model contains many more unknown parameters than the univariate AR model. Let d be the dimension of a MCAR model of order P. Then the AIC_c is

$$AIC_c(P) = N(\log|\hat{\Sigma}| + d) + 2b(Pd^2 + \frac{d(d+1)}{2}) , \tag{2.4.16}$$

where $b = N / (N-(Pd + d + 1))$ is a scale factor. If the ratio $(Pd+d+1) / N$ is significantly different from zero, the bias of AIC can be large and the scale factor b plays an important role in the performance of AIC_c.

The AIC_c for the univariate case can be derived from (2.4.16) with d, the dimension, is replaced by 1.

2.4.4 Subset Selection in Time Series Using The AIC

In Chapter 3, a parsimonious scalar autoregressive model is achieved by casting the AR model as a linear regression model and doing subset selection on that model. In general if there are p independent predictors in a linear regression model, there are 2^p possible subsets. The subset selection problem is to select the best of the alternative subsets. Akaike's AIC offers a solution to that problem. Because regression analysis is the single most extensively used statistical data analysis tool particular application issues have provoked a large alternative literature to the AIC on the topic of subset selection in regression models. Some of those contributions are by Sawa and Hiromatsu (1973), Amemiya (1980,1985), Shibata (1982), Leamer (1983), and Linhart and Zucchini (1986).

Consider fitting an autoregressive model of order p

$$y(n) = \alpha_1 y(n-1) + \ldots + \alpha_p y(n-p) + \varepsilon(n) \tag{2.4.17}$$

to the time series data $\{y(n), n=1, , , N\}$. That model can be put in the form of a standard linear regression model. Define the "design" matrix X and the vectors y, ε and α by

(2.4.18)

$$X = \begin{bmatrix} y(p) & y(p-1) & \ldots & y(1) \\ y(p+1) & y(p) & \ldots & y(2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y(N-1) & y(N-2) & \ldots & y(N-p) \end{bmatrix} \quad y = \begin{bmatrix} y(p+1) \\ y(p+2) \\ \cdot \\ \cdot \\ \cdot \\ y(N) \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon(p+1) \\ \varepsilon(p+2) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon(N) \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{bmatrix}.$$

Then the scalar autoregressive time series model fitting problem is recognized to be in the linear model vector form

$$y = X\alpha + \varepsilon. \tag{2.4.19}$$

From the point of view of conventional linear regression, y_i the i-th observation in (2.4.17), can be written as

$$y_i = x_i{}^T \alpha + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.4.20}$$

where the unobserved error ε_i is assumed to be $\varepsilon_i \sim N(0,\sigma^2)$. In the conventional linear regression problem, the y and X are assumed to be independent. That is not the case when the data is a time series and the model fitted is an AR. Neverless, for convenience but knowingly, under the penalty of

incurring a bias error in the estimated AR model parameters, we continue the development.

$$f(y_1, \cdots, y_n) = \prod_{i=1}^{n} f(y_i|\alpha) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} * \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x^T\alpha)^2). \quad (2.4.21)$$

Taking the natural logarithm on both side of (2.4.21) we obtain the log-likelihood of the model

$$\log\prod_{i=1}^{n} f(y_i|\alpha) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x^T\alpha)^2 . \quad (2.4.22)$$

The maximum likelihood estimates of the unknown parameters α and σ^2, a and $\hat{\sigma}^2$, are obtained by maximize the log-likelihood (2.4.22) with respect to α and σ^2.

$$\hat{\sigma}^2 = \frac{SSE}{n}, \quad SSE = \sum_{i=1}^{n}(y_i - x^T a)^2 . \quad (2.4.23)$$

Substituting above quantities into equation (2.4.22) we obtain the maximized log-likelihood of the model

$$2\log\prod_{i=1}^{n} f(y_i|a) = n\left[\log 2\pi + \log\frac{SSE}{n} + 1\right], \quad (2.3.24)$$

and thus, the AIC of a model with p independent variables ($|\alpha| = p+1$) is

$$AIC(p) = n(\log 2\pi + \log\frac{SSE}{n} + 1) + 2(p + 1) . \quad (2.4.25)$$

2.4.4.1 The Subset Selection Algorithm

With the problem cast into an ordinary regression model in the familiar linear model form, we exploit a modified S language (Becker et al. 1988) function "leaps" and the C_p subset selection algorithm to achieve subset selection using the AIC criterion. The C_p criterion is due to Mallows (1973) and the S C_p subset selection algorithm, "leaps" is due to Furnival (1954).

The S function yields the value of Mallows C_p, Mallows (1973), for each subset. We can easily relate the value of the C_p for each individual subset to the value of the AIC for that subset by extracting the value of the sum of squares of errors SSE from the C_p using the SSE as required in the AIC. The relationships between the C_p and the corresponding SSE are given by

$$C_p(p) = \frac{1}{\hat{\sigma}^2_{FULL}} SSE(p) - n + 2p \quad (2.4.26)$$

$$SSE(p) = \hat{\sigma}^2_{FULL}[C_p(p) + n - 2p]$$

where $\hat{\sigma}^2_{FULL}$ is the (unbiased) estimate of the unknown σ^2 on the assumption that the full model, that is the one with the maximum number of independent variables yields an unbiased estimate of the unknown σ^2.

The subset selection on the individual orthogonal innovations channels yields an "instantaneous" model. In turn a linear transformation of the instantaneous model yields the familiar MCAR model form, (see Section 2.2). The MCAR model is used to evaluate the tracking performance of the full and subset selection models. The tracking performance may be used as a "diagnostic" to compare the compatibility and relative performance of the subset selection model and the Yule-Walker models.

2.5 SMOOTHNESS PRIORS IN TIME SERIES

Smoothness priors is a normal distribution theory-linear model-stochastic regression treatment of stationary and nonstationary time series. The method is quasi-Bayesian. The Bayesianess provides a framework for doing statistical inference. A prior distribution on the model parameters is expressed in the form of a stochastic difference equation and is parameterized by hyperparameters which in turn have a crucial role in the analysis. The maximization of the likelihood of a small number of hyperparameters permits the robust modeling of a time series with relatively complex structure and a very large number of implicitly inferred parameters. The critical statistical ideas in smoothness priors are the likelihood of the Bayesian model and the use of likelihood as a measure of the goodness of fit of the model.

A conceptual predecessor of smoothness priors can be seen in a smoothing problem posed by Whittaker (1923). In that problem the observations y_n, $n=1,...,N$ are given. They are assumed to consist of the sum of a "smooth" function f and observation noise or,

$$y_n = f_n + \varepsilon_n, \tag{2.5.1}$$

where $\varepsilon_n \sim N(0,\sigma^2)$. The problem is to estimate the unknown f_n, $n=1, \ldots, N$. In a time series interpretation of this problem, f_n, $n=1, \ldots, N$ is the trend of a nonstationary mean time series. A typical approach to this problem is to approximate f_n by a class of parametric polynomial regression models. The quality of the analysis is dependent upon the appropriateness of the assumed model class.

A flexible model is desirable. In this context, Whittaker suggested that the solution balance a tradeoff of goodness of fit to the data and goodness of fit to a smoothness criterion. This idea was expressed by minimizing

$$[\sum_{n=1}^{N} (y_n - f_n)^2 + \mu^2 \sum_{n=1}^{N-k} (\nabla^k f_n)^2] \tag{2.5.2}$$

for an appropriately chosen smoothness tradeoff parameter μ^2. In (2.5.2), $\nabla^k f_n$ expresses a k-th-order difference constraint on the solution f_n, with $\nabla f_n = f_n - f_{n-1}$, $\nabla^2 f_n = \nabla(\nabla f_n)$, etc. (Whittaker's original solution was not expressed in a Bayesian context. Whittaker and Robinson (1924) is a Bayesian interpretation of this problem. Greville (1957) showed that there is a unique solution to (2.5.2).)

The properties of the solution to the problem in (2.5.1)-(2.5.2) are apparent. If μ^2=0, $f_n = y_n$ and the solution is a replica of the observations. As μ^2 becomes increasingly large, the smoothness constraint dominates the solution and the solution satisfies a k-th order constraint. For large μ^2 and k=1, the solution is a constant, for k=2, it is a straight line etc.. Whittaker left the choice of μ^2 to the investigator.

Shiller (1973) modeled a related problem, the distributed lag or impulse response relationship between the input and output of economic time series. Motivated by the sparsity of data, Shiller assumed an a priori smooth distribution on the unknown impulse response in the form of a difference equation constraint. (Hence the name smoothness priors.) The solution balanced a tradeoff of infidelity of the solution to the data and infidelity of the solution to the smoothness constraint. The selection of the smoothness tradeoff parameter was determined subjectively. Akaike, (1980), completed the analysis initiated by Shiller. Akaike developed and exploited the concept of the likelihood of the Bayesian model using a maximum likelihood procedure for determining the smoothness tradeoff parameter. In Bayesian terminology, the smoothness tradeoff parameter is referred to as the "hyperparameter", (Lindley and Smith, 1972.) The work of Kitagawa and Gersch was motivated by Akaike (1980). They applied Akaike's linear regression model method to a variety of time series modeling problems not considered by Akaike and also embedded that method into a state space model framework and applied the state space method to additional problems. That work was reviewed in Gersch and Kitagawa (1988) and Gersch (1992).

Smoothness priors relates to the ill-posed problems and problems of statistical regularization, Tikhonov (1963). Also related are the "bump hunting"-penalized likelihood methods, Good and Gaskins (1980), smoothing in the generalized linear model, O'Sullivan et al. (1986), and also the work of Wahba (1991). Vigorous work, primarily at the Institute of Statistical Mathematics, Tokyo, resulted in the application of smoothness priors methods to a variety of applications. These applications include the seasonal adjustment of time series, (Akaike 1980b), tidal analysis (Ishiguro et al. 1981), binary regression (Ishiguro and Sakamoto 1983), cohort analysis (Nakamura 1986), and density estimation (Tanabe et al. 1988).

A development of smoothness priors by the linear Gausian disturbances stochastic regression model approach and which leads to least squares algorithms is in Section 2.5.2. For illustrative purposes, the least squares computational approach is applied to trend estimation in Section 2.5.2.1. In Section 2.5.2.2, the least squares computational approach to smoothness priors is applied to the modeling of a long AR model for the purposes of spectral density estimation. This approach is relevant for the smoothness priors multichannel autoregressive modeling for multichannel spectral estimation. In Section 2.5.3 the state-space computational approach to realizing smoothness priors is described. The application of the state-space smoothness priors modeling is applied to the estimation of a trend in Section 2.5.3.1. This approach is used in PARCOR modeling in Chapter 5 on the SPMCTVAR modeling.

2.5.1 Smoothness Priors Bayesian Modeling

Consider the normal disturbances stochastic linear regression model , $y = X\theta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2 I)$. Let the prior distribution constraint on θ be $\frac{1}{\lambda} D\theta \sim N(0, I)$. The joint distribution of y and θ is,

$$\begin{bmatrix} y \\ \theta \end{bmatrix} \sim N\left(\begin{bmatrix} X\theta \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 I & 0 \\ 0 & \lambda^2 D^{-1} D^{-T} \end{bmatrix} \right). \tag{2.5.3}$$

In (2.5.3), y is the vector of observed data, X and D are assumed known while σ^2 and λ^2 are unknown. The dimensions of the matrices in (2.5.3) are y: nx1; X: nxp; θ: px1. In this conjugate family Bayesian situation (Berger 1985), the mean of the posterior normal distribution of the parameter vector θ

minimizes

$$\left[\left|\left|y-X\theta\right|\right|^2+\lambda^2\left|\left|D\theta\right|\right|^2\right]. \tag{2.5.4}$$

If λ^2 were known, the posterior mean of the distribution of θ could be computed by an ordinary least squares computation. For a smoothness priors interpretation of the problem, consider $-1/2\sigma^2$ times the quadratic form to be minimized and exponentiate. Then, minimizing the quadratic form for θ maximizes

$$\exp\left\{-\frac{1}{2\sigma^2}\left|\left|y-X\theta\right|\right|^2\right\}\exp\left\{-\frac{\lambda^2}{2\sigma^2}\left|\left|D\theta\right|\right|^2\right\}. \tag{2.5.5}$$

In (2.5.5), the posterior distribution interpretation of the parameter vector θ is that it is proportional to the product of the conditional data distribution (likelihood), $p(y|X,\theta,\sigma^2)$, and a prior distribution, $\pi(\theta|\lambda^2,\sigma^2)$ on θ,

$$\pi(\theta|y,\lambda^2,\sigma^2) \;\alpha\; p(y|X,\theta,\sigma^2)\pi(\theta|\lambda^2,\sigma^2). \tag{2.5.6}$$

Integration of (2.5.6) yields $L(\lambda^2,\sigma^2)$, the likelihood for the unknown parameters λ^2 and σ^2,

$$L(\lambda^2,\sigma^2)=\int_{-\infty}^{\infty}\pi(\theta|y,\lambda^2,\sigma^2)d\theta. \tag{2.5.7}$$

Since $\pi(\theta|y,\lambda^2,\sigma^2)$ is normally distributed, (2.5.7) can be expressed in the closed form, (Akaike 1980),

$$L(\lambda^2,\sigma^2)=(2\pi\sigma^2)^{-N/2}|\lambda^2D^TD|^{1/2}|X^TX+\lambda^2D^TD|^{-1/2}\exp\left\{\frac{-1}{2\sigma^2}SSE(\hat{\theta},\lambda^2)\right\}. \tag{2.5.8}$$

In (2.5.8), the maximum likelihood estimator of σ^2 is $\hat{\sigma}^2=SSE(\hat{\theta},\lambda^2)/N$ where SSE denotes the sum of squares of errors of the linear regression problem. Then $-2\log$ likelihood is,

$$-2logL(\lambda^2,\hat{\sigma}^2)= \tag{2.5.9}$$

$$N\log 2\pi+N\log SSE((\hat{\theta}^2,\lambda^2)/N)+\log|X^TX+\lambda^2D^TD|-\log|\lambda^2D^TD|+N.$$

A practical way to determine the value of λ^2 for which the $-2\log$-likelihood is minimized, is to compute the likelihood for discrete values of λ^2 and search the discrete $-2\log$ likelihood-hyperparameter space for the minimum. The Akaike (1980) paper was very likely the first practical application of the likelihood of the Bayesian model which used the likelihood of the hyperparameters as a measure of the goodness of fit of a model to data.

2.5.2 Smoothness Priors Trend Estimation

Consider the problem of estimating the mean of a nonstationary mean time series, t_n, which is the trend at time n. In terms of the problem treated by Whittaker, $t_n = f_n$. A critical observation is that from the stochastic regression or Bayesian point of view, the difference equation constraints, $\nabla^k f_n$, are stochastic. That is, $\nabla^k t_n = w_n$, where w_n is assumed to be normally distributed with zero mean and unknown variance τ^2. For example, for k=1 and k=2 the constraints are respectively:

$$t_n = t_{n-1} + w_n; \qquad t_n = 2t_{n-1} - t_{n-2} + w_n \ .$$

A parameterization which relates the trend estimation problem to the earlier development in this section is $\tau^2 = \sigma^2/\lambda^2$. Corresponding to the matrix D in (2.5.5), for k=1 and k=2, the smoothness constraints can be expressed in terms of the following NxN constraint matrices:

$$D_1 = \begin{bmatrix} \alpha & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdot & & \\ & 0 & \cdot\ \cdot & & \\ & & \cdot\ \cdot\ \cdot & & \\ & & & -1 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} \alpha & & & & & \\ -\beta & \beta & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & 0 & \\ & & \cdot & \cdot\ \cdot & & \\ & & & \cdot\ \cdot\ \cdot & & \\ & & & \cdot & \cdot & \\ & & & & 1 & -2 \quad 1 \end{bmatrix}. \tag{2.5.10}$$

In (2.5.10), α and β are small numbers that are chosen to satisfy initial conditions.

For fixed k and fixed λ^2 the least squares solution can be simply expressed in the form of (2.5.4). For example with $k = 2$, the solution $\{t_n, n = 1, \ldots, N\}$ satisfies

$$\left\| \begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} I \\ \lambda D_2 \end{bmatrix} t \right\|^2 . \tag{2.5.11}$$

Note that the problem in (2.5.11) is a version of the Bayesian linear stochastic regression problem in (2.5.4) with $\theta = t = (t_1,\ldots,t_N)^T$, $X = I$, the NxN identity matrix, and $D = D_1$ or D_2 . From (2.5.4), the solution to (2.4.11), with D=D_2, is

$$\hat{t} = [I + \lambda^2 D_2^T D_2]^{-1} y \tag{2.5.12}$$

and the value of $SSE(\hat{\theta},\lambda^2)$ is given by

$$SSE(\hat{t},\lambda^2) = y^T y - \hat{t}^T [I + \lambda^2 D_2{}^T D_2]\hat{t} \ .$$

$\hat{t} = (t_{1|N},\ldots,t_{N|N})^T$ is the solution vector of the smoothing problem expression of (2.5.11). The least

squares problem in (2.5.11) is solved for discrete values of λ and the -2 log likelihood-hyperparameter space is searched for a minimum. From (2.5.9), the minimized value of -2log likelihood for this problem is:

$$-2logL\,(\lambda^2,\sigma^2) = \tag{2.5.13}$$

$$N log\,2\pi + N\ \log(\ SSE(\hat{t},\hat{\lambda}^2)\ /\ N\ ) + log\,|\lambda^2 D_2^T D_2 + I\,| - log\,|\lambda^2 D_2^T D_2| + N.$$

The numerical values of SSE($\hat{t},\lambda^2$) and of the determinants in (2.5.13) are transparent in a least squares algorithm analysis of (2.5.11). Since $\lambda = \sigma/\tau$, λ^2 has a noise-to-signal ratio interpretation. Smaller λ corresponds to smoother trends.

2.5.2.1 A Smoothness Priors Long AR Model For Spectral Estimation

A smoothness priors-long autoregressive (AR) model approach is used here for spectral density estimation. The smoothness priors in the estimation of the mean value of a nonstationary time series was expressed in time domain as a stochastically perturbed difference equation constraint on the evolution of the trend. Smoothness priors constraints can also be expressed in the frequency domain. In this section, the use of frequency domain priors for the estimation of the power spectral density of a stationary time series is illustrated.

The classical windowed periodogram method of spectral estimation is satisfactory for spectral analysis when the data set is "long." The alternative of spectral estimation via the fitting of parametric models to moderate length data spans became popular in the last decade, Kesler(1986). When the data span is relatively short, three facts render parametric modeling methods of spectral estimation statistically unreliable. One is the instability or small sample variability of whatever statistic is used for determining the best order of parametric model fitted to the data. The second is that usually the "parsimonious" parametric model is not a very good replica of the system that generated the data. The third is that the spectral density of the fitted parametric model cannot possibly be correct. Independent of which parametric model order is selected, there is information in the data to select models of different orders. A Bayesian estimate of power spectral density requires that the spectral density of parametric

models of different model orders be weighted in accordance with the likelihood and the prior of the model order of different models.

The smoothness priors AR model of spectral estimation alleviates this problem. A particular class of frequency domain smoothness priors is assumed for the coefficients of AR model order M, with M relatively large. The likelihood of the hyperparameters that characterize the class of smoothness priors is maximized to yield the best AR model of order M with the best data dependent priors. (A more complete treatment of the modeling discussed here is in Kitagawa and Gersch, 1985a.)

Consider the scalar autoregressive model of order M,

$$y_n = \sum_{m=1}^{M} a_m y_{n-m} + \varepsilon_n \ . \tag{2.5.14}$$

In (2.5.14), $\{\varepsilon_n\}$ is a Gaussian white noise with mean zero and variance σ^2. A least squares fit of the AR model to the data, $y_1,...,y_N$, with the first M observations $y_{1-M},y_{2-M},...,y_0$ treated as given constants, leads to the minimization of $\sum_{n=1}^{N}[y_n - \sum_{m=1}^{M} a_m y_{n-m}]^2$.

The frequency response function of the whitening filter of the AR process is

$$A(f) = 1 - \sum_{m=1}^{M} a_m \exp\,[-2\pi i m f] \ . \tag{2.5.15}$$

Consider a frequency domain smoothness priors constraint on the AR model parameters. Let measures of the unsmoothness of the frequency response function be the integrated squared kth and zero-th derivative with respect to frequency of the AR process whitening filter. That is,

$$R_k = \int_{-1/2}^{1/2} \left|\frac{d^k A(f)}{df^k}\right|^2 df = (2\pi)^{2k} \sum_{m=1}^{M} m^{2k} a_m^2 \ , \quad R_0 = \int_{-1/2}^{1/2} |A(f)|^2 df = 1 + \sum_{m=1}^{M} a_m^2 \tag{2.5.16}$$

are the terms used as penalties to the whitening filter. With these constraints, and with λ^2 and ν^2 fixed, the AR model coefficients $\{a_m, m=1,...,M\}$, minimize

$$\sum_{n=1}^{N}[y_n - \sum_{m=1}^{M} a_m y_{n-m}]^2 + \lambda^2 \sum_{m=1}^{M} m^{2k} a_m^2 + \nu^2 \sum_{m=1}^{M} a_m^2 . \tag{2.5.17}$$

In (2.5.17), λ^2 and ν^2 are hyperparameters, (Lindley and Smith 1972). By a proper choice of these parameters, the estimates of the AR model coefficients balance the tradeoff between the infidelity of the model to the data and the infidelity of the model to the frequency domain smoothness constraints.

To within a constant, the Gaussian priors on the AR model coefficients corresponding to the R_0 and R_k constraints are

$$\exp\frac{-\lambda^2}{2\sigma^2}\sum_{m=1}^{p} m^{2k}a_m^2 \quad \exp\frac{-\nu^2}{2\sigma^2}\sum_{m=1}^{M} a_m^2. \tag{2.5.18}$$

Define the matrices D and a and the matrices X and y by

(2.5.19)

$$D = \begin{bmatrix} (\nu^2+\lambda^2)^{1/2} & & & \\ & (\nu^2+2^{2k}\lambda^2)^{1/2} & & \\ & & \ddots & \\ & & & (\nu^2+M^{2k}\lambda^2)^{1/2} \end{bmatrix}, \quad X = \begin{bmatrix} y_0 & \cdots & y_{1-M} \\ y_1 & \cdots & y_{2-M} \\ \vdots & & \vdots \\ y_{N-1} & \cdots & y_{N-M} \end{bmatrix},$$

$$a = (a_1, a_2, \ldots, a_M)^T, \; y = (y_1, y_2, \cdots, y_N)^T$$

Then in a least squares computational analysis, the fitted AR model coefficients and residual sum of squares satisfy

$$\hat{a} = (X^TX + D^TD)^{-1}X^Ty, \qquad S(\lambda^2,\nu^2) = y^Ty - \hat{a}^T(X^TX + D^TD)\hat{a}. \tag{2.5.20}$$

The likelihood of the hyperparameter model is

$$L(\lambda^2,\nu^2,\sigma^2) = (\frac{1}{2\pi\sigma^2})^{N/2}|D^TD|^{1/2}|X^TX + D^TD|^{-1/2}\exp\left\{\frac{-1}{2\sigma^2}S(\lambda^2,\nu^2)\right\}. \tag{2.5.21}$$

Given λ^2 and ν^2, the maximum likelihood estimate of σ^2 is, $\hat{\sigma}^2=S(\lambda^2,\nu^2)/N$. The maximum likelihood estimates of λ^2 and ν^2 are obtained by minimizing

$$-2logL(y|\lambda^2,\nu^2,\sigma^2) = N\log 2\pi\sigma^2 - \log|D^TD| + log|X^TX + D^TD| + N \tag{2.5.22}$$

with respect to λ^2 and ν^2. Computation of the likelihood over a discrete k,λ^2,ν^2 parameter grid and searching over the resulting discrete likelihood-hyperparameter space for the minimum of -2 log likelihood yields the desired smoothness priors long AR model.

The frequency domain smoothness priors constraints used here have interpretations as constraints on the smoothness of the whitening filter of the AR model. The 0th derivative has an energy constraint interpretation. These computations allow us to remain within the framework of the general linear model and are therefore computationally tractable.

Kitagawa and Gersch (1985a) show smoothness priors long AR model results that were superior to the minimum AIC-AR model method for estimating the power spectral density of a scalar time series. Two problem situations were studied. In one case, the simulation model was in the AR model class. The other was a two sine waves in noise case, in which the data actually correspond to an ARMA model. These results support the soundness of the empirical frequency domain smoothness priors constraint approach long AR modeling method to the estimation of the power spectral density matrix.

2.5.3 State Space Modeling

A state space method for the analysis of time series is reviewed. Our concern is with the "ordinary" linear Gaussian state space model. Many problems in time series analysis can be reduced to the estimation of the state of the model. The famous Kalman filter yields computationally efficient algorithms for recursive computation. A state space modeling approach for the linear model with Gaussian system and observation noise that is equivalent of the least squares computational approach to smoothness priors modeling, was shown in Brotherton and Gersch (1981) and Kitagawa (1981). The state space smoothness priors modeling method was applied to the modeling of nonstationary mean and nonstationary covariance time series, Gersch and Kitagawa (1983a,1985) and Kitagawa and Gersch (1984,1985b). Here in Section 2.5.2.1 a general treatment of the linear Gaussian state space model analysis is shown. In Section 2.5.2.2 the method is applied to the analysis of the estimation of the mean of a nonstationary time series. This method of analysis is used in the estimation of the PARCORS in Chapter 5.

Let a state space model be given by

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n w_n \\ y_n &= H_n x_n + \varepsilon_n, \end{aligned} \tag{2.5.23}$$

where $w_n \sim N(0,Q_n)$ and $\varepsilon_n \sim N(0,R_n)$. Given the observations $y_1, \ldots, y_N$ and the initial conditions $x_{0|0}$, $V_{0|0}$, the one-step-ahead predictor and the filter are obtained from the Kalman filter algorithm:

Time Update (Prediction)

$$x_{n|n-1} = F_n x_{n-1|n-1} \tag{2.5.24}$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T.$$

Observation Update (Filtering)

$$K_n = V_{n|n-1} H_n^T [H_n V_{n|n-1} H_n^T + R_n]^{-1}$$

$$x_{n|n} = x_{n|n-1} + K_n [y_n - H_n x_{n|n-1}] \tag{2.5.25}$$

$$V_{n|n} = [\, I - K_n H_n] V_{n|n-1}.$$

Using these estimates, the smoothed value of the state x_n given the entire observation set, $y_1,...,y_N$, is obtained by the fixed interval smoothing algorithm, (Anderson and Moore 1979),

$$A_n = V_{n|n} F_n^T V_{n+1|n}^{-1}$$

$$x_{n|N} = x_{n|n} + A_n [x_{n+1|N} - x_{n+1|n}] \tag{2.5.26}$$

$$V_{n|N} = V_{n|n} + A_n [V_{n+1|N} - V_{n+1|n}] A_n^T.$$

The state space representation and the Kalman filter yield an efficient algorithm for the likelihood of a time series model. The joint distribution of $y_1,...,y_N$ is,

$$f(y_{1,...,}y_N) = \prod_{n=1}^{N} f(y_n | y_1, \ldots, y_{n-1}), \tag{2.5.27}$$

with

$$f(y_n | y_1, \ldots, y_{n-1}) = (2\pi v_n)^{-1/2} \exp\left\{ \frac{-1}{2v_n} (y_n - H_n x_{n|n-1})^2 \right\}, \tag{2.5.28}$$

$$v_n = H_n V_{n|n-1} H_n^T + R_n.$$

Then, the log likelihood, l, of the model is given by

$$l = -\frac{1}{2} \left[N \log 2\pi + \sum_{n=1}^{N} \log v_n + \sum_{n=1}^{N} \frac{-1}{2v_n} (y_n - H_n x_{n|n-1})^2 \right]. \tag{2.5.29}$$

The maximum likelihood estimate of the model parameters are obtained by maximizing (2.5.29) with respect to those parameters. The AIC is defined by

$$AIC = -2(maximum\ log\text{-}likelihood) + 2(number\ of\ parameters)\ ..sp \tag{2.5.30}$$

2.5.3.1 State-Space Trend Estimation

In this section, a state space computational procedure for the modeling of a nonstationary mean time series is discussed. Consider the nonstationary mean time series $y_n = t_n + \varepsilon_n$. As done earlier, the nonstationary trend t_n is modeled as a stochastically perturbed difference equation.

Let the trend t_n satisfy a kth order stochastically perturbed difference equation $\Delta^k t_n = w_n$, where w_n is an i.i.d. sequence with $N(0\ ,\ \tau_1^2)$. The state or system noise w_n and the observation noise ε_n are assumed to be i.i.d. with zero mean and diagonal covariance matrix

$$\begin{bmatrix} w_n \\ \varepsilon_n \end{bmatrix} \sim N\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right\}. \tag{2.5.31}$$

For example for k=2, the state-space model (2.5.22) is,

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix},\ F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix},\ G = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ H_n = [1\ \ 0]\ . \tag{2.5.32}$$

The smoothness priors problem for the trend estimation above corresponds to the maximization of

$$\exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (y_n - t_n)^2 \right\} \exp\left\{ -\frac{\tau^2}{2\sigma^2} \sum_{n=1}^{N} |\Delta^2 t_n|^2 \right\}. \tag{2.5.33}$$

The first term in (2.5.33) is identified with the conditional data distribution, and the second corresponds to the priors on the trend.

The role of the hyperparameter τ^2 as a measure of the uncertainty in the belief of the priors is clear from (2.5.33). Small values of τ^2 imply wiggly trends and large values of τ^2 imply smooth trend with respect to the kth order difference constraint. The ratio of $\frac{\tau^2}{\sigma^2}$ can be interpreted as signal-to-noise ratio.

CHAPTER 3

PARSIMONIOUS MULTICHANNEL AUTOREGRESSIVE MODELING

In this chapter the one channel at-a-time paradigm is exploited to achieve parsimonious modeling of multivariate time series. Starting with a d channel time series, via the instantaneous response-orthogonal innovations multichannel autoregressive model, these data are modeled as d independent scalar autoregressive time series in linear regression form. Subset selection is achieved within this linear regression model framework and the resulting scalar models are recombined by a linear transformation to realize the parsimonious multichannel autoregressive model. Parsimonious autoregressive modeling of both multivariate nonstationary mean and multivariate second order stationary time series is considered. In both cases our method achieves more parsimonious modeling than previously achieved. With respect to parsimonious autoregressive modeling of multivariate nonstationary mean time series our modeling method is compared to the work by Tiao and Tsay (1986). They attempted a parsimonious modeling of three simultaneous nonstationary mean time series. With respect to parsimonious autoregressive modeling of multivariate second order stationary time series (both the mean and covariance are stationary), the primary objective is to reduce the one-step-ahead prediction variance as compared to more conventional MCAR modeling of stationary time series. Our modeling method is compared to a previous attempt at achieving parsimonious modeling of bivariate stationary time series by Hsiao (1979). Our method achieves a smaller mean square tracking error of the Hsiao data example. That result indicates that ours is a superior fit to the data than is that by Hsiao. In some additional examples, we compare the mean square tracking error and one step-ahead prediction variance performance of our one channel at-a-time subset selection method with that achieved using the more conventional Yule-Walker equation-AIC model order selection solution to MCAR modeling.

3.1 BACKGROUND

As shown in the previous chapter, the unknown parameters of a stationary time series MCAR model can be estimated from the observed data by solving the multivariate Yule-Walker equations (Whittle 1963, Lutkepohl 1991). In practice, Whittle's algorithm solution of the Yule-Walker equations is employed to recursively fit successively increasing MCAR order models to the observations. The

MCAR model of order M for a d-channel time series is, (Jenkins and Watts 1968, Anderson 1971)

$$\sum_{m=0}^{M} A_m X(n-m) = \varepsilon(n), \quad E[\varepsilon(n)] = 0, \quad Var[\varepsilon(n)] = V\delta_{n,0}. \tag{3.1}$$

In (3.1), $A_m, m = 0, \ldots, M$, are the $d \times d$ AR coefficient matrices, A_0 is the $d \times d$ identity matrix and $A_m \neq 0$ for $m = M$. The $d \times d$ innovations matrix V is referred to as the innovations matrix. In fitting AR models to data, the MCAR order M can be determined with the Akaike's AIC criterion (Akaike 1974) or the AIC_C criterion (Hurvich and Tsai 1992). For the d dimensional MCAR model in (3.1), there are $d \times d \times M$ AR coefficient parameters and $d(d+1)/2$ innovations matrix parameters to be estimated. In terms of the dimension d, the number of parameters to be estimated increases as the square of d while the number of observations increases linearly in proportion to d.

In fitting MCAR model to data, the conventional Yule-Walker equations solution methods have an inherent tendency to overparameterize. The prediction of a future value of a time series uses the fitted MCAR model. Each of the MCAR model parameters is necessarily estimated with an accompanying error which in turn contributes to the prediction error variance. Hence modeling with an excessive number of estimated parameters is costly in prediction error variance and we are motivated to achieve a parsimonious MCAR modeling. We note that neither the direct Yule-Walker method nor the Whittle's algorithm method of solving the Yule-walker equations address the inherent tendency to overparameterize in MCAR model fitting. Also relevant to note here, the AR coefficient estimates in the Yule-Walker solution methods are computed using the estimated covariance function of the observed time series data. When only small number of observations are available the numerical inaccuracy in computing the empirical covariances can also contribute significantly to the prediction error. In contrast with that, the one channel at-a-time instantaneous response-orthogonal innovations modeling is performed on instantaneous data and does not bear the burden of the short data span estimation errors in the covariance computations.

3.2 METHOD

Our solution to the problem of overparameterization in the conventional Yule-Walker MCAR modeling exploits the one channel at-a-time paradigm in developing a parsimonious MCAR modeling

procedure. We fit the instantaneous response-orthogonal innovations covariance model to the multichannel time series. The instantaneous response-orthogonal innovations MCAR model permits multichannel data to be modeled one AR channel at-a-time. By virtue of the one channel at-a-time instantaneous response-orthogonal innovations representation, each orthogonal innovations channel is modeled as a scalar autoregressive linear combination of itself and the other data channels. A parsimonious MCAR model is realized by expressing each of the individual scalar AR models in the form of a linear regression model and employing a subsequent subset selection procedure. The one channel at-a-time instantaneous response-orthogonal innovations variance model is related to the usual MCAR model with general innovations variance by an invertible algebraic transformation. In the one channel at-a-time MCAR modeling of stationary time series, the order of each of the scalar AR models is determined using the Akaike AIC, (Akaike 1973,1974). Shibata (1980) demonstrated that use of the AIC for *AR* model order determination is optimal with respect to prediction variance. Here, and wherever else the AIC is referred to we refer to it generically and actually consider it as well as the Hurvich and Tsai small sample corrections for scalar and multivariate series, (Hurvich and Tsai 1989, 1992). With each scalar AR model represented in a linear regression model subset selection is achieved via an all-subsets search also using Akaike's AIC. Kitagawa (the program MULMAR in Akaike et al. 1979), was another attempt at achieving parsimonious multichannel AR modeling. Kitagawa's method was a one-channel at-a-time exploratory method. It differs from ours primarily in the subset selection procedure. There aren't very good examples of Kitagawa's method. For comparison of performance purposes, the subset instantaneous MCAR model with orthogonal innovations are algebraically transformed to the conventional MCAR model type, and model performance measures are then computed from the conventional MCAR model.

3.2.1 Some Details

The one channel at-a-time modeling of d-channel multichannel stationary time series yield d individual AR modeled data channels. To each individual channel consider fitting an autoregressive model of order p

$$y(n) = \alpha_1 y(n-1) + \ldots + \alpha_p y(n-p) + \varepsilon(n) \tag{3.3}$$

to the time series data $\{y(n), n=1, , , N\}$. In (3.3), $\{\varepsilon(n)\}$ is a zero-mean uncorrelated sequence Gaussian process. Define the design matrix X and the vectors y, ε and α by

$$X = \begin{bmatrix} y(p) & y(p-1) & \ldots & y(1) \\ y(p+1) & y(p) & \ldots & y(2) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ y(N-1) & y(N-2) & \ldots & y(N-p) \end{bmatrix} \tag{3.4}$$

$y = (y(p+1) , y(p+2) , \ldots , y(N))^T$, $\varepsilon = (\varepsilon(p+1) , \varepsilon(p+2) , \ldots , \varepsilon(N))^T$, $\alpha = (\alpha_1 , \ldots , \alpha_p)^T$.

Then the scalar autoregressive time series model fitting problem is in the linear model form

$$y = X\alpha + \varepsilon. \tag{3.5}$$

In detail, the structure of the of the linear model regression equation components corresponding to the individual channels for $d = 3$, and MCAR model order $p = 3$ are

$$y_1 = \begin{bmatrix} x_1(10) \\ x_1(11) \\ . \\ . \\ . \\ x_1(N) \end{bmatrix}, \quad X_1 = \begin{bmatrix} x_3(9) & x_2(9) & x_1(9) & . . & x_3(1) & x_2(1) & x_1(1) \\ x_3(10) & x_2(10) & x_1(10) & . . & x_3(2) & x_2(2) & x_1(2) \\ . & . & . & . . & . & . & . \\ . & . & . & . . & . & . & . \\ . & . & . & . . & . & . & . \\ x_3(N-1) & x_2(N-1) & x_1(N-1) & . . & x_3(N-9) & x_2(N-9) & x_1(N-9) \end{bmatrix}, \tag{3.6}$$

$\alpha_1 = (\alpha_1(1) , \alpha_1(2) , \ldots , \alpha_1(9))^T$.

$$y_2 = \begin{bmatrix} x_2(10) \\ x_2(11) \\ . \\ . \\ . \\ x_2(N) \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_1(10) & x_3(9) & x_2(9) & x_1(9) & \ldots & x_3(1) & x_2(1) & x_1(1) \\ x_1(11) & x_3(10) & x_2(10) & x_1(10) & \ldots & x_3(2) & x_2(2) & x_1(2) \\ . & . & . & . & . . . & . & . & . \\ . & . & . & . & . . . & . & . & . \\ . & . & . & . & . . . & . & . & . \\ x_1(N) & x_3(N-1) & x_2(N-1) & x_1(N-1) & \ldots & x_3(N-9) & x_2(N-9) & x_1(N-9) \end{bmatrix},$$

$\alpha_2 = (\alpha_2(1) , \alpha_2(2) , \cdots , \alpha_2(9) , \alpha_2(10))^T$.

$$y_3 = \begin{bmatrix} x_3(10) \\ x_3(11) \\ . \\ . \\ . \\ x_3(N) \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_2(10) & x_1(10) & x_3(9) & x_2(9) & x_1(9) & . . & x_3(1) & x_2(1) & x_1(1) \\ x_2(11) & x_1(11) & x_3(10) & x_2(10) & x_1(10) & . . & x_3(2) & x_2(2) & x_1(2) \\ . & . & . & . & . & . . & . & . & . \\ . & . & . & . & . & . . & . & . & . \\ . & . & . & . & . & . . & . & . & . \\ x_2(N) & x_1(N) & x_3(N-1) & x_2(N-1) & x_1(N-1) & . . & x_3(N-9) & x_2(N-9) & x_1(N-9) \end{bmatrix},$$

$$\alpha_3 = (\alpha_3(1) , \alpha_3(2) , \cdots , \alpha_3(9) , \alpha_3(10) , \alpha_3(11))^T .$$

In (3.6), the notation y_1, y_2, y_3 refer respectively to the dependent variable associated with the individual channels, X_i, i = 1,2,3 refers to the design matrix of the i-th channel and $\alpha_1,\alpha_2,\alpha_3$ refer to the estimated i-th channel AR parameters. Equation (3.6) clearly shows that each dependent variable, y_1, y_2, and y_3 is a linear function of the interleaved data from each of the data channels. Also note that if p is the order of the MCAR model then $p_k = d \times p + (k - 1)$, k = 1,2,3 where k is the channel number and p_k is the scalar AR model order in the k-th channel.

With the problem cast into an ordinary regression model in the familiar linear model form, we exploit a modified S language (Becker et al. 1988) function "leaps" and the C_p subset selection algorithm to achieve subset selection using the AIC_C (Hurvich and Tsai 1989), criterion. The AIC_C is a small sample modification of the Akaike AIC, (Akaike 1974). The C_p criterion is due to Mallows (1973) and the S language subset selection algorithm, "leaps" is due to Furnival (1954). (We also note that it is only a nominal exercise to implement an alternative nonparametric model selection criterion such as PRESS, Allen 1974, in the S environment without resorting to the leaps algorithm and without invoking the assumption of the normality of the $\{\varepsilon(n)\}$ process.)

The subset selection on the individual orthogonal innovations channels yields an instantaneous model which is then transformed into the conventional MCAR form. The MCAR is used to evaluate the tracking performance of the full and subset selection models. Mean square error tracking performance may be used as a heuristic to compare the relative performance of the instantaneous response-orthogonal innovations subset selection model and the Yule-Walker models.

3.3 SOME MODELING EXAMPLES

Here we compare the attempts at achieving parsimonious time series modeling in the $d = 3$ channel nonstationary mean time series modeling by Tiao and Tsay (1986) and the bivariate stationary time series modeling by Hsiao (1979) with our own approach. While our model is more parsimonious than that by Hsiao, his data set is not a very satisfactory example. We show a more satisfactory $d = 3$ stationary time series example of the modeling and tracking of bathymetry data.

3.3.1 The Tiao and Tsay Nonstationary Mean Data Example

In this comparative study we illustrate the parsimonious property of our MCAR model instantaneous response-orthogonal innovation subset selection procedure as an application to the $d = 3$, $n = 100$ flour price data considered by Tiao and Tsay (1985), (1989). The Tiao-Tsay series data is illustrated in Figure 3.1. The nonstationarity of the mean is evident. Respectively Tiao and Tsay (1985), (1989) consider transformations of the original series, canonical correlation analyses, AR_2 and AR_1MA_1 models. The fitted AR_2 model was a full model with 18 AR parameters and 5 innovation variance parameters. The principal objective of the 1989 paper is to achieve parsimonious modeling. The 1989 AR_1MA_1 model conceptually can have as many as 18 AR-MA parameters and 6 innovation matrix parameters. The Tiao-Tsay model fitted to the flour price data consists of 9 AR-MA parameters and 5 innovations variance parameters. In comparison with that latter model our AR_2 instantaneous response-orthogonal innovation subset selection procedure model fitted 11 AR parameter, 2 parameters in the B_0 matrix, and 3 instantaneous variance matrix parameters. That is a total of 16 parameters or 2 parameters more than the Tiao-Tsay model. Of course we should take into account the fact that the Tiao-Tsay model also requires computation, via a fairly complex analysis procedure, of a 3×3 transformation matrix. Consequently, we feel that our own rather simple computational procedure is at least a reasonable parsimonious model contender if not superior to the Tiao-Tsay modeling of the flour price data. The excellent tracking results achieved by our model of the Tiao-Tsay data is shown in Figure 3.2. For completeness the B_i, i=0,1,2 parameter matrices fitted to the flour price data are

$$B_0 = \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 \\ 1.0393 & 0.0000 & 0.0000 \\ 0.9832 & 0.0000 & 0.0000 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0.0000 & 1.0902 & 0.0000 \\ -0.5466 & 0.5259 & 0.0000 \\ 0.0000 & -0.8081 & 0.8956 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 1.1132 & -1.4888 & 0.2326 \\ -0.4569 & 0.4357 & 0.0000 \\ 0.0000 & -0.0865 & 0.0000 \end{bmatrix}. \quad D = \begin{bmatrix} 44.0000 & 00.0000 & 00.0000 \\ 00.0000 & 11.8192 & 00.0000 \\ 00.0000 & 00.0000 & 12.1898 \end{bmatrix}.$$

As a result of the non-negligible feed through terms B_{120} and B_{130} the innovations variance in the first channel, $D[1,1]$, is considerably larger than in channels 2 and 3, $D[2,2]$,$D[3,3]$.

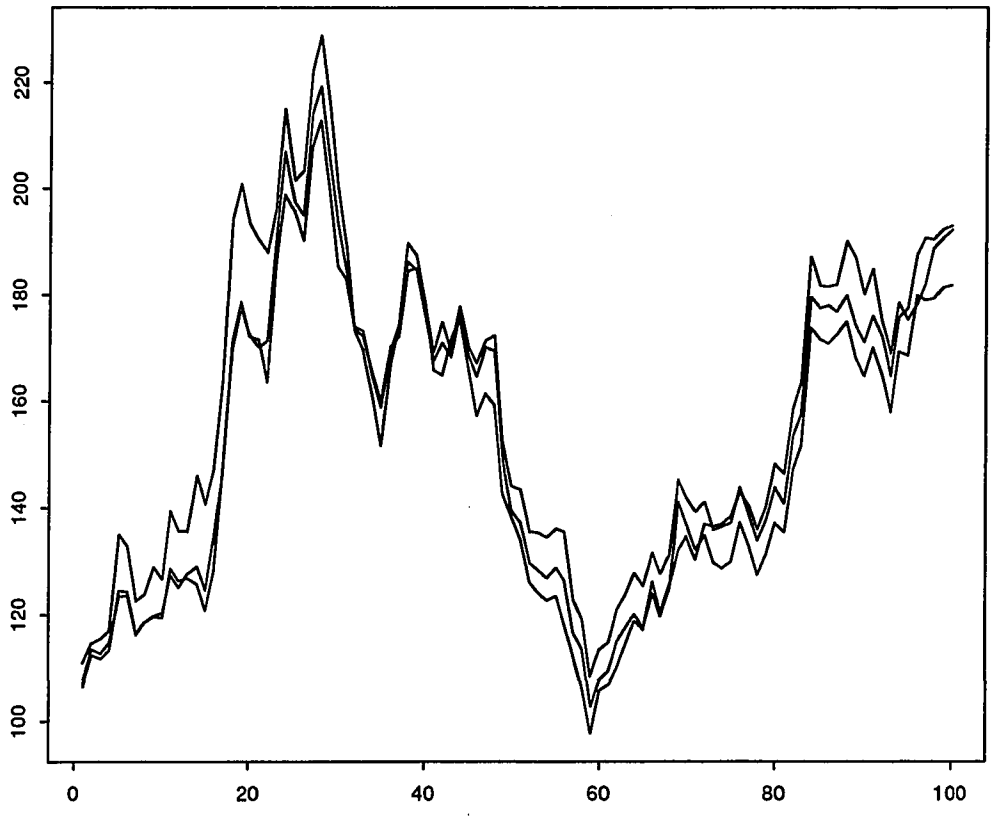

Figure 3.1: Flour Price Time Series Data

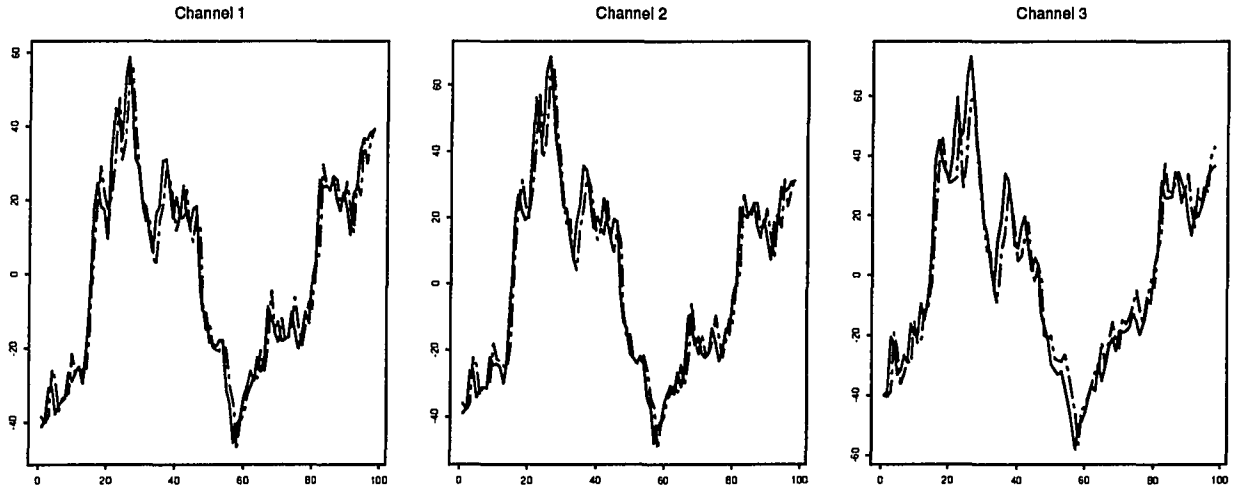

Figure 3.2: Flour Price Data Tracking

3.3.2 The Hsiao Second Order Stationary Time Series Model

Here we compare results of our one channel at-a-time subset selection approach with the MCAR stationary time series modeling in Hsiao(1979). Hsiao's model for Canadian post war money (M1) and income (GNP) series is in the form:

$$\begin{bmatrix}(1-L)^2\log GNP \\ (1-L)^2\log M1\end{bmatrix} = \begin{bmatrix}\psi_{11}^{m_{11}}(L) & \psi_{12}^{m_{12}}(L) \\ \psi_{21}^{m_{21}}(L) & \psi_{22}^{m_{22}}(L)\end{bmatrix} \begin{bmatrix}(1-L)^2\log GNP \\ (1-L)^2\log M1\end{bmatrix} + \begin{bmatrix}a \\ b\end{bmatrix} + \begin{bmatrix}u \\ v\end{bmatrix}, \tag{3.7}$$

where $\psi_{ij}(L) = \sum_{l=1}^{M_{ij}} \psi_{ijl} L^l$, L is a lag operator, $Ly_t = y_{t-1}$, and u, v are zero mean white noise with $cov(u,v)_{s,t} = \delta_{t,s}\Omega$ and a and b are constants (mean values).

To determine the orders m_{11}, m_{12}, m_{21}, and m_{22}, the Hsiao procedure is as follows:

First consider modeling the GNP series.

1) Determine the order of GNP series as a scalar autoregressive process using Akaike's FPE criterion. Assume that order is p and refer to this as the FPE(p,0) model.

2) With the order of GNP fixed at p, add components of ψ_{12}, (the $M1$ process) to the p components of the GNP series, one-component at-a-time and use the FPE to determine q the FPE best number of $M1$ components to be added. This yields a $FPE(p,q)$ component model.

3) Now "reverse" the process and fixing the order of the $M1$ components at q, starting at zero components, add components of ψ_{11}, the GNP series, one lag component at-a-time and using the FPE, determine FPE best FPE(p',q) model.

4) Finally, using the FPE criterion select the FPE best of the FPE(p',q) and FPE(p,0) models.

Repeat the process for the M1 series.

Using the procedure specified above, Hsiao identified the model orders m_{11}, m_{12}, m_{21} and m_{22} to be 6, 8, 4 and 9, respectively. The resulting total number of AR parameters estimated is 27. In addition, 3 parameters were estimated to obtain the innovation variance matrix.

In our approach, because the use of an instantaneous response model, the choice of which series is modeled first is significant. We chose the $M1$ series first and allow that series to have an instantane-

ous effect on GNP. Here we use a notation that is similar to that of Hsiao's and the model is

$$\begin{bmatrix} (1-L)^2 log(M1) \\ (1-L)^2 log(GNP) \end{bmatrix} = \begin{bmatrix} \psi_{11}^{m_{11}}(L) & \psi_{12}^{m_{12}}(L) \\ \psi_{21}^{m_{21}}(L) & \psi_{22}^{m_{22}}(L) \end{bmatrix} \begin{bmatrix} (1-L)^2 log(M1) \\ (1-L)^2 log(GNP) \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix}, \tag{3.8}$$

where $\psi_{21}(L) = \sum_{l=0}^{M_{21}} \psi_{21l} L^l$, and everything else remains the same as defined in Hsiao's model.

Computational experiments suggested that the maximum value of 10 for m_{11}, m_{12}, m_{21} and m_{22} was sufficient. This maximum value is also compatible with Hsiao's model. We then apply our procedure and we found that only 4 of the 10 ψ_{11}'s, only 4 of the 10 ψ_{12}'s, 6 of the 11 ψ_{21}'s, and 7 of the 10 ψ_{22}'s were nonzero values. That is, a total of 21 AR parameters were estimated using our approach as compared to 27 AR parameters using Hsiao's approach. Also only 2 additional parameters are required to identify the innovations variance matrix in our model.

We do not consider the M1, GNP series to be a very good example of AR model fitting. Nevertheless, the mean square errors for the M1 and GNP series respectively for Hsiao's model are 0.000181 and 0.000123. Correspondingly the means square errors for the M1 and GNP series for our model are 0.000197 and 0.000110.

This kind of model fit or tracking performance result revealed in this example is typical of performance that can be expected from our procedure in comparison with other procedures. The first channel modeled generally will have a tracking performance comparable with that of other procedures, however subsequent channels will be better fitted by our procedure in the sense of smaller means square error than other procedures. Very likely this is a consequence of the "feedthrough" of the instantaneous response.

3.3.3 Stationary Tirvariate Time Series Bathymetry Data

In this example we consider 3-channels of stationary time series bathymetry data, (bathymetric data is a measure of the depth of the ocean). The data is a 3 channel subset of the 15 channel data in Figure 18 in Goff and Jordan (1988). The tracking performance achieved by our method is compared to that achieved by the more conventional Yule-Walker AIC MCAR modeling method. We used a

reduced data set of length, N = 60. The conventional MCAR model was fitted by Yule-Walker AIC_C modeling method. An AR_2, $d = 3$ model was fitted to the data. The number of parameters estimated in the conventional MCAR model is 24; $d \times d \times p = 3 \times 3 \times 2$ AR parameters and $d \times (d + 1)/2 = 3 \times (3+1)/2$ innovations matrix paramenters. On the other hand, using our procedure only 8 AR parameters and 3 parameters of the matrix D, or a total of 11 parameters were required to fit the instantaneous response-orthogonal innovations model.

The tracking performance of both the Yule-Walker method of modeling and the one channel at-a-time subset procedure are shown in Figure 3.3. The one-channel at-a-time model was converted to a conventional MCAR model. This permits the tracking for both procedures to be computed in the same way.

The mean square tracking errors for the Yule-Walker AIC model and our subset instantaneous response-orthogonal innovations MCAR model are shown in Table 3.1.

Table 3.1: Bathymetry Data Tracking Error Performance

	MEAN SQUARED TRACKING ERRORS		
	Channel 1	Channel 2	Channel 3
Whittle-AIC	3140.527	3408.540	4265.773
Subset MCAR	2965.613	3300.645	3977.945

As is seen in Figure 3.3, and the numerical results in Table 3.1, the tracking performance of our model is superior to that achieved by conventional MCAR modeling.

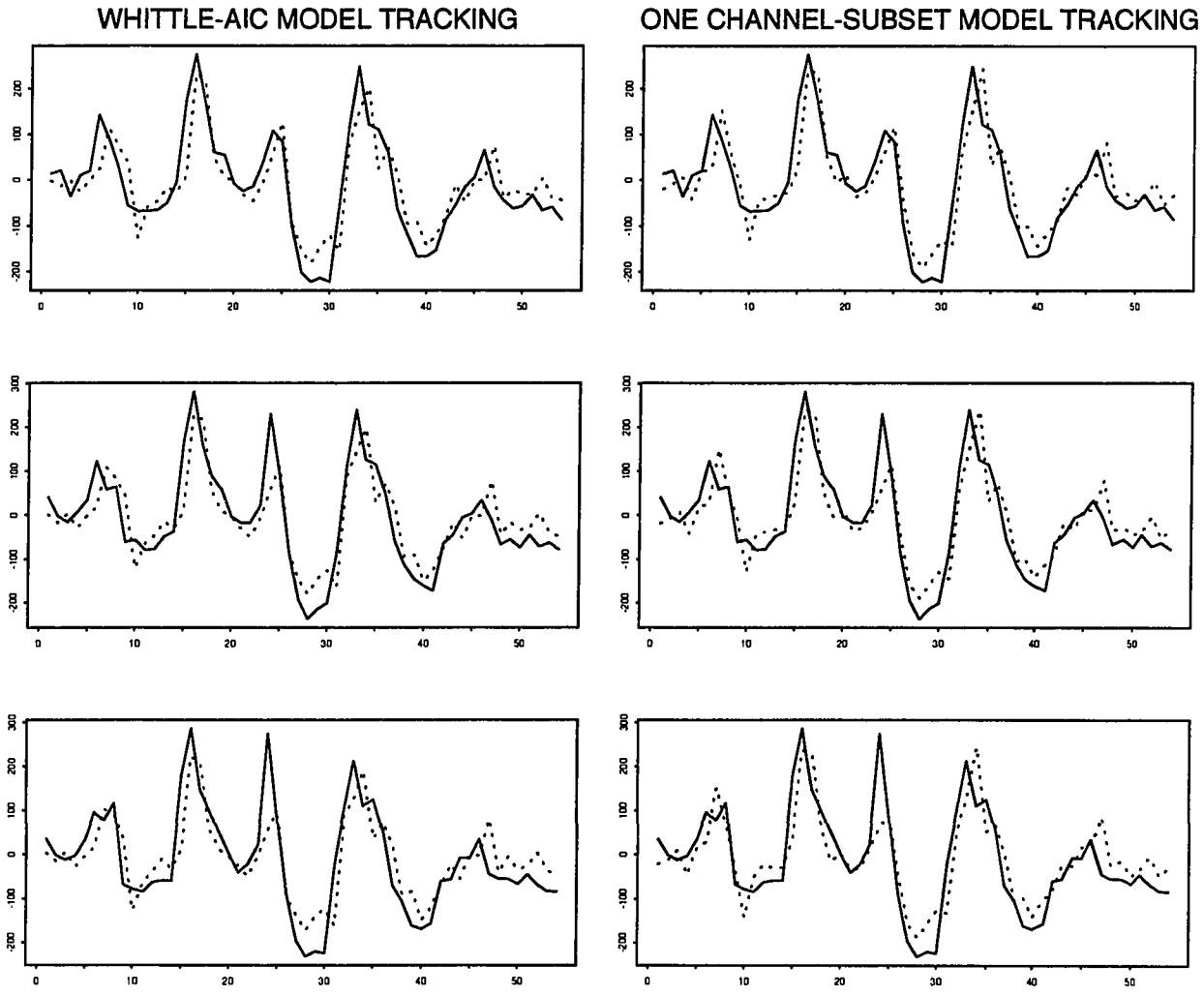

Figure 3.3: Bathymetry Data Tracking

3.4 THE NEWTON MODEL OF THE BOX-JENKINS SERIES J DATA

The Box-Jenkins Series J data (Box-Jenkins 1970, p. 532-533) is a bivariate time series data record of the input and output of a real physical gas furnace plant. That data has been used in numerous studies. A stationary autoregressive model of order 4 of that data is in Newton (1988), p. 342. We use that model in several examples. In one, a detailed example of the workings of the instantaneous response-orthogonal innovations model and subset selection is shown for an $n = 300$ simulated data set from the Newton model. In that context we compare the tracking of that data by both the Yule-Walker AIC MCAR model and our parsimonious model. Separately, we also do a Monte Carlo simulation study, based on data simulated from the Newton model, to compare one one-step ahead prediction performance of the Yule-Walker AIC MCAR procedure and our procedure.

3.4.1 A Detailed Example of Parsimonious MCAR Modeling

The Newton bivariate autoregressive model of order 4 is in the form

$$x(n) = \sum_{i=1}^{4} A_i x(n-i) + \varepsilon(n), \quad E[e(n)] = 0, \quad Cov[e(n), e(n')] = V\delta_{n,n'}. \tag{3.9}$$

The parameters of that model are

$$V = \begin{bmatrix} 0.35155 & -0.07314 \\ -0.07314 & 0.97240 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} -1.925889 & 0.001240 \\ -0.050508 & -1.299794 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1.201669 & -0.004222 \\ 0.020493 & 0.327703 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} -0.116929 & 0.008672 \\ 0.711791 & 0.257010 \end{bmatrix}, \quad A_4 = \begin{bmatrix} -0.104232 & -0.003270 \\ -0.195398 & -0.133420 \end{bmatrix}.$$

The equivalent instantaneous response-orthogonal innovations matrix representation may be computed by simple algebraic manipulation as shown in Chapter 2. The instantaneous response AR matrix coefficient B_0 is

$$B_0 = \begin{bmatrix} 1.0000 & 0.0000 \\ 0.2081 & 1.0000 \end{bmatrix}$$

where $b_{210} = -V(1,2)/V(1,1)$. With $B_m = B_0 A_m$, $m = 1, \ldots, 4$ and $D = B_0 V B_0^T$, the equivalent

theoretical instantaneous response-orthogonal innovations model. B_i and D matrices are,

$$B_1 = \begin{bmatrix} -1.9259 & 0.0012 \\ -0.4512 & -1.2995 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1.2016 & -0.0042 \\ 0.2705 & 0.3268 \end{bmatrix}, \quad B_3 = \begin{bmatrix} -0.1169 & 0.0087 \\ 0.6875 & 0.2588 \end{bmatrix},$$

$$B_4 = \begin{bmatrix} -0.1042 & -0.0033 \\ -0.2171 & -0.1341 \end{bmatrix}, \quad D = \begin{bmatrix} 0.3516 & 0.0000 \\ 0.0000 & 0.9572 \end{bmatrix}.$$

An $n = 300$ sample time series was simulated using Newton's model and the conventional Whittle-AIC MCAR model was fitted to that data. The simulated data for this example are shown in Figure 3.4.

Whittle's algorithm yields successively increasing order multivariate autoregressive models recursively. The model order was selected using the Akaike's AIC. The AIC best model is the one with the smallest AIC where the AIC computed for a multivariate AR model of order p with d channels of data is

$$AIC(p) = N[d\log(2\pi) + \log|\hat{V}| + 1] + 2(pd^2 + \frac{d(d+1)}{2})$$

and $\hat{V}$ is the estimated innovations variance. In this example, a maximum order 12 MCAR model was considered for the data. The values of the AIC computed for each MCAR order model fitted are shown in Table 3.2.

Table 3.2: MCAR Model Order Selection

AIC VALUES FOR EACH MODEL ORDER CONSIDERED						
Order	1	2	3	4*	5	6
AIC	346.561	-134.782	-154.677	-165.190	-164.963	-157.589
Order	7	8	9	10	11	12
AIC	-150.524	-145.820	-138.933	-135.038	-131.969	-124.723

The best AIC model selected for this example was AR_4. With this model a total of 16 AR parameters and 3 innovations variance parameters were fitted. The estimated AR matrix coefficients and the innovations matrix are

$$\hat{V} = \begin{bmatrix} 0.41430 & -0.18941 \\ -0.18941 & 1.32092 \end{bmatrix} \tag{3.10}$$

$$\hat{A}_1 = \begin{bmatrix} -1.76670 & -0.04970 \\ -0.08796 & -1.16936 \end{bmatrix}, \hat{A}_2 = \begin{bmatrix} 1.02914 & 0.06975 \\ 0.17659 & 0.20347 \end{bmatrix},$$

$$\hat{A}_3 = \begin{bmatrix} -0.11175 & -0.01232 \\ 0.32766 & 0.32180 \end{bmatrix}, \hat{A}_4 = \begin{bmatrix} -0.10972 & -0.01870 \\ 0.21605 & -0.13952 \end{bmatrix}.$$

The equivalent instantaneous response-orthogonal innovations representation for the estimates derived as before by algebraic transformations yields for $\hat{V}_D$ and B_m , $m = 0, \cdots, 4$

$$\hat{V}_D = \begin{bmatrix} 0.41430 & 0.00000 \\ 0.00000 & 1.23432 \end{bmatrix}, \quad B_0 = \begin{bmatrix} 1.00000 & 0.00000 \\ 0.45718 & 1.00000 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} -1.76670 & -0.04970 \\ -0.89566 & -1.19208 \end{bmatrix}, B_2 = \begin{bmatrix} 1.02914 & 0.06975 \\ 0.64709 & 0.23536 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -0.11175 & -0.01232 \\ 0.27657 & 0.31617 \end{bmatrix}, B_4 = \begin{bmatrix} -0.10972 & -0.01870 \\ 0.16589 & -0.14807 \end{bmatrix},$$

The data in Figure (3.4) was also modeled with the parsimonious MCAR modeling scheme. Under the one channel at-a-time paradigm, the two channel time series data were modeled as instantaneous response MCAR. That modeling yields a set of two independent scalar AR time series which in a linear regression model setting are suitable for subset selection and yields the subset AR coefficients. Akaike's AIC was used to determine the best subset. The *AR* coefficients from the two scalar models are combined to give the instantaneous MCAR model.

Correspondingly, the parameters of that reconstituted MCAR model are

$$\hat{V} = \begin{bmatrix} 0.32351 & -0.08685 \\ -0.08685 & 0.88340 \end{bmatrix}$$

$$\hat{A}_1 = \begin{bmatrix} -1.87739 & 0.00000 \\ -0.007875 & -1.20384 \end{bmatrix}, \quad \hat{A}_2 = \begin{bmatrix} 1.28118 & -0.00000 \\ 0.03960 & 0.22415 \end{bmatrix},$$

$$\hat{A}_3 = \begin{bmatrix} -0.34029 & 0.00000 \\ 0.63205 & 0.31630 \end{bmatrix}, \quad \hat{A}_4 = \begin{bmatrix} -0.00000 & -0.00000 \\ -0.00000 & -0.13934 \end{bmatrix}.$$

As anticipated, the diagonal terms in the estimated innovations variance in the reconstituted MCAR model are smaller than the corresponding diagonal terms in the original Yule-Walker AIC model fitted to the sample data. Also as might have been anticipated in the reconstituted MCAR model, zeros replace the relatively small values in the original Yule-Walker AIC model fitted to the sample data.

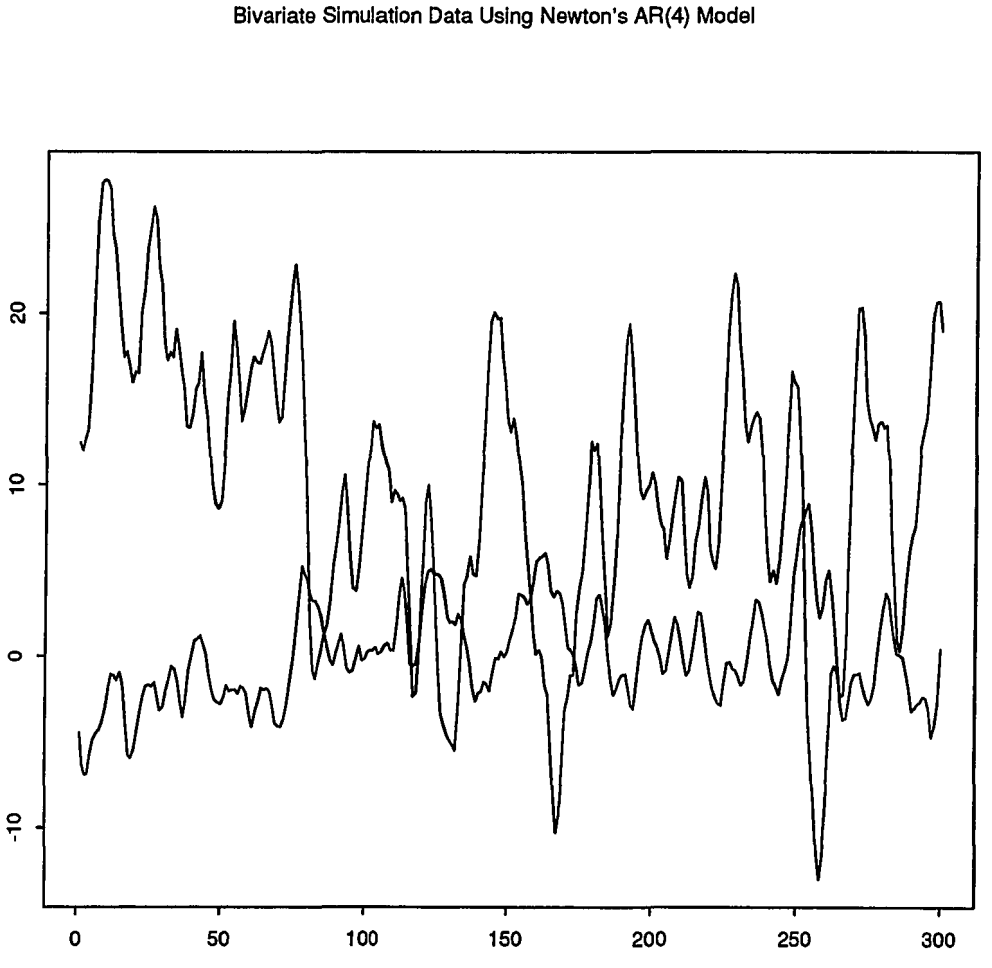

Figure 3.4: Bivariate Simulation Data Using Newton's AR(4) Model

3.4.1.1 Subset Selection Documentation for the Example

The 2 scalar time series obtained from the instantaneous response-orthogonal innovations model are:

$$y_1(n) = \sum_{i=1}^{p_1} \alpha_1(i) x_1(n-i) + \sum_{i=1}^{p_1} \alpha_2(i) x_2(n-i) + \varepsilon_1(n)$$

$$y_2(n) = \alpha_{210} x_1(n) + \sum_{i=1}^{p_2} \alpha_1(i) x_1(n-i) + \sum_{i=1}^{p_2} \alpha_2(i) x_2(n-i) + \varepsilon_2(n).$$

For the sample data set, $p_1 = 4$ and $p_2 = 4$. Correspondingly, a total of 8 variates can be fitted to the model for $y_1(n)$ and and 9 variates can be fitted to the model for $y_2(n)$. We point out that for a regression model with m independent variables there are 2^m possible subsets. In our computations an all-subset-search is conducted. The subset returned from the procedure for y_1 contains only 3 (three) AR coefficients $\{\alpha_1(1), \alpha_1(2), \alpha_1(3)\}$. The AR model coefficients of the best subset for y_1 were respectively, {1.877386, -1.281184, 0.3402851}.

For y_2, the subset contains 8 variables with the instantaneous term α_{210}, and $\alpha_1(1) \cdots \alpha_1(3)$, and $\alpha_2(1) \cdots \alpha_2(4)$ included in the subset. Correspondingly, the AR model coefficients of the best subset for y_2 were {-0.2684586, 0.5827482, -0.383548, -0.540701, 1.203842, -0.2241519, -0.3162988, 0.139338}.

The resulting parsimonious MCAR in the instantaneous response-orthogonal innovations MCAR form is

$$\hat{D} = \begin{bmatrix} 0.35538 & 0.00000 \\ 0.00000 & 0.91810 \end{bmatrix}, \; B_0 = \begin{bmatrix} 1.00000 & 0.00000 \\ -0.26846 & 1.00000 \end{bmatrix}, \; B_1 = \begin{bmatrix} 1.87739 & 0.00000 \\ 0.58275 & 1.20384 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} -1.28118 & 0.00000 \\ -0.38355 & -0.22415 \end{bmatrix}, \; B_3 = \begin{bmatrix} 0.34029 & 0.00000 \\ -0.54070 & -0.31630 \end{bmatrix}, \; B_4 = \begin{bmatrix} 0.00000 & 0.00000 \\ 0.00000 & 0.13934 \end{bmatrix}.$$

Comparing the number of parameters fitted with the conventional MCAR model, a reduction of 3 (three) AR parameters in the model was realized. An important point to note is that that the diagonal elements of the estimate $\hat{D}$ for the subset model are, respectively smaller than their counter parts in $\hat{V}_D$.

The mean squared tracking errors are computed for both models and the results are tabulated in Table 3.3. Respectively for channel 1 and 2, the tracking errors are smaller with the parsimonious

MCAR.

Table 3.3: Simulation Example Tracking Error Performance

	MEAN SQUARED TRACKING ERRORS	
	Channel 1	Channel 2
Whittle-AIC	0.32765	0.90120
Subset MCAR	0.32351	0.88340

3.4.2 A Monte Carlo Prediction Study

An anticipated property of the parsimonious MCAR is that it will have a smaller one-step-ahead prediction error variance than the conventional Yule-walker AIC method of MCAR modeling. In order to verify that property we conducted a Monte Carlo simulation study. 100 different simulations of 300 data points each were done from the Newton model. For each simulation trial the 300th data point is predicted using the two models, the conventional and the parsimonious MCAR models. The difference between the predicted value and the theoretical value is recorded, and the variances of the differences are computed as an estimate of the one-step ahead prediction error variance. Typically the conventional MCAR model fits of the simulated AR_4 data have an order of 4 with 16 AR model parameters and 3 innovations variance parameters. Typically, the parsimonious MCAR subset selection model required only 12 parameters to be fitted to the data. The prediction variance of the subset selection method was approximately 70% and 89% respectively of that achieved by the Yule-Walker solution method for channels one and two. The one step-ahead prediction variances for the Yule-Walker AIC MCAR method and the subset selection method are shown in Table 3.4.

Table 3.4: Prediction Error Variance Performance

	PREDICTION VARIANCES	
	Channel 1	Channel 2
Whittle-AIC	0.41777	1.28508
Subset MCAR	0.29445	1.15461

CHAPTER 4

SPMCAR MODELING FOR SPECTRUM ESTIMATION

In this chapter the one channel at-a-time paradigm is exploited in conjunction with the smoothness priors modeling of scalar autoregressive time series for the purposes of achieving multivariate autoregressive spectrum estimation for stationary covariance time series.

The use of multivariate autoregressive models for spectrum estimation is mentioned in Jenkins and Watts (1968). Early uses of multivariate autoregressive models for spectrum estimation appear in Gersch (1970), Gersch and Goddard (1970), (applied to the analysis of electroencephalograms), Akaike (1972), (in a control theory application), and Jones (1974), (also in electroencephalogram analysis). The analysis of the statistical properties of multivariate spectrum estimation has primarily been concentrated on Fourier analysis methods, (Brillnger 1975, Koopmans 1974). Those methods lead more naturally to sharp analytic results. Gersch (1978) is an early treatment of the statistical performance of the autoregressive model method for multidimensional spectral analysis. Contemporary treatment in engineering texts, Kay (1988) and Marple (1988) for example, confine their discussion only to the simplest, of $d = 2$, autoregressive model of order 1, examples in the use of autoregressive models for spectral analysis.

Our emphasis here is on examining the conjecture that the one channel at a-time smoothness priors modeling approach might yield superior power spectrum estimation than the more conventional approach in which the autoregressive modeling is achieved via the solution of the multivariate version of the Yule-Walker equations.

4.1 BACKGROUND

Given N observations of a d-dimensional zero-mean stationary time series $\{x(t), t=1, \cdots, N\}$, the power spectral density matrix may be estimated by fitting a MCAR model to the observed data. The MCAR model is

$$\sum_{i=0}^{P} A_i x(n-i) = \varepsilon(n), \quad E[\varepsilon(n)] = 0, \quad COV[\varepsilon(n),\varepsilon(n')] = V\delta_{n,n'}. \tag{4.1}$$

In (4.1), the $d \times d$ matrix V is referred to as the innovations matrix and $A_0 = I$ the $d \times d$ identity matrix. The model order P along with the unknown matrix parameters $V, A_1, \cdots, A_P$ are estimated from the data. Akaike's AIC may be used to determine the model order P. As discussed in Chapter 2, the AR matrix coefficients are obtained by solving the Yule-Walker equations. The power spectral density matrix $S(f)$ of the stationary time series $\{X(t), t=1, \cdots, N\}$ is given by

$$S(f) = A^{-1}(f) V A^{*}(f)^{-T} . \tag{4.2}$$

In (4.2), $A(f)$, the system "whitening filter" is $A(f) = \sum_{k=0}^{P} A_k \exp(-2i\pi fk)$, $A^{*}(f)$ is the complex conjugate of $A(f)$ and the normalized frequency domain over which the spectrum is computed is $-1/2 \leq f \leq 1/2$.

A justification for attempting to achieve spectrum estimation by means other than a direct application of (4.2) to the fitted MCAR model in (4.1) can be appreciated from the following argument: First, from the Bayesian point of view, the spectral density computed from the fitted parametric model cannot possibly be optimum. A Bayesian estimate of power spectral density would require that the spectral density of parametric models of different model orders be weighted in accordance with the likelihood and the prior of the model order of different models. Secondly, for short data spans, the small sample variability of whatever statistic used for determining the best order of the parametric model fitted to the data makes parametric model fitting statistically unreliable. Thus we are motivated in the problem of multivariate spectral estimation to attempt to exploit the quasi-Bayesian smoothness priors method that worked very well in the spectrum estimation of scalar autoregressive time series (Kitagawa and Gersch 1985b).

4.2 THE APPROACH

Our solution to the estimation of the power spectrum density using MCAR model is realized by extending the smoothness priors long AR modeling method for scalar time series to the multichannel case. Such an extension is made possible by exploiting the one channel at-a-time paradigm. As in Chapters 2 and 3, the data from the individual channels is interlaced to permit an instantaneous response-orthogonal innovations representation of the multichannel AR model. In that way the data can

be modeled by a collection of scalar AR models to which the frequency domain smoothness priors can be applied. The concept of smoothness priors-Bayesian modeling of scalar time series was reviewed in Chapter 2. Kitagawa and Gersch (1985b) demonstrated a smoothness priors computational procedure for the fitting of long AR models to stationary time series data for the purposes of spectrum estimation. There the smoothness priors long AR modeling method was compared to the minimum AIC-AR modeling for scalar time series power spectrum estimation. It showed superior results in application to real physical data and to simulations. In the simulations, two cases were examined. In one the simulation model was in the AR model class. In the other the simulation model was an ARMA model. The ARMA model corresponds to an infinite order AR model. Thus, in this case, the AR model was only an approximation to the modeling class.

In attempting to achieve enhanced power spectrum estimation in the multivariate stationary time series case, the one channel at-a-time and smoothness priors paradigms, are combined into an SPMCAR (smoothness priors multichannel autoregressive modeling) model. As in Kitagawa and Gersch (1985b), we fit a long AR order model with smoothness priors constraints on the AR model coefficients to each of the d scalar AR time series. The individual scalar AR models are then combined to yield the SPMCAR model which in turn yields the estimation of the power spectral density matrix of a stationary d-channel time series. In each autoregressive time series channel, the priors on the AR parameters are the integrated squared kth derivative with respect to frequency of the departure from model smoothness. The estimation of the autoregressive model parameters and an additional small number of hyperparameters is required. The maximization of the likelihood of the hyperparameters is the critical computation.

4.3 THE METHOD

The smoothness priors multichannel autoregressive modeling for the estimation of stationary multichannel time series power spectrum density is shown in this section. Much of the background development have been given in Chapter 2. Given observations from a d-channel stationary covariance time series, the instantaneous response-orthogonal innovations MCAR model (4.1) is fitted to the data. Let $\{X(n)\}$ be the set of observations from the d-channel stationary time series where $X(n) = (x_1(n), x_2(n), \cdots, x_d(n))'$. Then for example, the instantaneous response-orthogonal

innovations MCAR model for $d = 3$ time series $X(n)$ is

$$\begin{bmatrix} 1 & 0 & 0 \\ \alpha_2(1) & 1 & 0 \\ \alpha_3(2) & \alpha_3(1) & 1 \end{bmatrix} \begin{bmatrix} y(1+3(t-1)) \\ y(2+3(t-1)) \\ y(3+3(t-1)) \end{bmatrix} + \begin{bmatrix} \alpha_1(3) & \alpha_1(2) & \alpha_1(1) \\ \alpha_2(4) & \alpha_2(3) & \alpha_2(2) \\ \alpha_3(5) & \alpha_3(4) & \alpha_3(3) \end{bmatrix} \begin{bmatrix} y(1+3(t-2)) \\ y(2+3(t-2)) \\ y(3+3(t-2)) \end{bmatrix} \tag{4.3}$$

$$+ \quad . \; . \; . \quad = \begin{bmatrix} \varepsilon(1+3(t-1)) \\ \varepsilon(2+3(t-1)) \\ \varepsilon(3+3(t-1)) \end{bmatrix}$$

In (4.3), the $d = 3$ multivariate time series, $\{Y(n)\}$ is a rearrangement of the original $d = 3$ time series, $\{X(n)\}$ according to the formula, $y(k + d\ (n-1)) = x_k(n)$, $k = 1$, $\cdots$, d (Pagano 1978).

Each of the d scalar time series $y(k + d\ (n-1))$ is modeled individually as a scalar autoregression in the form .

$$y(k + d(n-1)) = \sum_{m=1}^{p_k} \alpha_m y(k + d\ (n-1)-m) + \varepsilon(n) . \tag{4.4}$$

4.4 EXAMPLES

Results of Monte Carlo simulation studies of power spectrum estimation performance are shown in this section. The studies were designed to explore whether or not the performance of our SPMCAR modeling method was superior to the more conventional AIC-MCAR modeling method for power spectrum estimation for multichannel stationary time series. The experiments are conducted using the two theoretical models, *AR* (2) and *ARMA* (2,1), from which simulation data will be generated with different data length of $N = 100$ and $N = 200$. Using the *AR* and *ARMA* theoretical models the SPMCAR model performance was examined under two different modeling conditions. When fitting MCAR models to data generated from the *AR* (2) model, the fitted MCAR and SPMCAR models are in the same model class of the true model. Modeling simulated data from *ARMA* (2,1), the fitted MCAR and SPMCAR models are in a different model class and are only approximations to the data generating model.

The theoretical AR_2MA_1 considered was the bivariate time series model

$$\sum_{m=0}^{2} A_m x(n-m) = \sum_{m=0}^{1} B_m \varepsilon(n-m) , \quad E[\varepsilon(n)] = 0 , \quad COV[\varepsilon(n),\varepsilon(n')] = V\delta_{n,n'} \tag{4.5}$$

with parameters $A_0 = I$, $B_0 = I$ and

$$A_1 = \begin{bmatrix} 0.130 & 0.130 \\ -0.130 & -1.130 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.850 & 0.085 \\ -0.085 & 0.850 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0.600 & -0.300 \\ -0.300 & 0.600 \end{bmatrix}, \quad V = \begin{bmatrix} 1.000 & 0.500 \\ 0.500 & 1.250 \end{bmatrix}.$$

The AR_2 model considered was a variation of the model in (4.5) with the moving average parameter B_1 identically zero. The AR coefficient matrices are given below,

$$A_1 = \begin{bmatrix} 0.130 & 0.330 \\ -0.430 & -1.030 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.550 & 0.085 \\ -0.185 & 0.650 \end{bmatrix}.$$

The innovations variance matrix V for the AR_2 is the same as in (4.5) for the *ARMA* (2,1) model.

The simulated data from the *AR* and *ARMA* models are fitted by two different modeling methods, the Whittle-AIC MCAR method and the smoothness priors multichannel AR, (SPMCAR) modeling method. The resultant estimated multichannel autoregressive models were used to compute the estimated power spectrum density. A sum of the relative squares of the difference between the estimated model function and the true model function was computed as a goodness of fit criterion. Results were computed for the four functions, power spectral density in channels one and two, coherence squared and transfer function as functions of frequency.

For completeness a detailed worked SPMCAR example is shown for the case in which the data was simulated from the *ARMA* (2,1) model.

4.4.1 A Detailed Worked Example

For this example, N = 200 data points were generated using the bivariate *ARMA* (2,1) theoretical model. That data is illustrated in Figure 4.1. The conventional Yule-Walker AIC MCAR model was fitted to the simulated data. In this example, a maximum order 10 MCAR model was considered for the data. The values of the AIC computed for each MCAR order model fitted are shown in Table 4.1.

Table 4.1: MCAR Model Order Selection

MODEL ORDER SELECTION

Order	1	2	3	4	5
AIC	1386.635	1042.949	971.205	951.686	947.401
Order	6·	7	8	9*	10
AIC	941.209	940.880	944.824	939.630	944.929

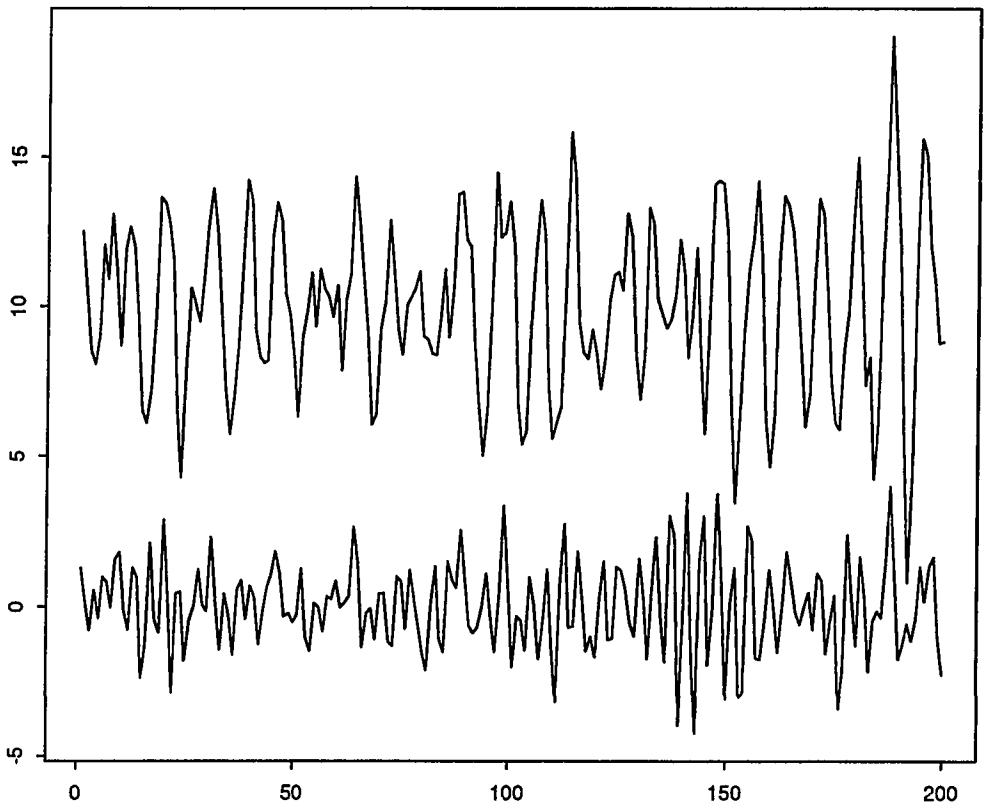

Figure 4.1: Bivariate ARMA(2,1) Simulation Data

The AIC best model selected for this data was AR_9. The AR coefficient matrices and innovations matrix for the AIC best model are

$$\hat{V} = \begin{bmatrix} 0.75767 & 0.43059 \\ 0.43059 & 1.17855 \end{bmatrix},$$

$$\hat{A}_1 = \begin{bmatrix} -0.42739 & -0.00918 \\ -0.46985 & -0.47093 \end{bmatrix}, \quad \hat{A}_2 = \begin{bmatrix} 0.94103 & -0.12186 \\ -0.75692 & 0.48918 \end{bmatrix}, \quad \hat{A}_3 = \begin{bmatrix} -0.74757 & 0.20585 \\ -0.13036 & 0.39469 \end{bmatrix},$$

$$\hat{A}_4 = \begin{bmatrix} 0.45101 & 0.01553 \\ -0.51996 & 0.43092 \end{bmatrix}, \quad \hat{A}_5 = \begin{bmatrix} -0.37082 & -0.00907 \\ -0.14085 & 0.09078 \end{bmatrix}, \quad \hat{A}_6 = \begin{bmatrix} 0.14038 & 0.07625 \\ -0.44093 & 0.26263 \end{bmatrix},$$

$$\hat{A}_7 = \begin{bmatrix} 0.09319 & -0.05220 \\ 0.18248 & 0.09573 \end{bmatrix}, \quad \hat{A}_8 = \begin{bmatrix} 0.07531 & -0.01302 \\ -0.08786 & -0.01498 \end{bmatrix}, \quad \hat{A}_9 = \begin{bmatrix} 0.15887 & -0.01493 \\ 0.25918 & 0.12210 \end{bmatrix},$$

We also used the SPMCAR modeling method to model this data. In that method the data is interpreted as an instantaneous response-orthogonal innovations matrix MCAR model. The resulting two scalar time series are modeled individually by the smoothness priors long *AR* method. For an order M MCAR model the resulting scalar AR model orders are respectively, $p_1 = M$ and $p_2 = M+1$. Two different SPMCAR models, with AR order $M = 10$ and order $M = 13$, and frequency domain smoothness priors constraints on the *AR* parameters were fitted to the data. The orders of the difference constraints considered are $k = 1, \cdots, 7$. As was described in Chapter 2, for each k, the hyperparameters ν^2 and λ^2 are searched over a discrete parameter space to minimize the log-likelihood of the hyperparameters in a least squares computation. Actually we compute -2 log-likelihood by the formula

$$-2logL(y|\lambda^2,\nu^2,\sigma^2) = N\log 2\pi\sigma^2 - \log|D^TD| + log|X^TX + D^TD| + N$$

where the quantities D and X respectively refer to the matrix of constraints and the design matrix in the smoothness priors long AR modeling. The results for the two scalar time series are tabulated below in Table 4.2 and Table 4.3.

Table 4.2: SPMCAR Order 10: Hyperparameters and log-likelihoods

Hyperparameters and likelihoods for channel 1

k	ν	λ	-2log-likelihood
1	1.69955	0.84977	527.22989
2	1.69955	0.10622	520.16009
3*	1.69955	0.00664	518.45444
4	3.39910	0.00005	532.81718
5	3.39910	0.00000	537.58772
6	3.39910	0.00000	537.84689
7	3.39910	0.00000	537.85288

Hyperparameters and likelihoods for channel 2

1	1.77547	0.44387	555.25417
2	1.77547	0.05548	548.98809
3*	1.77547	0.00347	547.36331
4	1.77547	0.00005	555.13799
5	1.77547	0.00000	567.40852
6	1.77547	0.00000	569.05522
7	1.77547	0.00000	569.10563

Table 4.3: SPMCAR Order 13: Hyperparameters and log-likelihoods

Hyperparameters and likelihoods for channel 1

k	ν	λ	-2log-likelihood
1	0.00000	0.81356	510.20666
2	1.62711	0.10169	501.07728
3*	1.62711	0.00635	499.11307
4	3.25422	0.00004	514.32195
5	3.25422	0.00000	525.16529
6	4.88133	0.00000	527.95041
7	4.88133	0.00000	528.01544

Hyperparameters and likelihoods for channel 2

1	1.72213	0.43053	531.26104
2	1.72213	0.05382	521.80137
3	1.72213	0.00673	520.04506
4	1.72213	0.00005	530.99815
5	1.72213	0.00000	553.01845
6	3.44425	0.00000	554.44225
7	5.16638	0.00000	562.04388

The two orthogonal smoothness priors long *AR* models are combined into the conventional MCAR form by the algebraic manipulations described in Chapter 2. The resulting AR matrices and the innovations matrix computed for the SPMCAR order 10 model are,

$$\hat{V} = \begin{bmatrix} 1.19884 & 0.50481 \\ 0.50481 & 1.31951 \end{bmatrix}, \quad \hat{A}_1 = \begin{bmatrix} -0.24777 & -0.1316 \\ -0.4158 & -0.64669 \end{bmatrix},$$

$$\hat{A}_2 = \begin{bmatrix} 0.77718 & 0.11137 \\ -0.56408 & 0.49594 \end{bmatrix}, \quad \hat{A}_3 = \begin{bmatrix} -0.40183 & 0.08403 \\ 0.11742 & 0.30497 \end{bmatrix}, \quad \hat{A}_4 = \begin{bmatrix} 0.07921 & 0.01221 \\ -0.09411 & 0.12415 \end{bmatrix},$$

$$\hat{A}_5 = \begin{bmatrix} -0.0554 & 0.09188 \\ 0.02538 & 0.07793 \end{bmatrix}, \quad \hat{A}_6 = \begin{bmatrix} -0.00167 & 0.04146 \\ 0.005 & 0.08316 \end{bmatrix}, \quad \hat{A}_7 = \begin{bmatrix} 0.01078 & 0.00419 \\ 0.00722 & 0.0483 \end{bmatrix},$$

$$\hat{A}_8 = \begin{bmatrix} 0.00062 & -0.00151 \\ 0.00840 & 0.00812 \end{bmatrix}, \quad \hat{A}_9 = \begin{bmatrix} -0.00376 & -0.00069 \\ 0.00459 & 0.01596 \end{bmatrix}, \quad \hat{A}_{10} = \begin{bmatrix} 0.00049 & -0.00012 \\ -0.00066 & 0.00648 \end{bmatrix}.$$

The resulting AR matrices and the innovations matrix computed for the SPMCAR order 13 model are,

$$\hat{V} = \begin{bmatrix} 1.21738 & 0.48592 \\ 0.48592 & 1.2946 \end{bmatrix}, \quad \hat{A}_1 = \begin{bmatrix} -0.24068 & -0.12288 \\ -0.40626 & -0.62747 \end{bmatrix},$$

$$\hat{A}_2 = \begin{bmatrix} 0.76392 & 0.09901 \\ -0.57652 & 0.4813 \end{bmatrix}, \quad \hat{A}_3 = \begin{bmatrix} -0.39227 & 0.08647 \\ 0.13549 & 0.31147 \end{bmatrix}, \quad \hat{A}_4 = \begin{bmatrix} 0.08111 & 0.01145 \\ -0.09447 & 0.12829 \end{bmatrix},$$

$$\hat{A}_5 = \begin{bmatrix} -0.05679 & 0.09283 \\ 0.02678 & 0.08269 \end{bmatrix}, \quad \hat{A}_6 = \begin{bmatrix} -0.00087 & 0.04249 \\ 0.00399 & 0.08718 \end{bmatrix}, \quad \hat{A}_7 = \begin{bmatrix} 0.00939 & 0.00335 \\ 0.00359 & 0.05162 \end{bmatrix},$$

$$\hat{A}_8 = \begin{bmatrix} 0.00075 & -0.00267 \\ 0.01113 & 0.00648 \end{bmatrix}, \quad \hat{A}_9 = \begin{bmatrix} -0.00325 & -0.0013 \\ 0.00613 & 0.01486 \end{bmatrix}, \quad \hat{A}_{10} = \begin{bmatrix} 0.00014 & 0.00024 \\ -0.00264 & 0.008580 \end{bmatrix},$$

$$\hat{A}_{11} = \begin{bmatrix} -0.00018 & 0.00065 \\ -0.00065 & 0.00002 \end{bmatrix}, \quad \hat{A}_{12} = \begin{bmatrix} -0.00047 & -0.00023 \\ 0.00105 & -0.00154 \end{bmatrix}, \quad \hat{A}_{13} = \begin{bmatrix} 0.00017 & -0.00071 \\ -0.00037 & -0.00205 \end{bmatrix}.$$

As expected, examination of the values in the *AR* coefficient matrices reveals that the effect of the priors is to produce successively higher order coefficient matrices with diminishing entries.

The power spectral density function $S(f)$ is related to the *AR* coefficient matrices by

$$S(f) = A^{-1}(f)VA^{*-T}(f), \quad -1/2 \le f < 1/2$$

where $A(f) = \sum_{k=0}^{M} A_k \exp(-2\pi i f k)$ is the whitening filter of the AR modeled time series. The power spectral density of the two channel time series are estimated with the AR matrices obtained from the following three different models of the sample data: The Yule-Walker-AIC, SPMCAR 10 and SPMCAR 13 models. The estimates of the power spectrum density for channel 1 and channel 2, coherence, and the transfer function are shown in Figure 4.2 and Figure 4.3, respectively. The first row of four plots in

Figure 4.2 and Figure 4.3 are the estimates obtained using the conventional Yule-Walker-AIC MCAR model. The solid line in each plot is the theoretical value and the dotted line is the corresponding estimated value. The second row of four plots in the figures are the estimates obtained using the SPMCAR models. The general appearance of the graphs are similar for the estimates obtained by the SPMCAR and Yule-Walker-AIC models. However, the estimates obtained from the two SPMCAR models, are almost identical in appearance.

The goodness of the estimates to the theoretical functions is measured with the sum of relative squared errors between the true and the estimated value, $\sum_{i=1}^{N}[(y(i)-\hat{y}(i))/y(i)]^2$, where $y(i)$ is the theoretical value and $\hat{y}(i)$ is an estimate of $y(i)$. The sum of relative squared errors are tabulated in Table 4.4.

Table 4.4: PSD Estimation Performance

	Sum of Relative Squared Errors			
	PSD CHN 1	PSD CHN 2	Coherence	Transfer Func
Yule-Walker	54.773	14.924	30.021	18.426
SPMCAR 10	29.293	9.204	15.910	9.069
SPMCAR 13	32.279	8.946	18.001	10.069

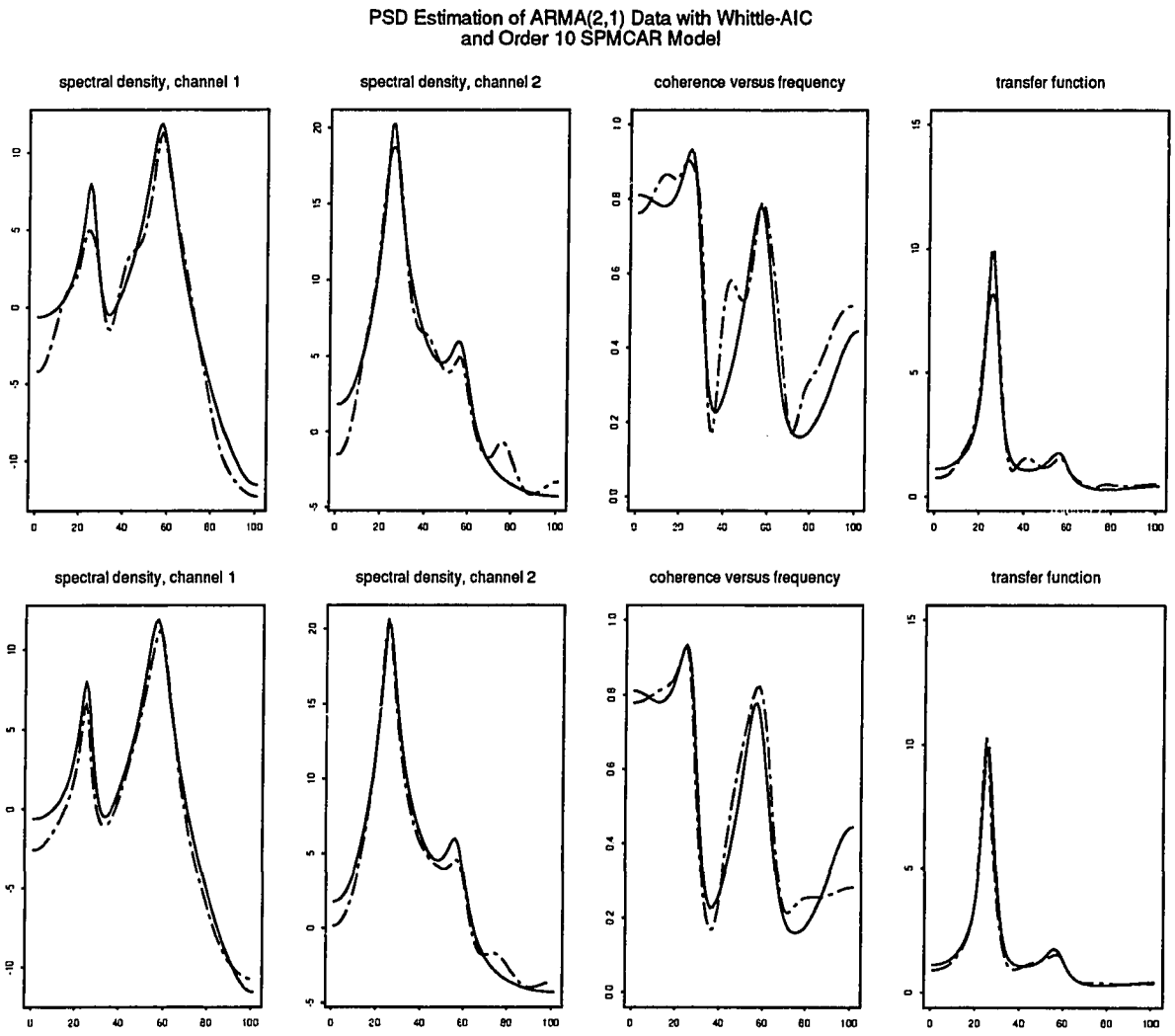

Figure 4.2: PSD Estimation With Order 10 SPMCAR Model

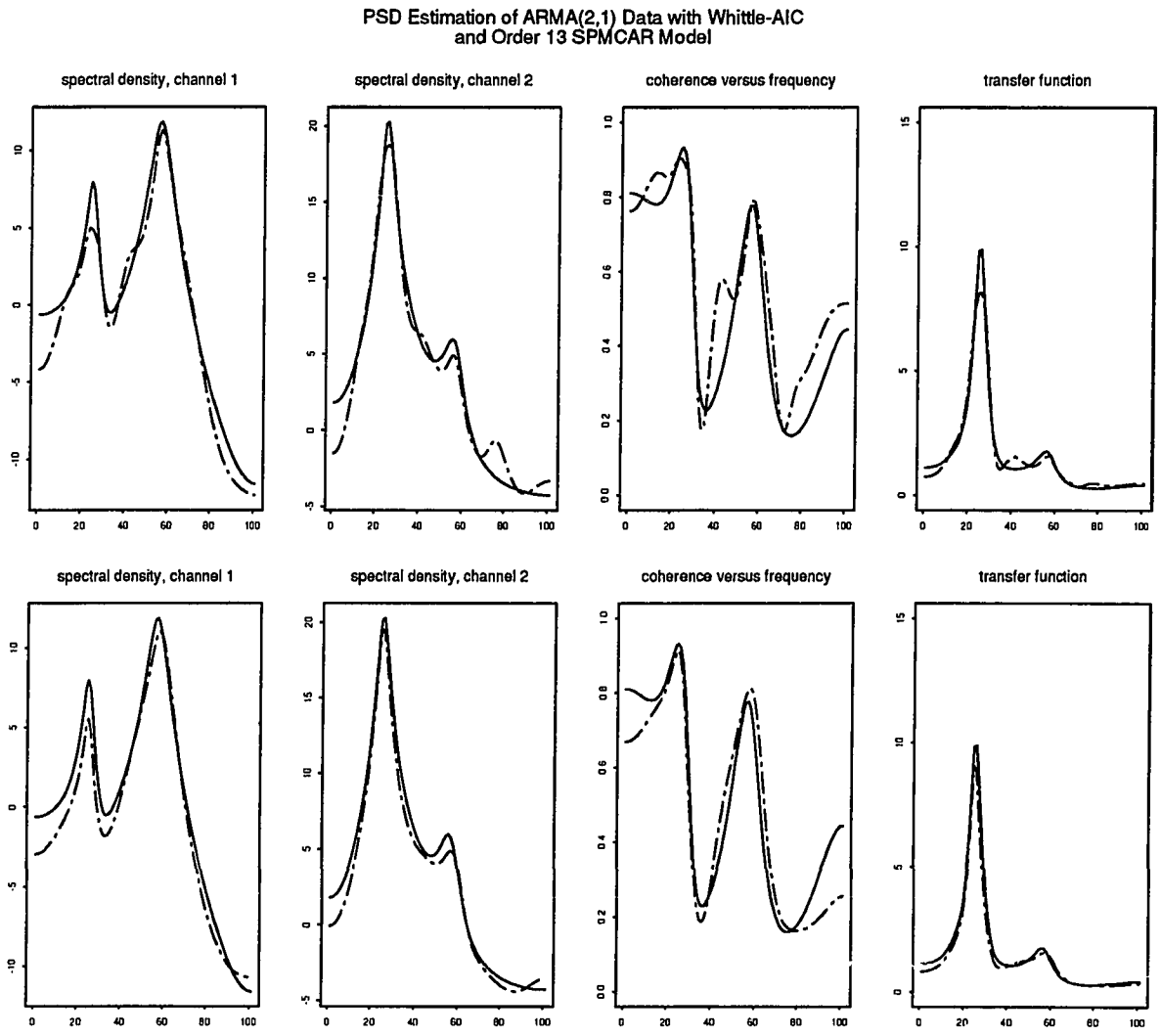

Figure 4.3: PSD Estimation With Order 13 SPMCAR Model

4.4.2 AR Model Simulation Results

In this simulation study the bivariate AR model of order 2 is used as the theoretical model from which simulation data are generated. Data from 30 (thirty) simulations were computed with $N = 100$. The results obtained using the bivariate AR_2 simulation model are:

Table 4.5: AR_2 MCAR Model:
Comparative spectral function estimate performances:
Whittle-AIC MCAR and SPMCAR modeling.

Simulation Results: Relative MSE, 30 Trials			
	Whittle	SP Order 3	SP Order 4
Channel 1 Mean	10.160	9.371	10.905
S.D.	6.062	5.091	5.636
Channel 2 Mean	12.087	9.736	11.881
S.D.	10.187	6.424	7.614
Coherence Mean	1130.324	672.272	836.100
S.D.	2452.186	818.449	1005.377
T.funct. Mean	77.574	63.809	72.950
S.D.	97.945	49.330	56.380

The results of our experiments suggest that the SPMCAR method of power spectrum estimation is superior to the Whittle-AIC method in a mean square sense and that it is more stable in the sense that the estimated function variances are smaller. Furthermore and quite important, the SPMCAR method is relatively insensitive to SPMCAR model order and it can realize good model estimation results with relatively short duration data sets.

Figure 4.4 shows, respectively the theoretical spectral densities, coherence and transfer function (dark lines) and the empirical mean and plus and minus one standard deviation (dotted lines) for the Whittle-AIC and SPMCAR modeled data. The figure shows the results for the Whittle-AIC MCAR model and the SPMCAR model order 3. The graphical results for SPMCAR model orders 3 and 4 were visually indistinguishable.

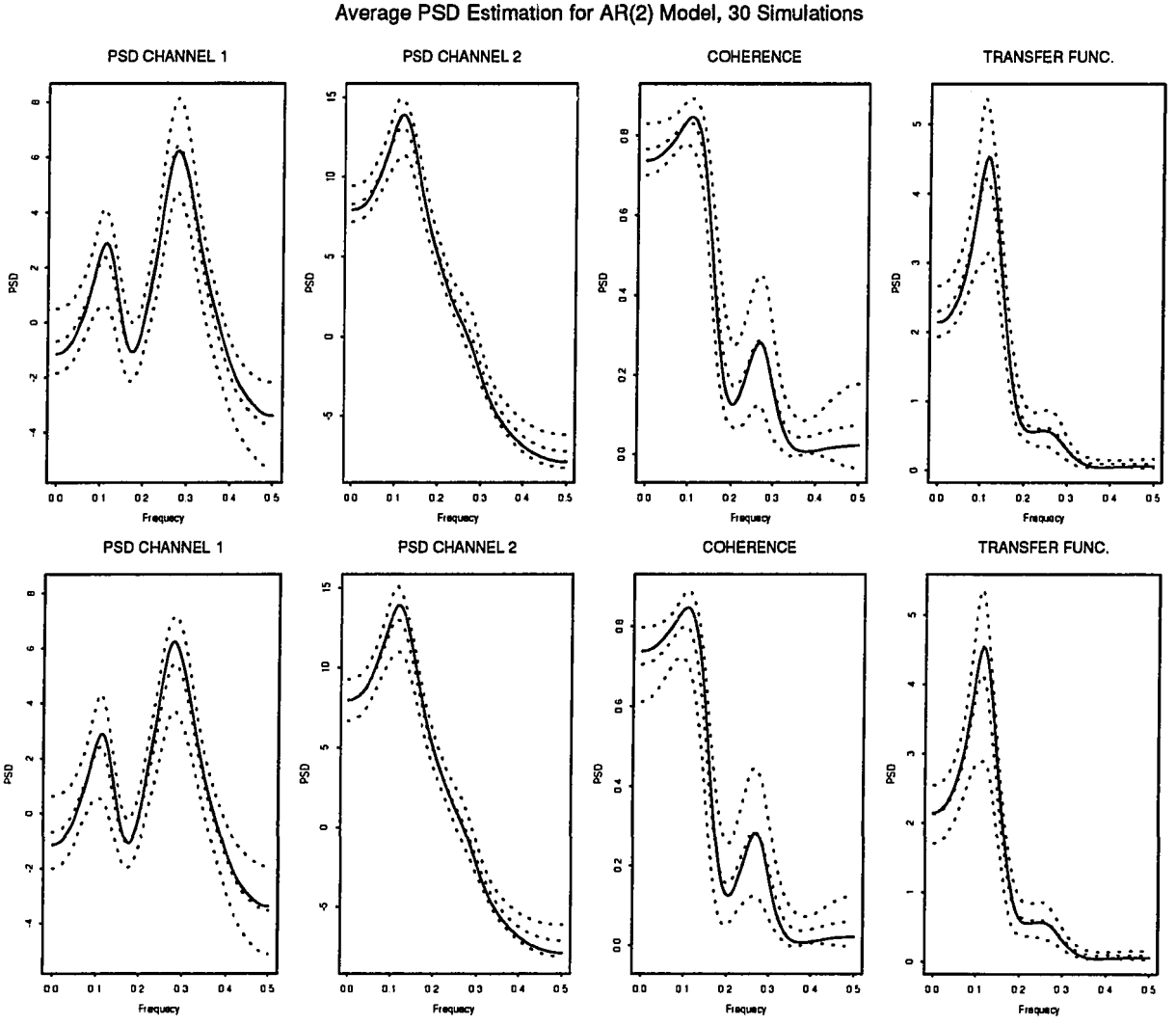

Figure 4.4: PSD Estimation For AR_2 Model

4.4.3 ARMA Model Simulation Results

In this simulation study the bivariate AR_2MA_1 model is used as the theoretical model from which simulation data are generated. Data from 30 (thirty) simulations was computed with $N = 200$. The results obtained using the bivariate AR_2MA_1 simulation model are:

Table 4.6: AR_2MA_1 MCAR Model:
Comparative spectral function estimate performances:
Whittle-AIC MCAR and SPMCAR modeling.

Simulation Results: Relative MSE, 100 Trials			
	Whittle	SP Order 10	SP Order 13
Channel 1 Mean	37.768	18.369	21.060
S.D.	5340.213	8.581	109.410
Channel 2 Mean	28.935	7.732	8.102
S.D.	2543.022	8.581	14.830
Coherence Mean	14.396	14.168	14.396
S.D.	74.773	39.074	43.410
T.funct. Mean	10.735	7.556	7.814
S.D.	85.775	9.613	10.310

Figure 4.5 shows, respectively the theoretical spectral densities, coherence and transfer function (dark lines) and the empirical mean and plus and minus one standard deviation (dotted lines) for the Whittle-AIC and SPMCAR modeled data. Figure 4.5 corresponding to the simulated AR_2MA_1 model shows the results for SPMCAR model order 13. The graphical results for SPMCAR model orders 10 and 13 were visually indistinguishable.

The results of our experiments suggest that the SPMCAR method of power spectrum estimation is superior to the Whittle-AIC method in a mean square sense and that it is more stable in the sense that the estimated function variances are smaller. Furthermore and quite important, the SPMCAR method is relatively insensitive to SPMCAR model order and it can realize good model estimation results with relatively short duration data sets.

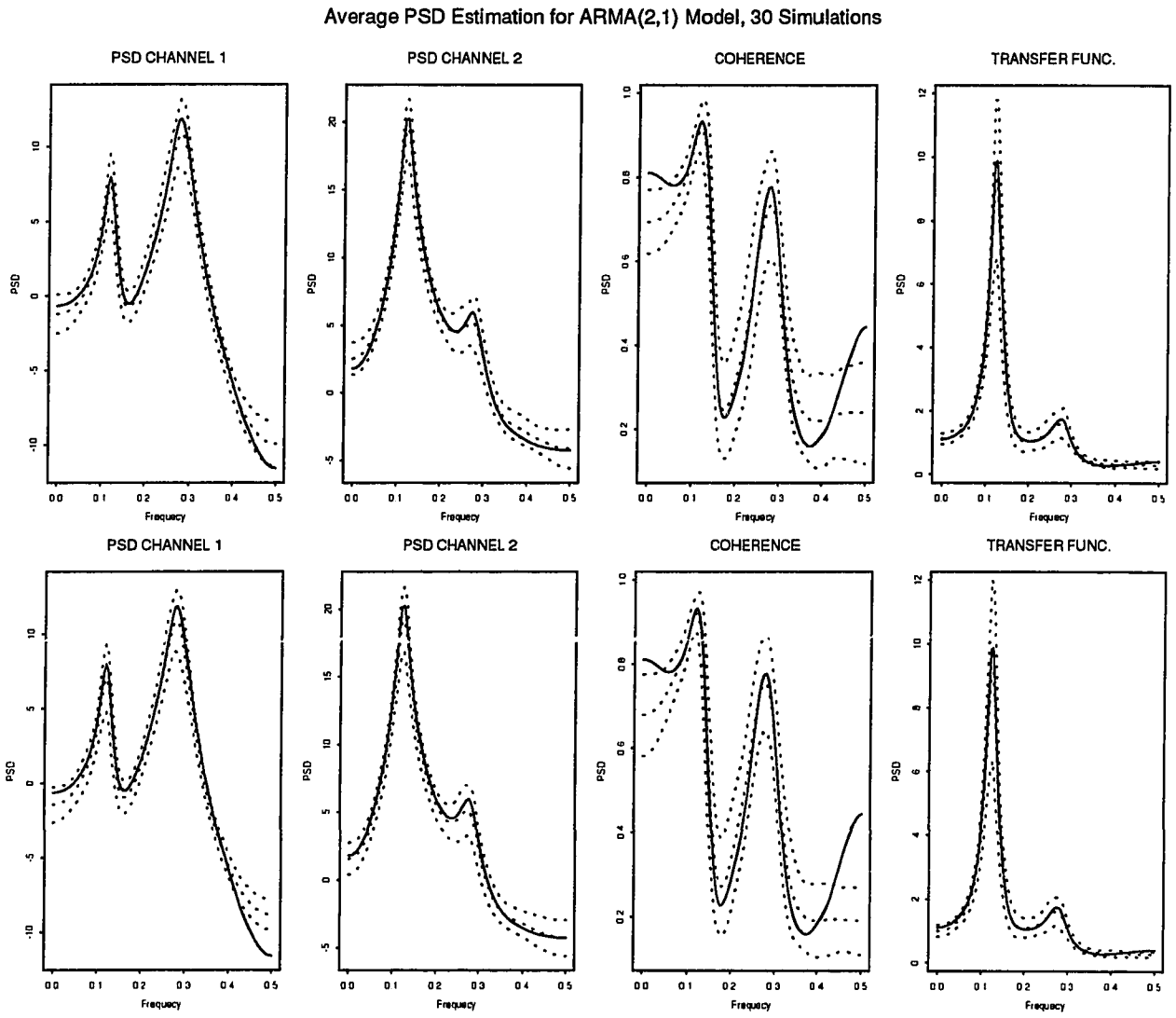

Figure 4.5: PSD Estimation for *ARMA* (2,1) Model

CHAPTER 5

SPMCTVAR MODELING

In this chapter the one channel at-a-time and smoothness priors paradigms are exploited to achieve the modeling of multivariate nonstationary covariance time series using multivariate time varying autoregressive models. Modeling of multichannel nonstationary covariance time series is relevant to many econometric, human electroencephalogram, oceanographic, meteorological and seismic data analysis situations.

In sections 5.1 and 5.2 some additional historical and theoretical background materials are provided. The development is outlined in Section 5.3, and examples are shown in Section 5.4.

5.1 BACKGROUND

An earlier and in fact still used approach to the modeling of nonstationary covariance time series is to segment the series into approximately stationary time series segments and to model each segment as a stationary time series model. The problem with this approach is that it does not adequately capture the local and global statistical properties of the time series. On the basis of success with previous use of multivariate autoregressive models of stationary time series, Gersch and Kitagawa (1983) considered the fitting of a multivariate time varying autoregressive model,

$$x(n) = \sum_{m=1}^{P} A_{n,m}\, x(n-m) + \varepsilon(n), \quad E[\varepsilon(n)] = 0, \quad COV[\varepsilon(n),\varepsilon(n+k)] = V(n)\delta_{n,n+k},$$

to nonstationary covariance economic data. That method was subsequently applied to the modeling of nonstationary covariance human epileptic episode electroencephalograms, (Gersch 1985, 1987).

In the model above for a d channel N observation time series there are $N \times d$ observations and $d \times d \times P \times N$ autoregressive model parameters and $d \times (d+1) \times N/2$ innovations matrix parameters. In order to fit such a model with so many more parameters than observations it is necessary to efficiently parametrize the time evolution of the parameters. The method used by Gersch and Kitagawa was to characterize the evolution of each of the $d \times d \times P$ autoregressive parameters in terms of a discrete Laguerre polynomial in a least squares computational solution. The order of the multivariate AR model

and the degree of the Laguerre polynomial were determined using Akaike's AIC statistic. That approach appeared to be reasonable and the results obtained in the aforementioned papers also seemed to be quite reasonable. On the other hand in that procedure, the Laguerre polynomial coefficients were determined by global statistical considerations. Ultimately it became clear that it would be more desirable to develop a multivariate time varying AR parameter modeling procedure that was more sensitive to local time variations. The circular lattice filter for the fitting of multivariate stationary autoregressive time series models to data using sample covariances, (Sakai 1982), was very suggestive of such a procedure. In fact our smoothness priors multichannel time varying autoregressive, SPMCTVAR, modeling method is an adaptation of Sakai's modeling procedure. Two innovative developments that permit our extension of Sakai's method to the fitting of multichannel autoregressive time varying models to data are the modeling on instantaneous data and the use of a stochastic regression (smoothness priors) modeling of the partial autocorrelations, PARCORS, of individual scalar autoregressive time series. In our algorithm, the $d \times d \times P \times N$ autoregressive model parameters and $d \times (d + 1) \times N/2$ innovations matrix parameters. parameters are expressed implicitly in terms of only $2 \times d$ hyperparameters that are computed explicitly.

5.2 BACKGROUND: TOWARDS THE SPMCTVAR

Whittle (1963) was a singular development in the fitting of multichannel autoregressive (AR) models to stationary time series data. In extending the Levinson algorithm for fitting AR models to scalar stationary time series to the multichannel case, Whittle found it necessary to employ both forward and backward multichannel AR models. The latter was critical in the implementation of a multichannel lattice-form algorithm, Wiggins and Robinson (1965). That realization and subsequent multichannel realizations by Lee et al. (1981), Friedlander (1982), and Porat et al. (1982) involved matrix computations. Scalar implementations of multichannel recursions, which required no matrix processing, were realized by Sakai (1982), Lev-Ari (1983, 1987), Lev-Ari and Kailath (1984), and Ling and Proakis (1984). Sakai's implementation transforms multichannel AR models into scalar AR models, (motivated by Pagano 1978), by converting the multichannel Levinson algorithm into a scalar equivalent which Sakai refers to as the "circular lattice filter". Lev-Ari's realization emphasizes the regularity or modularity of the architecture. His 1987 presentation is the most elegant of the multichannel lattice

algorithm developments. (The structure of Lev-Ari's algorithm and Sakai's algorithm are equivalent.) Sakai's algorithm is based on sample correlation function computations.

The Sakai and Lev-Ari multichannel lattices are in fact one channel at-a-time structures. However Sakai only deals with stationary time series modeling and Lev-Ari concern was with a conceptual theoretical structure, not with implementation. The development of our own SPMCTVAR algorithm was directly influenced by Sakai's paper and his use of the Levinson algorithm type lattice structure.. The variations in our work are that our algorithm is computed on the instantaneous data whereas Sakai, being concerned only with stationary time series computed with covariances. Another conceptual difference is that we think of the partial correlation coefficients, (PARCORS) in a stochastic regression model context. In fact, we exploit a random walk trend model in the regression equation for the PARCORS. Instantaneous PARCORS are computed for each channel which in turn yields an instantaneous multichannel AR model. The use of a stochastic regression model for the PARCORS follows from the concept of smoothness priors time series modeling. The smoothness priors specify a prior distribution on the time series model parameters in the form of stochastic difference equation constraints. The estimate of the time series model parameters, in this case the estimate of the instantaneous PARCORS is the mean of the posterior distribution of the PARCORS. Our algorithm for computing the instantaneous PARCORS is a version of that in Kitagawa (1988), used in the scalar AR modeling of non-Gaussian nonstationary covariance time series.

5.2.1 Instantaneous Spectral Density

Primarily the applications we are concerned with here, in the fitting of time varying autoregressive models to data, involves the estimation of an instantaneous power spectrum. That particular term is due to Page (1952) in a pionering paper. Page's concept was that the instantaneous spectrum, roughly speaking, is the difference between the power distributions of the process over the interval (0,t) and over the interval (0,t + dt). An alternative concept introduced in Priestley (1965) and reviewed and summarized in Priestley (1981), is that of the evolutionary spectrum. The evolutionary spectrum represents the power distribution of the process within the interval (t,t + dt). That is, the evolutionary spectrum corresponds roughly to what we would measure if the portion of time series lying between t

and $t + dt$ were passed through a wave analyzer. Priestley's estimation of the evolutionary spectrum is done via an adaptation of windowed periodogram analysis. The analysis are only applicable to very slowly varying spectra and are not of great interest today. Another approach to the estimation of an instantaneous spectrum is the Wigner-Wille spectrum, (introduced in Ville 1930, recent theoretical work and references are in Flandrin and Martin 1989). At any particular time instant, that spectrum is defined as the Fourier transform of the covariance function over a small interval time interval around that particular time instant. The Wigner-Ville spectrum does not have a meaningful physical interpretation and may in fact take on negative values for certain processes. Our own approach is to define an instantaneous spectrum operationally. That is, the power spectral density of a scalar process modeled by an autoregressive model may be computed via a formula that involves the AR model. In a natural way we extend that definition to the estimation of an instantaneous spectrum in terms of the time varying autoregressive model. That approach, while not mathematically formal appears to yield meaningful computational results. Examples were shown for the scalar case in Kitagawa and Gersch (1985b) in the modeling of scalar nonstationary covariance time series and in Gersch and Kitagawa (1983) in the modeling of multivariate nonstationary covariance time series. Our concept of instantaneous spectrum is similar to the evolutionary spectrum concept used by Priestley.

The instantaneous power spectral density matrix is the distribution of energy in the time series as a function of frequency at an instant in time. The instantaneous power spectral density matrix is a generalization of the more familiar (time independent) power spectral density matrix that is used in the analysis of stationary time series. The more conventional analysis of stationary time series by parametric or Fourier transform methods, do not have the time-frequency resolution properties necessary to capture the transitory characteristics of relatively rapidly changing nonstationary covariance data.

Recall that the multivariate time varying autoregressive model of order M may be written as, $x(n) = A(1,n)x(n-1) + A(2,n)x(n-2) + ...+ A(M,n)x(n-M) + \varepsilon(n)$. The $x(n)$ denotes the observations of a d-vector time series at the time points $n=1,2,\ldots,N$. The $\{A(m,n), m=1,\ldots,M, n=1,\ldots,N\}$ denote the $d\,x\,d\,x\,M\,x\,N$ coefficient matrices of the model, and $\{\varepsilon(n)\ n=1,\ldots,N\}$ is a zero mean, uncorrelated in time, d-channel time series with instantaneous symmetric $d\,x\,d$ covariance matrix $V(n)$. The instantaneous power spectral density matrix is computed

from the fitted SPMCAR model by the formula, $S(f,n) = A(f,n)^{-1}V(n)A^{*}(f,n)^{-T}$. where $A(f,n) = I - \sum_{m=1}^{M} A(m,n)\exp[-2\pi imf]$ is a dxd polynomial operator, the symbol * denotes the complex conjugate and the instantaneous power spectral density matrix $S(f,n)$ is a dxd matrix defined for $-1/2 \le f \le 1/2$ and $n=1,\ldots,N$.

5.3 DEVELOPMENT

The multichannel time varying AR coefficient model for modeling the d-channel N duration data, $\{x(1),\ldots,x(N)\}$, is in the form

$$x(n) = \sum_{m=1}^{M} A_{n,m}x(n-m) + \varepsilon(n). \tag{5.1}$$

In (5.1), the $x(n)$'s are d-dimensional vectors and the $\{\varepsilon(n), n = 1,\ldots,N\}$ is a zero-mean uncorrelated normally distributed d-dimensional vector sequence with instantaneous covariance matrix, V.

In the practical analysis of data, we are usually interested in the evolution with time of transfer functions, coherences etc. between channels as a function of frequency. Such quantities can be readily computed from the evolutionary power spectral density matrix. That is given by, (Gersch and Kitagawa 1983, Gersch 1985, Gersch 1988),

$$S(f,n) = A(f,n)^{-1}V(n)A^{*}(f,n)^{-T}. \tag{5.2}$$

In (5.2), the multichannel AR model instantaneous system function $A(f,n)$ is given by,

$$A(f,n) = I - \sum_{m=1}^{M} A_{n,m}\exp[-2\pi imf]. \tag{5.3}$$

In order to simplify the presentation, much of the development is shown in terms of multichannel stationary AR time series. Correspondingly, the presentation follows substantially from Sakai (1982). Nonstationarity is re-invoked in computing the instantaneous partial correlation coefficients.

5.2.1 The Pagano-Sakai Algorithm

Consider the stationary multichannel AR time series model

$$x(n) = \sum_{m=1}^{M} A_m x(n-m) + w(n), \quad n = 1,\ldots,N \tag{5.4}$$

where $\{w(n),\ \ n = 1, \ldots, N\}$ is an zero-mean time uncorrelated d-vector sequence with covariance matrix W.

The "trick" of doing multivariate AR things one channel at-a-time follows from Pagano (1978). Interlace the d-channel time series $x(t)$ into a scalar series $y(t)$ using, $x_j(t) = y(j + d(t-1)),\ j = 1, \ldots, d$. Then, express the $\{y(t)\}$ sequence into a d-vector AR model according to $y(k + d(t-1)) + \sum_{j=1}^{p_k} \alpha_k(j)y(k + d(t-1) - j) = \tilde{\varepsilon}(k + d(t-1))\ \ k = 1,...,d.$ with $\tilde{\varepsilon}(k + d(t-1)) \sim N(0,\sigma_k^2)$ i.i.d. with p_k the order of the k-th channel AR model. An example of this procedure of interlacing the data of original time series into a multivariate instantaneous response-orthogonal innovations autoregressive model is shown in Chapter 2.

Consider the individual autoregressive channels. The key idea of the Levinson algorithm is next invoked to obtain a recursive order updating algorithm in terms of a reflection coefficient parameterization, (Morf et al. 1978a,b). Toward that end, define the j-th order k-th channel forward and backward prediction errors,

$$\varepsilon(j,k + nd) = y(k + nd) + \sum_{i=1}^{j} \alpha_k(j,i)y(k + nd - i) \tag{5.5}$$

$$\eta(j,k + nd) = y(k + nd - j) + \sum_{i=1}^{j} \beta_k(j,j + 1 - i)y(k + nd - i + 1)$$

for $k = 1, \ldots, d$ and integer n respectively in terms of the forward model $\{\alpha_k(j,1), \ldots, \alpha_k(j,j)\}$ parameters where k is the channel number and j is the AR model order and backward model parameters $\{\beta_k(j,1), \ldots, \beta_k(j,j)\}$.

Invoking the recursive order update relations yields

$$\varepsilon(j + 1,k + nd) = \varepsilon(j,k + nd) + \alpha_k(j+1,j+1)\eta(j,k-1+nd) \tag{5.6}$$

$$\eta(j+1,k+nd) = \eta(j,k-1+nd) + \beta_k(j+1,j+1)\varepsilon(j,k+nd)$$

with

$$\sigma_{k(j)}^2 = E[\varepsilon^2(j,k+nd)], \quad \tau_{k(j)}^2 = E[\eta^2(j,k+nd)]. \tag{5.7}$$

The parameter update equations then become

$$\alpha_k(j+1,i) = \alpha_k(j,i) + \alpha_k(j+1,j+1)\beta_{k-1}(j,j+1-i) \tag{5.8}$$

$$\beta_k(j+1,i) = \beta_{k-1}(j,i) + \beta_k(j+1,j+1)\alpha_k(j,j+1-i)$$

$$\sigma_k(j+1) = \sigma_k^2(j)[1 - \alpha_k(j+1,j+1)\beta_k(j+1,j+1)]$$

$$\tau_k(j+1) = \tau_{k-1}^2(j)[1 - \alpha_k(j+1,j+1)\beta_k(j+1,j+1)]$$

$$\sigma_k^2(0) = E[\varepsilon(0,k+nd]^2, \quad \tau_k^2(0) = E[\eta(0,k+nd]^2,$$

with the following constraint

$$0 \le \alpha_k(j+1,j+1)\beta_k(j+1,j+1) < 1.$$

In Sakai's algorithm, the zero-th order innovations variances, $\sigma_k^2(0)$ and $\tau_k^2(0)$ are specified by zero-lag covariances of corresponding k-th channel data variances. Also the order updated partial correlation coefficients are computed as linear combinations of sample channel lagged covariances. Our algorithm works on the instantaneous data.

5.2.2 Instantaneous PARCORS

In our algorithm, the k th channel zero-th order forward and backward innovations are

$$\varepsilon(0,k + nd) = y(k + nd), \quad \eta(0,k + nd) = y(k + nd). \tag{5.9}$$

From (5.6), imposing the nonstationarity of the PARCORS, the relations between the order update innovations and the instantaneous PARCORS are given by

$$\varepsilon(j + 1,k + nd) = \varepsilon(j,k + nd) + \alpha_{n,k}(j+1,j+1)\eta(j,k-1+nd) \tag{5.10}$$

$$\nabla\alpha_{n,k} = v_{n,k} \quad v_{n,k} \sim N(0,\tau_1^2)$$

$$\eta(j+1,k+nd) = \eta(j,k-1+nd) + \beta_{n,k}(j+1,j+1)\varepsilon(j,k+nd)$$

$$\nabla\beta_{n,k} = u_{n,k} \quad u_{n,k} \sim N(0,\tau_2^2)$$

The instantaneous updated forward and backward model parameters become,

$$\alpha_{n,k}(j+1,i) = \alpha_{n,k}(j,i) + \alpha_{n,k}(j+1,j+1)\beta_{n,k-1}(j,j+1-i) \tag{5.11}$$

$$\beta_{n,k}(j+1,i) = \beta_{n,k-1}(j,i) + \beta_{n,k}(j+1,j+1)\alpha_{n,k}(j,j+1-i)$$

The order updated instantaneous PARCORS, $\alpha_{n,k}(j+1,j+1)$, $\beta_{n,k}(j+1,j+1)$ in (5.10) are scalar regression coefficients and the updated forward and backward innovations are the residuals of the regressions. The key innovation in (5.10) is the stochastic difference equation representation for the order updated instantaneous PARCORS.

To elaborate on this point, consider the stochastic time-varying linear regression model:

$$y_n = \phi_n^T \theta_n + \eta_n, \quad n = 1,...,N \tag{5.12}$$

where y_n, η_n, ϕ_n, and θ_n are respectively, the observation and the (unobserved) observation noise, a d-dimensional stochastic regression vector and an unknown time-varying parameter. For example, consider the parameter variation at time n to be governed by the stochastic difference equation,

$$\nabla^k \theta_n = u_n \tag{5.13}$$

where $\nabla\theta_n = \theta_n - \theta_{n-1}$, $\nabla^2\theta = \nabla(\nabla\theta)$ etc. and $\{\eta_n\}$ and $\{u_n\}$ are normally distributed uncorrelated sequences with variances σ^2 and τ^2 respectively. For example, if $\phi_n = 1$ and $\nabla^k\theta_n = u_n$, (i.e. ϕ_n and θ_n are scalars), then $\{y_n, \ n = 1,...,N\}$ is a nonstationary mean time series. Let a general model be $y_n = \sum_{i=1}^{p} \phi_{i,n}\theta_{i,n} + \eta_n, \quad n = 1,...,N$ and $\nabla^k\theta_{i,n} = u_{i,n}$, with the $\phi_{i,n} = 1$. Then $y_n = \theta_{1,n} + \theta_{2,n} + \ldots \theta_{p,n} + \eta_n$, is a nonstationary mean time series-variance component model that has proven to be useful for econometric time series modeling, (Gersch and Kitagawa 1983, Kitagawa and Gersch 1984). Alternatively, the model $y_n = \sum_{i=1}^{M} \theta_{i,n} y_{n-i} + \eta_n$, with $\nabla^k\theta_{i,n} = u_{i,n}$, (i.e. the time varying AR coefficient model, with stochastically varying AR coefficients), is useful for modeling scalar nonstationary covariance time series. Note that in both the variance component and time varying AR coefficient model, the number of parameters exceeds the number of data. The key idea in fitting such models is to express the parameters implicitly in a prior distribution as functions of a hyperparameter. Only the likelihood for the hyperparameter needs to be computed explicitly.

Returning to (5.12) and (5.13), in the model $y_n = \theta_n + \eta_n$, $\nabla\theta_n = u_n$, $\eta_n \sim N(0,\sigma^2)$, $u_n \sim N(0,\tau^2)$, the normal distribution for the $\{\theta_n\}$ sequence is the prior and τ^2 is the hyperparameter of the prior distribution. The Kalman filter algorithm-likelihood computations for the hyperparameter τ^2 and the implicit estimation of the time-varying parameter sequence $\{\theta_n, \ n = 1,...,N\}$ are shown in Chapter 2. Kitagawa (1988) shows a variation of the time-varying stochastic linear regression model for the computation of PARCORS for a scalar time varying AR coefficient model. A related approach is adapted for the algorithm shown here.

In (5.10), the innovations variances τ_1^2 and τ_2^2 are hyperparameters of a smoothness priors regression-trend model for the forward and backward PARCORS. In our algorithm, they are computed via a maximized likelihood of the Bayesian-smoothness priors model. We compute the sum of squares of forward prediction errors of each channel, one channel at-a-time, for each model order for use with the AIC_C statistic in order to determine the AIC_C best order PARCOR model. Following that model order selection, the instantaneous multichannel AR model parameters are computed from the instantaneous PARCORS.

5.4 EXAMPLES

In this section the SPMCTVAR modeling of $d = 2$ and $d = 3$ simulated nonstationary covariance time series and the SPMCAR modeling of real physical $d = 2$ and $d = 3$ data are shown. The simulations are from nonstationary time varying autoregressive models. The examples demonstrate that our SPMCTVAR modeling procedure does capture both the local and the global statistical properties of known multichannel nonstationary covariance time series. The application of the SPMCTVAR modeling to an analysis of a human EEG for an epileptic episode is also shown.

5.4.1 SPMCTVAR d = 2 Examples

In this section we consider the modeling of a bivariate nonstationary covariance time series with SPMCAR models. First we show some details in the fitting of a simulated bivariate quasi-periodic time varying AR model. Secondly we show the application of the SPMCTVAR algorithm to the modeling of a real physical data example.

5.4.1.1 Fitting the SPMCTVAR Model to Simulated Bivariate Data

We choose a time varying $d = 2, AR_2$ quasi-periodic time series model,

$$x(n) + \begin{bmatrix} 1.73 - 1.05*sin(\pi*n/200) & 0.2 \\ 0.2 & -1.55 + 1.25*sin(\pi*n/200) \end{bmatrix} x(n-1)$$

$$+\begin{bmatrix} 0.85 & 0.2 \\ 0.2 & 0.85 \end{bmatrix} x(n-2) = \varepsilon(n),$$

$n = 200$ data points were simulated using this model with $\varepsilon(n)$, $n = 1, \ldots 200$ independent normal $d = 2$ vectors with identity covariance matrix. The simulated time series is shown in Figure 5.1. Also shown in Figure 5.1 are the evolution of the theoretical and fitted instantaneous power spectral density in each channel, and the coherence between the channels for every 10 th data point.

Following the development in Section 5.3, the simulation model corresponds to AR orders 4 and 5 for channels one and two respectively. First, the two channel time series data are interlaced using $y(k+2(n-1)) = x_k(n)$, where $x(n) = (x_1(n), x_2(n))'$ is the original bivariate time series. Next the individual scalar time series are modeled in a Levinson-type lattice structure model in which the PARCORS are modeled with a smoothness priors estimation procedure. The stochastic regressions of the forward and backward instantaneous PARCORS are realized using state-space modeling and a Kalman filter algorithm. (State-space modeling was reviewed in Chapter 2.) The generic state-space model is

$$x_n = F_n x_{n-1} + G_n w_n$$
$$y_n = H_n x_n + \varepsilon_n .$$

In the specific case of modeling the forward instantaneous PARCORS, $F_n = 1$, $G_n = 1$, $x_n = \alpha_{n,k}$, $H_n = \eta(j,k-1+nd)$, and $y_n = \varepsilon(j,k+nd)$. Within the Kalman filter algorithm it is assumed that $w_n \sim N(0,\tau_1^2)$ and $\varepsilon_n \sim N(0,1)$. (This is a convenient normalization. The actual variance of the $\{\varepsilon(n)\}$ is a byproduct of the computation.) For the backward instantaneous PARCORS the state space model becomes, $F_n = 1$, $G_n = 1$, $x_n = \beta_{n,k}$, $H_n = \varepsilon(j,k+nd)$, and $y_n = \eta(j,k+nd)$. We assume in this case that $w_n \sim N(0,\tau_2^2)$ and $\varepsilon_n \sim N(0,1)$. The unknown parameters τ_1^2 and τ_2^2 are hyperparameters. They are estimated in conjunction with an optimization routine that maximizes the log-likelihood of the model that is external to the Kalman filter. For optimization we use a Davidon-Fletcher-Powell variation of a quasi-Newton method, (Davidon 1959, Fletcher and Powell 1963, Fletcher 1981). The topic of optimization in statistical computations in nicely treated in Thisted (1988). The hyperparameters τ_1^2

and τ_2^2 are expressed implicitly in the model. From Chapter 2, the likelihood of the state-space model that is maximized is

$$l = -\frac{1}{2}\left[N\log 2\pi + \sum_{n=1}^{N}\log v_n + \sum_{n=1}^{N}\frac{-1}{2v_n}(y_n - H_n x_{n|n-1})^2\right].$$

In our example, the maximum order considered for each of the two scalar *AR* regressions was 7. Correspondingly the optimal values of the hyperparameters τ_1^2 and τ_2^2 for the forward and backward PARCORS, for each channel, and for each model order are listed in table below.

Table 5.1: Optimal hyperparameters

		Optimal τ_1^2 and τ_2^2	
Order	Channel	Forward	Backward
1	1	$8.854\text{x}10^{-07}$	$3.836\text{x}10^{-13}$
	2	$2.875\text{x}10^{-11}$	$1.965\text{x}10^{-17}$
2	1	$3.793\text{x}10^{-04}$	$2.242\text{x}10^{-03}$
	2	$4.671\text{x}10^{-04}$	$4.270\text{x}10^{-04}$
3	1	$1.996\text{x}10^{-12}$	$8.408\text{x}10^{-11}$
	2	$4.465\text{x}10^{-14}$	$1.395\text{x}10^{-12}$
4	1	$2.522\text{x}10^{-04}$	$6.275\text{x}10^{-09}$
	2	$2.004\text{x}10^{-14}$	$4.306\text{x}10^{-26}$
5	1	$2.387\text{x}10^{-05}$	$1.883\text{x}10^{-11}$
	2	$3.077\text{x}10^{-16}$	$3.744\text{x}10^{-12}$
6	1	$7.101\text{x}10^{-11}$	$2.373\text{x}10^{-16}$
	2	$3.646\text{x}10^{-04}$	$3.524\text{x}10^{-04}$
7	1	$3.074\text{x}10^{-08}$	$4.959\text{x}10^{-08}$
	2	$8.522\text{x}10^{-05}$	$1.827\text{x}10^{-04}$

The instantaneous PARCORS and forward and backward prediction errors $\alpha_{n,k}$, $\beta_{n,k}$, $\varepsilon(j,k+2n)$ and $\eta(j,k+2n)$ were then computed using the optimal values of τ_1^2 and τ_2^2 as identified in Table 5.1. In order to determine the AIC_C best order PARCOR model we compute the sum of squares of forward prediction errors $\varepsilon(j,k+2n)$ for k=1,2, that is, of each channel, one channel at-a-time, for each model order. The results are tabulated in Table 5.2 and the minimum values of AIC_c are indicated with "*" characters.

Table 5.2: SPMCTVAR Order Selection Using AIC_c

	AIC_c for d = 2 TVAR model			
Order	$\hat{\sigma}_1^2$	AIC_c	$\hat{\sigma}_2^2$	AIC_c
1	5.524	526.931	6.142	547.397
2	2.856	401.646	4.374	483.920
3	2.825	401.655	3.561	446.354
4	0.972	197.876*	1.105	222.674
5	0.966	198.806	1.104	224.621
6	0.966	200.959	1.076	221.793*
7	0.953	200.588	1.068	222.555

The AIC best model fitted for channel 1 is AR_4 and for channel 2 it is AR_6.

The individual scalar time varying AR models are then combined into a time varying multivariate AR model. In order to illustrate some details, we show the theoretical and estimated AR matrix coefficients at a particular instant in time, $n = 20$. The theoretical AR matrix coefficients computed from the quasi-periodic simulation model, at time n=20 are

$$A_{(20,1)} = \begin{bmatrix} -1.40865 & 0.20000 \\ 0.20000 & 1.16008 \end{bmatrix}, \quad A_{(20,2)} = \begin{bmatrix} 0.85000 & 0.20000 \\ 0.20000 & 0.85000 \end{bmatrix}.$$

The estimated AR matrix coefficients at time n=20 are

$$\hat{A}_{(20,1)} = \begin{bmatrix} -1.43377 & 0.25406 \\ 0.01668 & 1.25591 \end{bmatrix}, \quad \hat{A}_{(20,2)} = \begin{bmatrix} 0.92745 & 0.36521 \\ 0.31784 & 0.96614 \end{bmatrix}, \quad \hat{A}_{(20,3)} = \begin{bmatrix} 0.00000 & 0.00000 \\ 0.00000 & 0.08686 \end{bmatrix}.$$

That is, the simulation model was a $d = 2$, AR_2 model. The fitted model was a $d = 2$, AR_3 model.

Having computed the fitted time varying multivariate AR model we can use that model to estimate the instantaneous power spectral densities. They are estimated with the formula

$$S(f,n) = A(f,n)^{-1} V(n) A^*(f,n)^{-T},$$

where

$$A(f,n) = I - \sum_{m=1}^{P} A_{n,m} \exp[-2\pi i m f].$$

Figure 5.2 illustrates the evolution of the fitted instantaneous power spectral density in each channel, and the coherence between the channels for every 10 th data point. The graphs in Figure 5.2 closely approximates graphs of the corresponding theoretical values in Figure 5.1.

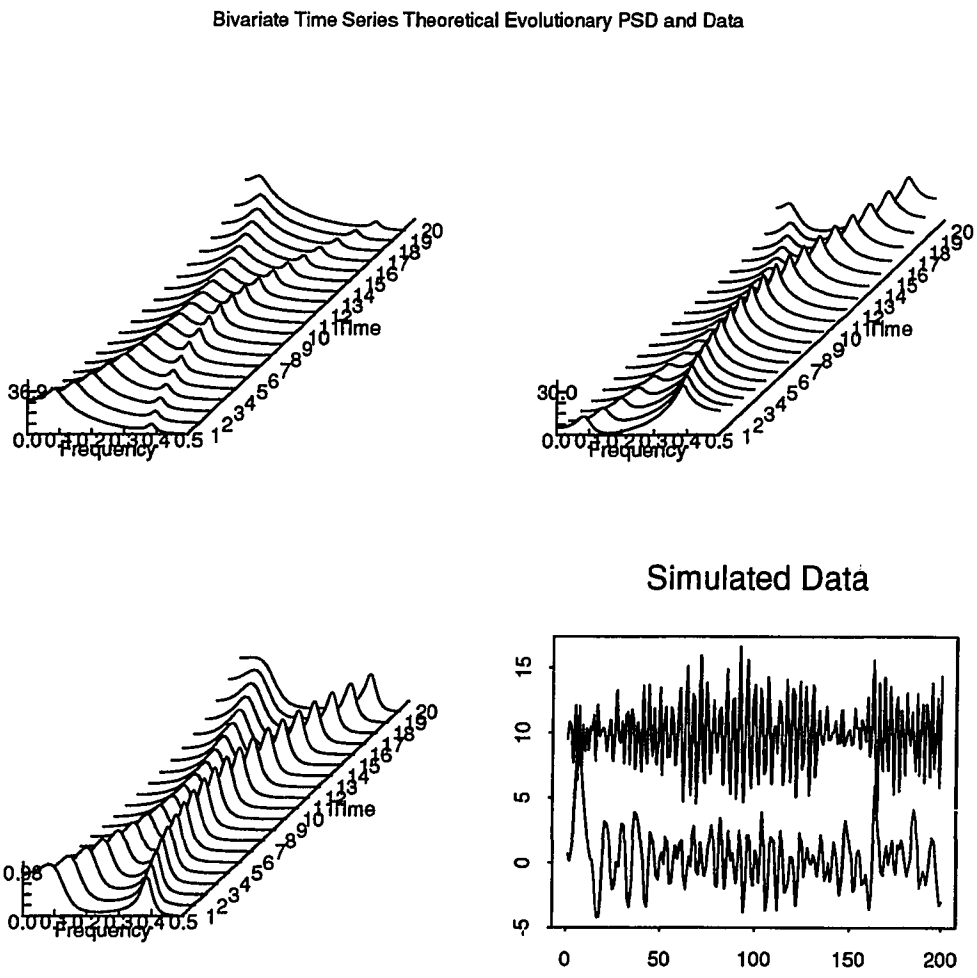

Figure 5.1: Bivariate Time Series Theoretical Evolutionary PSD and Data

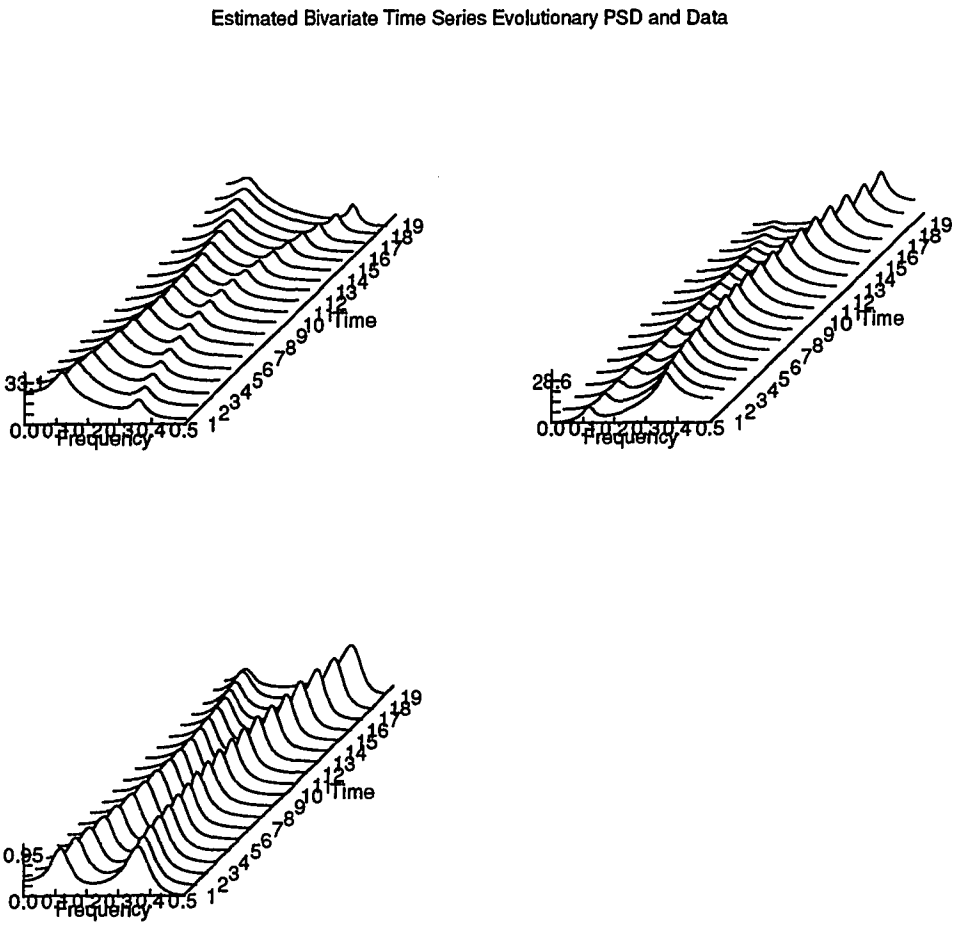

Figure 5.2: Estimated Bivariate Time Series Evolutionary PSD

5.4.2.1 SPMCTVAR d = 2 Real Data Example

In this section the results of an analysis of data provided to us by Professor George Moore, Department of Biomedical Engineering, University of Souther California, are shown. The data is that of the time intervals between alternating percussive taps of the right and left hands by a studio percussionist, that is a player with putative high temporal precision. Initially, for two seconds the performer was cued by a metronome. After that for a 3 minute interval, the performer relied on her recollection of the time intervals between beats. The notes were played on a Yamaha grand piano with MIDI interface that records the note accuracy with a resolution of about 1 msec. The original data, and the results of our analyses are shown in Figure 5.3. We analyzed the data twice. In one case we took the left hand as leading the right hand. In the other we took the right hand as leading the left hand.

Professor Moore remarked that our results appeared to be compatible with his own and in fact extended his analyses. His analysis was quite tentative being based on a serial correlation analysis which could not properly account for the time varying structure of the data. Professor Moore's explanation of the difference between the intervals in the two hands is interpreted as a hierarchical difference. That is one hand is dominant. For example, the performer may be counting the beat with one finger and the "and" with the other. In effect it is as if one hand resets the other.

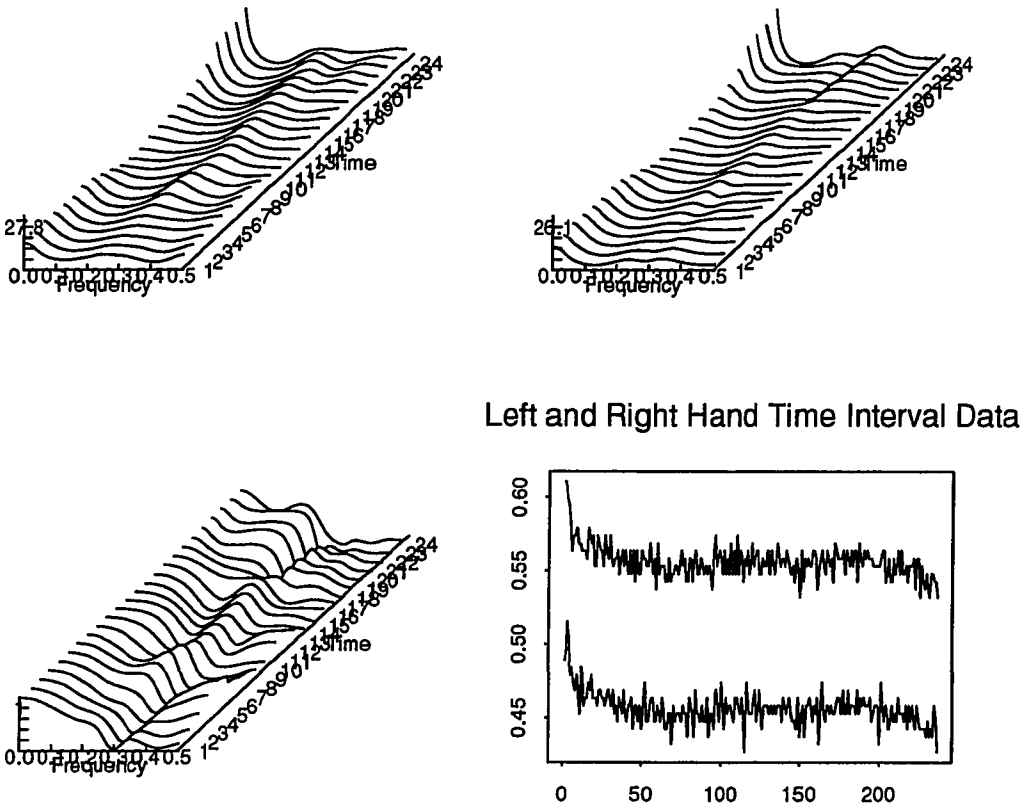

Figure 5.3: Bivariate Time Series Real Data Example

5.4.2 SPMCTVAR d = 3 EXAMPLE

A $d = 3$ *TVAR* model of order 2 is used in this example. The model is similar to the $d = 2$ TVAR model used in the previous section. A stationary AR time series that is independent of the first two channels is added as the third channel. The $d = 3$, AR_2 model is:

$$x(n) + \begin{bmatrix} 1.73 - 1.05*sin(\pi*n/200) & 0.20 & 0.00 \\ 0.20 & -1.55 + 1.25*sin(\pi*n/200) & 0.00 \\ 0.00 & 0.00 & 0.89 \end{bmatrix} x(n-1)$$

$$+ \begin{bmatrix} 0.85 & 0.20 & 0.00 \\ 0.20 & 0.85 & 0.00 \\ 0.00 & 0.00 & -0.85 \end{bmatrix} x(n-2) = \varepsilon(n),$$

$n = 200$ data points were generated from this theoretical trivariate time varying *AR* coefficient model. The one channel at-a-time SPMCTVAR method was applied to model the simulated data. Theoretically, the additional third time series has no interaction with the other two. Figure 5.4 illustrates the theoretical evolutionary power spectral density for each channel, the pairwise coherences and the pairwise partial coherences at every tenth point in time. Figure 5.5 illustrates the SPMCTVAR estimates of the theoretical model. The channel model order determined by the AIC for each of the three channels are respectively, 7 (seven), 7 (seven) and 7 (seven).

For this model, the theoretical pairwise instantaneous coherences and the theoretical pairwise instantaneous partial coherences are identically zero. In fact the maximum values of the estimated pairwise coherences, 0.11, and the estimated pairwise partial coherences, 0.08, were observed. In the application of SPMCTVAR modeling to EEG data the detection of zero partial coherence is of most concern to us. From Koopmans (1974), the distribution of the partial spectral coherence, in the vicinity of zero partial coherence is given by

$$F = \frac{(\nu-1)W^2_{12|3}(f)}{(1 - W^2_{12|3}(f))} \sim F_{2,2(\nu-1)},$$

where ν is the number of degrees of freedom in the F distribution with 2 degrees of freedom in the numerator and $2(\nu-1)$ degress of freedom in the denominator and a reasonable estimate of ν for an N observation MCTVAR order P modeled time series is $\nu = \frac{N}{P}$, (Gersch 1978). For the data analyzed, the 90% confidence interval level for zero partial coherence is 0.10. A computed partial coherence of

less than 0.10 is statistically indistinguishable from zero partial coherence. Therefore the worst case observed partial coherence of 0.08 is statistically indistinguishable from zero partial coherence.

Figure 5.6a are graphs of the instantaneous PARCORS for each of the three scalar time series in separate columns. The PARCORS are printed successively down the column for increasing *AR* model order, from *AR* order 1 (one) to *AR* order 7 (seven). The solid lines are the forward PARCORS and the dotted lines are the backward PARCORS. Figure 5.6b are graphs of the instantaneous PARCORS for each of the three scalar time series of *AR* order 7 (seven) determined by the AIC. The PARCORS for channel 1 and channel 2 vary slowly with time. This is as anticipated from Figures 5.4 and 5.5, the evolutionary power spectral density for channel 1 and channel 2. The PARCORS for channel 3 are constant. This is compatible with the model since the third channel is a stationary time series.

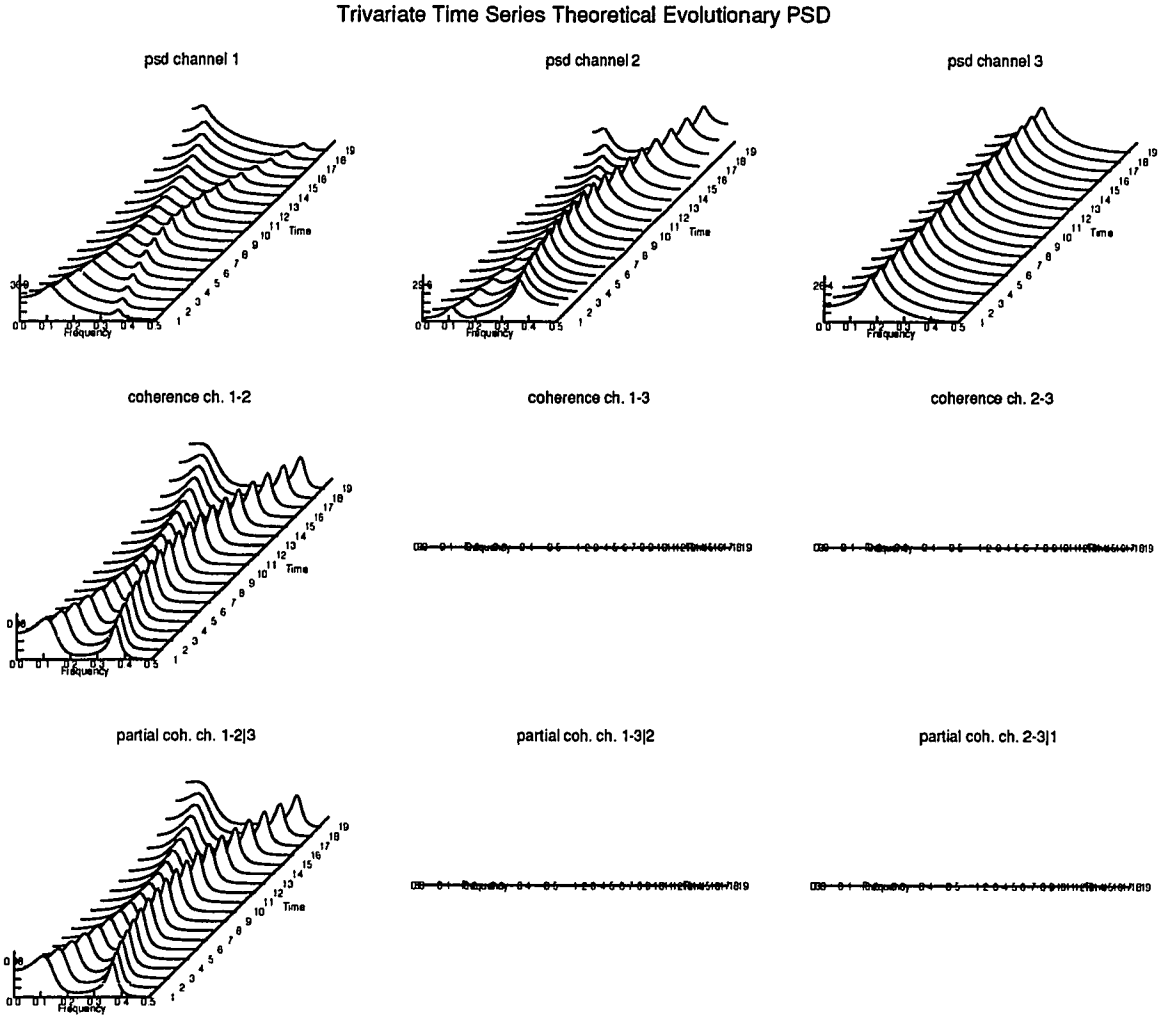

Figure 5.4: Trivariate Time Series Theoretical Evolutionary PSD

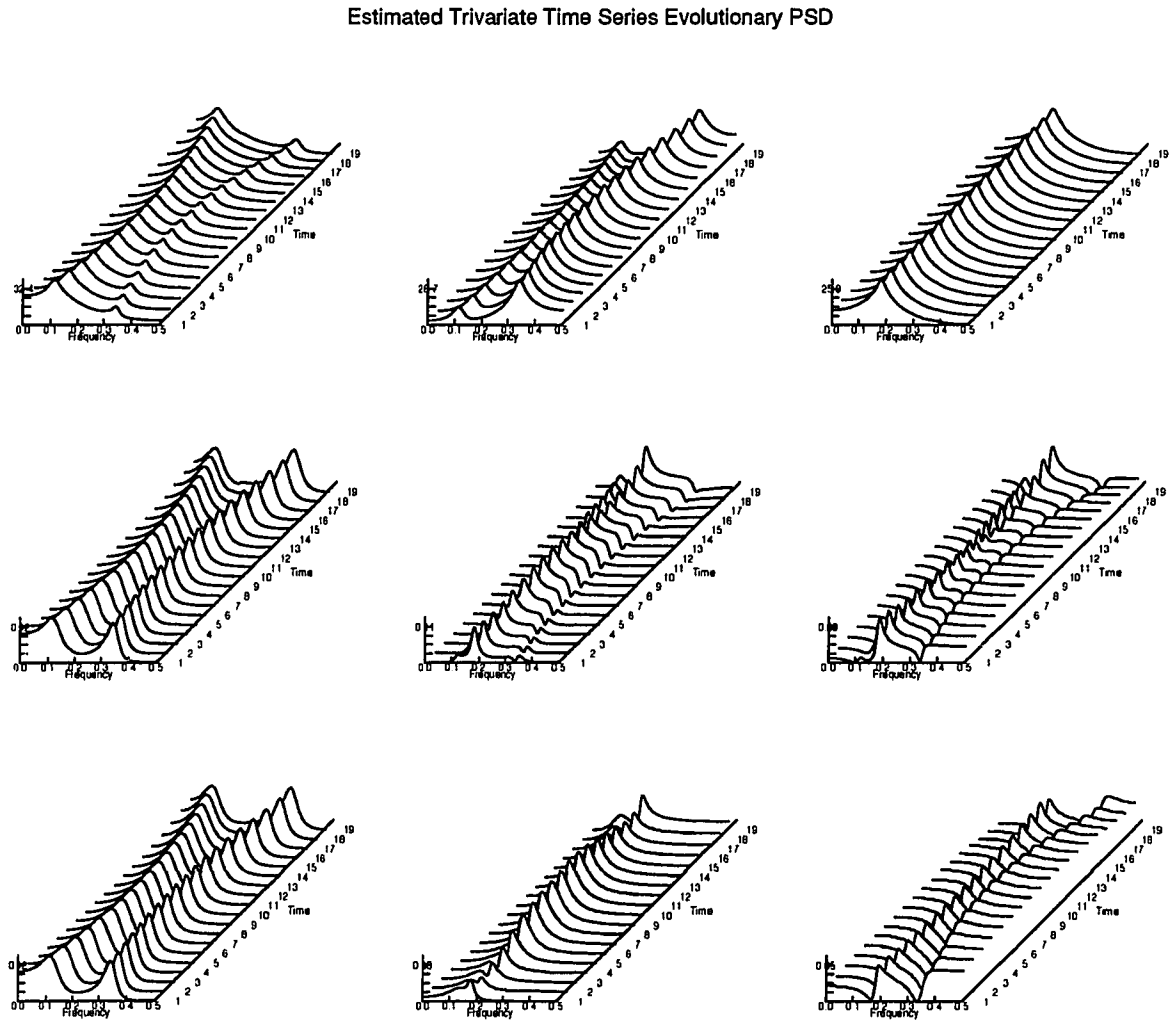

Figure 5.5: Estimated Trivariate Time Series Evolutionary PSD

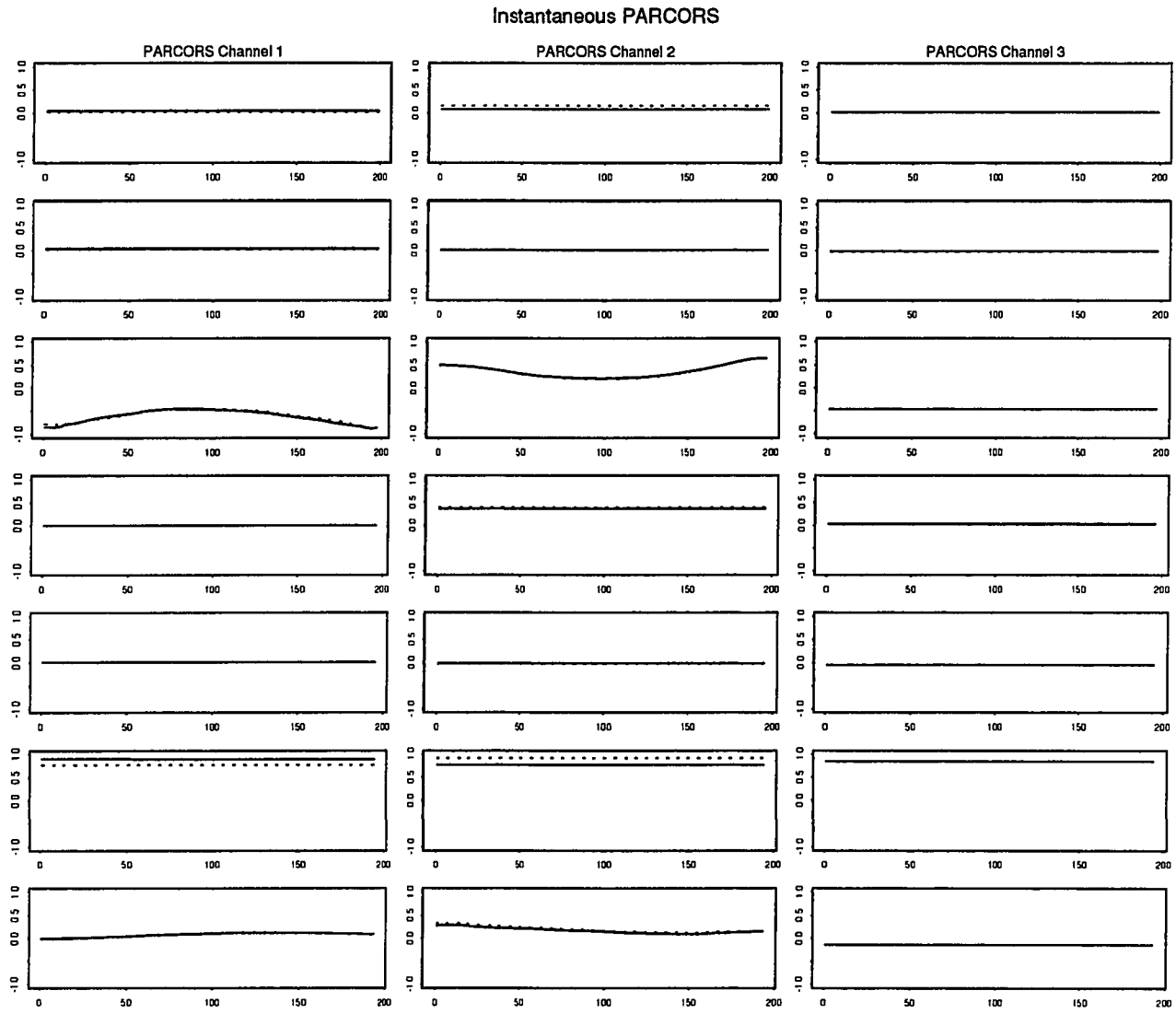

Figure 5.6a: Estimated Instantaneous PARCORS for the Trivariate Time Series

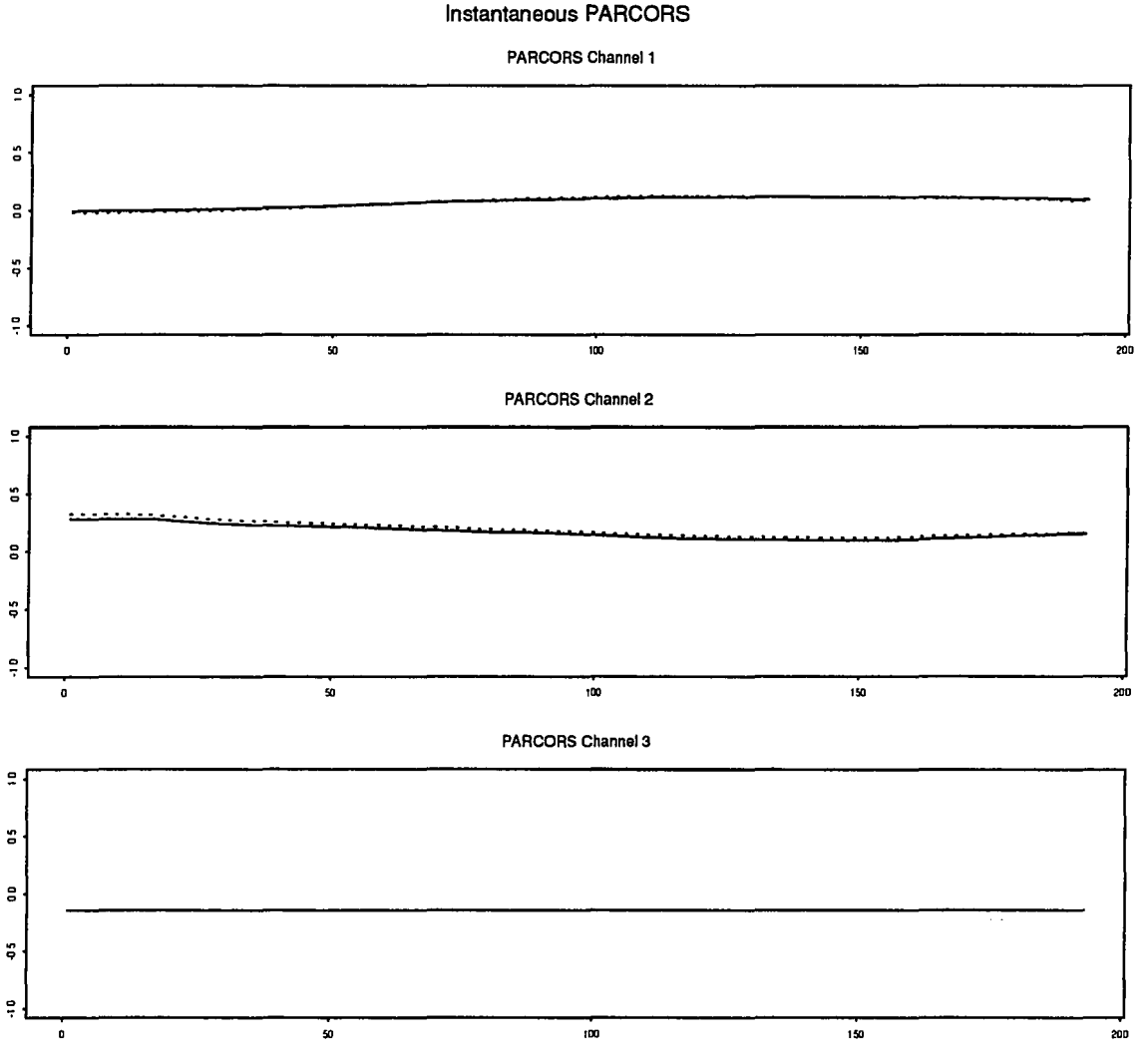

Figure 5.6b: Estimated Instantaneous PARCORS for the Trivariate Time Series

5.4.3 Analysis of a Human EEG Event

Next an example of the data analysis of a three-channel data episode of an "aura" event, a pre-epileptic seizure EEG event is shown. Some previous work in the study of human epileptic EEGs appears in Gersch (1985, 1987). The epileptic episode analyzed here is the same as that analyzed in those papers.

We illustrate a case in which an epileptic focus was located and determined to act as a pacemaker during a short time interval from an analysis of an epileptic EEG event from stereotaxically implanted electrodes in a human. Approximately 60% of individuals with epilepsy have seizures initiated from an anatomically localized brain region or seizure focus. The focus initiated seizure propagates through the brain. On the basis of several variables, (social disability arising from seizures, refractoriness to drug treatment, neurophysiological status, electroclinical correlation, radiological evidence), certain of these individuals can be identified as suitable candidates for surgical treatment. Intracranial recordings of EEG accompanying spontaneous seizures in these patients can be useful for seizure focus localization, and for guiding surgical treatment, (P.H. Crandall 1987).

Figure 5.7 illustrates data from a 7-second 6-site data record of the electrical activity at the onset of a spontaneous seizure event. In fact, 16 channels of data were continuously monitored by telemetry. The recorded data was sampled at the rate of 200 samples per second. The data was provided to us by Dr. J.P. Lieb, Reed Neurological Research Center, University of California at Los Angeles. Every 4th data point (Nyquist frequency = 25 Hertz), was used for analysis. Spectral analysis done at U.C.L.A., revealed that the effects of aliasing were inconsequential at this sampling rate. The data in Figure 5.7 is that of the dramatic onset of a spontaneous seizure in which only right mesial temporal lobe activity was implicated, (only right sided activity is shown in Figure 5.7). The clinical diagnosis was that the patient had medically refractory complex partial seizures that appeared to initiate in the right amygdala and hippocampal formation.

The objective of the analysis was to determine whether the electrical activity at any one of the observed anatomical recording sites could be interpreted as "driving" the electrical activity at the other

sites. The data is complex, and there is no obvious visual clue that might unambiguously identify an initiating or driving site or identify when driving is present.

The appearance of each individual EEG trace reveals that the local structure of the time series changes with time. Clearly what we observe, is a nonstationary covariance time series. That is, the statistical properties of the series vary with time. Our analysis of these nonstationary time series involves the fitting of a multivariate time varying autoregressive model, via the SPMCTVAR approach, to the data and performing subsequent spectral analysis using the concept of an instantaneous power spectral density matrix.

The simultaneous EEG's in the channels labeled 1,2 and 3 in Figure 5.7 are analyzed here by the SPMCVAR method. The top row of Figure 5.8 shows the instantaneous power spectral density matrix as estimated from the fitted SPMCTVAR model versus frequency for channels 1,2 and 3. Initially there are relatively sharp concentrations of energy in each of channels 1,2 and 3 at 7.25 Hertz and 14.5 Hertz. As the EEG evolves those peaks become diminished in amplitude and the distribution of spectral energy becomes more diffuse. Later in the record the emergence of a spectral peak at 4.25 Hertz can be observed in each channel.

We identify causality or driving at an instant in time by a frequency domain analysis via the instantaneous power spectral density matrix. The concept of causality that we use was introduced for stationary time series in Gersch and Goddard, (1970). Gersch (1972), expressed the concept of causality in a mathematical model. D. Brillinger (1975) used that concept for identifying causality between neural activity in Aplaysia. Our concept of causality differs from the "feedback free" definition of causality in econometrics (Geweke 1982), or the time precedence definition used in the location of epileptic foci from EEG time series. The concepts of instantaneous coherence and instantaneous partial coherence versus frequency are involved in our concept of instantaneous causality or instantaneous driving. Driving is operationally defined in terms of a statistical test for zero partial coherence.

Our definition of causality at an instant in time exploits the concept of causality for stationary time series. It follows from conditions placed on a measure of instantaneous information between vectors of times series, (Gersch 1987), and is equivalent to the following conditions on spectral coherences and partial coherences

$$W^2_{x,y}(f_A,t) \neq 0, \quad W^2_{x,w}(f_A,t) \neq 0, \quad W^2_{y,w}(f_A,t) \neq 0,$$

$$W^2_{x,y|w}(f_A,t) \neq 0, \quad W^2_{x,w|y}(f_A,t) \neq 0, \quad W^2_{y,w|x}(f_A,t) = 0.$$

where $W^2_{x,y}(f_A,t)$, and $W^2_{x,y|w}(f_A,t)$ are respectively the coherence between $x(t),y(t)$ and the coherence between $x(t),y(t)$ regressed on $w(t)$, (partial coherence), at frequency f_A and time t. That is, the identification of causality at frequency f_A at time t is determined by the detection of zero partial instantaneous coherence. The instantaneous coherences and partial coherences are computed from the instantaneous power spectral density matrix.

Spectral coherence is a measure of the linear dependence between two time series at frequency f. It is the square of the correlation coefficient between the energies of two different time series at a frequency f, (for any f in the interval $0 \leq f \leq 25$ Hertz). Partial coherence is the coherence between two time series after the influence of a third time series is removed by regression from each of the two time series. The SPMCTVAR model yields estimates of the spectral coherence and partial spectral coherence at every instant in time. Computational results of both the evolving spectral coherence and evolving partial spectral coherence for each of the three possible pairs of three different data channels are shown in Figure 5.8 in row 2 and row 3. The computational results are used to determine whether or not and when one channel drives the other two.

The second row of Figure 5.8 illustrates the time-changing structure of pairwise coherence versus frequency between pairs of channels and the corresponding time-changing partial coherence versus frequency. At one and two seconds into the record, there are sharp coherence peaks between each of the of the channel pairs at both 7.25 and 14.5 Hertz, the frequencies of largest concentration of spectral energy in each channel. The partial coherence between channels 1 and 2 and between channels 1 and 3 with the partialing done on the excluded third channel remain significantly correlated. On the other hand, the partial coherence between channels 2 and 3 with the influence of channel 1 removed is statist-

ically indistinguishable from zero over the frequency interval between 7 and 11 Hertz. We conclude that the pairwise coherence between the three channels is a consequence of the fact that channel 1 was driving channels 2 and 3 for the first 2 1/2 seconds of the record in the vicinity of 7.25 Hertz. Channels 2 and 3 have very little explanatory power.

The relatively low partial coherence between channels 2 and 3 partialed on channel 1 throughout the epoch, (third row of Figure 5.8), results in the appearance of a relatively flat partial coherence surface and suggests that this kind of illustration may be a potentially powerful diagnostic aid to identify candidate driving channels.

The covariance structure and hence the ability to identify the location of the focal driving site of the epileptic episode changes over a very short time interval. That fact makes a proper method of analysis of multichannel nonstationary covariance data critical for the analysis of human focal epilepsy EEG data.

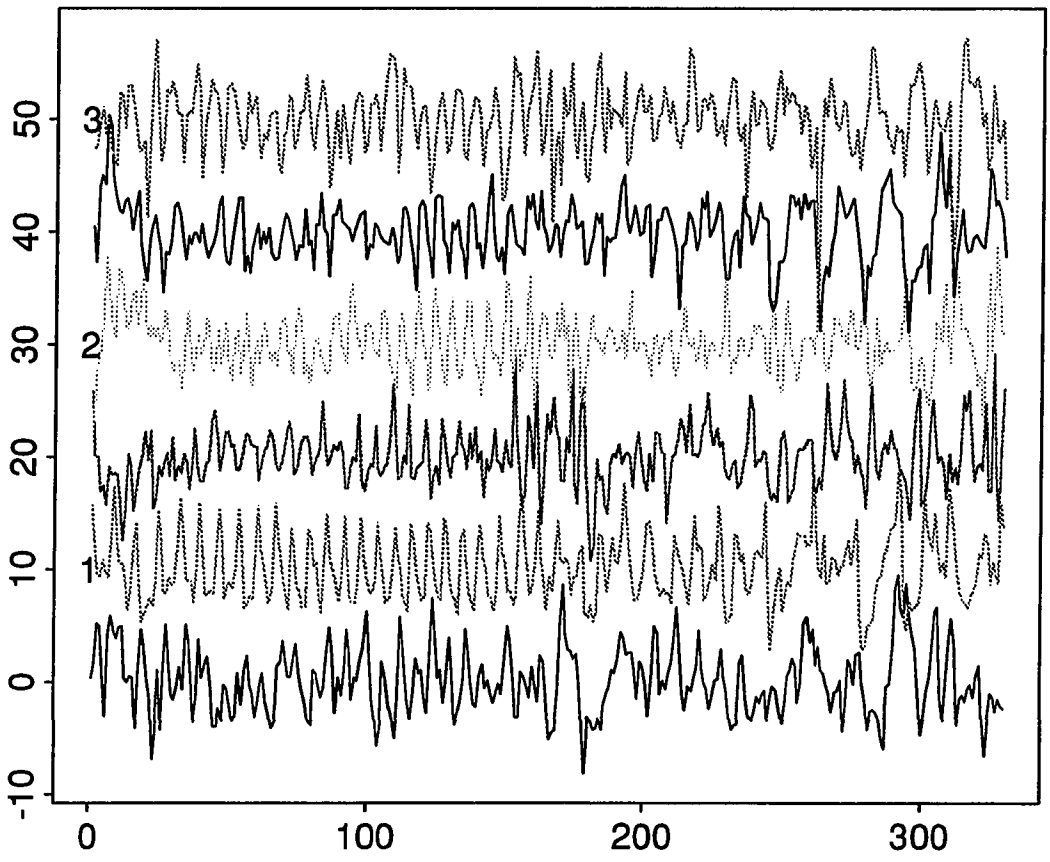

Figure 5.7: Six Channel Human EEG Data

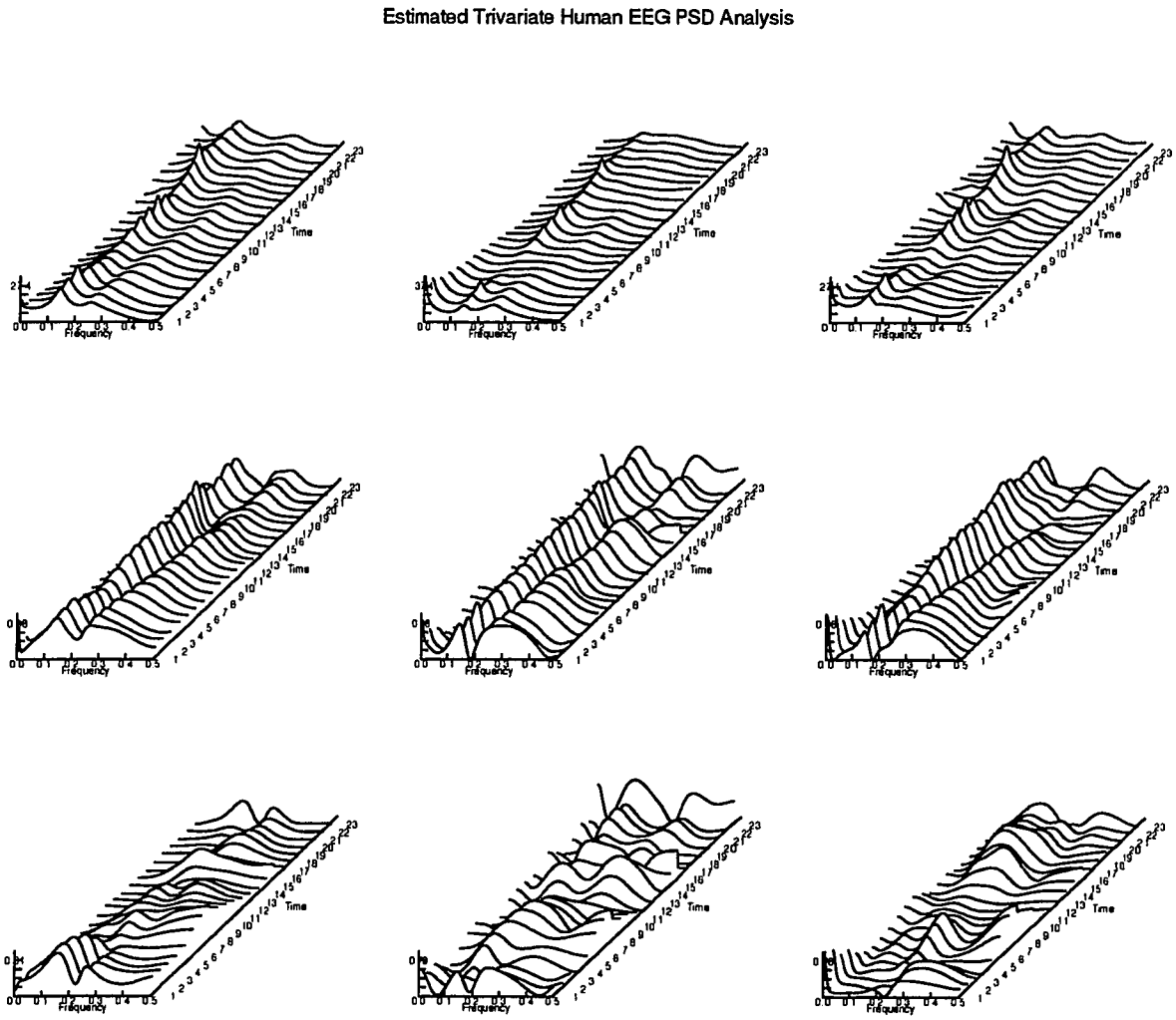

Figure 5.8: Estimated Trivariate Human EEG Evolutionary PSD Analysis

CHAPTER 6

SUMMARY, SUGGESTED RESEARCH AND CONCLUSIONS

This thesis addressed some problems in the modeling of multichannel stationary and nonstationary time series. A unifying and motivating idea in this thesis was to "do multichannel things one channel at-a-time". That is, we considered fitting multichannel autoregressive models to both stationary and nonstationary covariance data by exploiting an instantaneous response-orthogonal innovations multichannel autoregressive model. That approach yields a set of orthogonal scalar time series which we could then model using scalar time series modeling techniques on instantaneous data.

A basic objective of the thesis was to verify the potential usefulness and seeming generality of the "do multichannel things one channel at-a-time" approach. Toward that end only the simplest distributional assumptions that everything is multivariate normally distributed, were made. We proved no theorems. We examined the practical applicability of the "do multichannel things one channel at-a-time" approach empirically in addressing the several different areas of application by working examples. That involved a rather substantial amount of programming and computing.

The problems addressed were those of achieving parsimonious modeling of both nonstationary mean time series and second order stationary time series, (and thereby achieving a decreased prediction variance for stationary time series as compared to what one might obtain using more conventional MCAR modeling), of achieving enhanced spectrum estimation for multichannel stationary covariance time series as compared to that of more conventional MCAR modeling methods, and of modeling nonstationary covariance time series. In this chapter summaries of results and suggested further research are discussed for each of the aforementioned topics. A brief conclusion is also offered.

6.1 SUMMARY OF RESULTS

Specifically, three major problems in the fitting of multivariate time series to data are addressed. One is the tendency in multivariate *AR* models of both nonstationary mean and stationary mean covariance stationary time series to overparameterize. Our objective was to attempt to achieve a more parsimonious modeling of MCAR modeled time series. Our approach was to exploit an instantaneous

response orthogonal innovations representation of multivariate autoregressive time series. This results in a one scalar autoregressive channel at-a-time modeling paradigm. That and subsequent use of subset selection in a linear regression modeling of the scalar autoregressive time series realizes a relatively parsimonious MCAR modeling. Another topic treated in the thesis is the attempt at achieving an enhancement of power spectral density estimation in multichannel stationary time series as compared to more conventional methods of MCAR modeling. Our approach to achieving such enhancement is by using the one channel at-a-time paradigm in conjunction with the use of smoothness priors distributions on the parameters of the individual scalar autoregressive models. Finally the problem of the modeling of multivariate nonstationary covariance time series is treated using the idea of one channel at-a-time modeling and smoothness priors distributions on the partial correlation coefficients (PARCORS) in each scalar lattice AR model.

A summary of the results obtained for each of individual problem areas follows.

6.1.1 Parsimonious MCAR Modeling

Our efforts in parsimonious multichannel autoregressive modeling were motivated by the inherent tendency to overparameterize that is present in the conventional MCAR modeling. We showed a procedure for achieving a relatively parsimonious MCAR modeling that was suitable for the modeling of both nonstationary mean and second order stationary time series. The application of the one channel at-a-time paradigm and subsequent subset selection of the individual scalar AR modeled channels achieves a "thinning" of the overparametrized MCAR models.

We achieve multichannel modeling doing things one channel at-a-time using only scalar computations on instantaneous data. The one channel at-a-time modeling is achieved with an instantaneous response-orthogonal innovations variance matrix. Each channel is thought of as a scalar autoregressive (AR) time series and is expressed in a linear regression model form which is conducive to subset selection procedures. Akaike's AIC was used to determine the order of the scalar autoregressions as well as to achieve subset selection in a modification of Furnival's "leaps" all-subsets search function in the S language.

Comparative studies of other published parsimonious modeling attempts were done. Our modeling was in fact more parsimonious in the modeling of trivariate nonstationary mean time series data than that achieved by Tiao and Tsay (1986) and that achieved on bivariate stationary time series data by Hsiao (1978). The mean square tracking error performance of Hsiao's data by our method was superior to that achieved by his method. In addition, using our parsimonious modeling method, we demonstrated superior mean square tracking error performance of the trivariate bathymetry data as compared to that achieved by the more conventional Yule-walker AIC method of MCAR modeling. We also conducted a Monte Carlo simulation study of one step ahead prediction performance using both the conventional MCAR and our one channel at-a-time parsimonious modeling. The Newton model of the Box-Jenkins Series J data was used for the simulation model. The results were that our method did in fact have smaller one-step-ahead prediction variances than the conventional MCAR modeling method. A detailed example of the parsimonious modeling of simulated Newton model data was shown.

6.1.2 SPMCAR Model Spectrum Estimation

The acronym SPMCAR refers to a smoothness priors multichannel autoregressive method of MCAR modeling. That is a new method for MCAR modeling that is intended specifically for multichannel power spectrum density estimation. Our strategy combines the one channel at-a-time and smoothness priors paradigms. For the estimation of the power spectral density matrix of a stationary d-channel time series, d scalar AR time series are obtained using the instantaneous response orthogonal innovations model. That permits us to model each of the d scalar AR time series one channel at-a-time as a long AR order model with smoothness priors constraints on the AR model coefficients. In each channel, the priors on the AR parameters are computed from the integrated squared kth derivative with respect to frequency of the departure from model smoothness. The AR model parameters and two hyperparameters are estimated via a maximization of the likelihood of the hyperparameters.

SPMCAR modeling achieves an approximate Bayesian estimate of power spectral density. Such modeling ordinarily requires that the spectral density of parametric models of different model orders be weighted in accordance with the likelihood and the prior of the model order of different models. SPMCAR modeling achieves such a modeling in one step. In comparison with conventional MCAR

modeling with short data spans, SPMCAR modeling ameliorates the problem of the small sample variability of whatever statistic used for determining the best order of the parametric model fitted to the data.

Monte Carlo simulation studies of power spectral estimation performance on bivariate time series for both the conventional MCAR and SPMCAR models were conducted. A sum of relative squares of the difference between the estimated model function and the true model function was computed as a goodness of fit criterion. Results were computed for the four functions, power spectral density in channels one and two, coherence squared and transfer function as functions of frequency. In modeling multichannel simulated AR model data, the model fitted is in the same model class as the model which generated the data. In the case of the simulated ARMA model data, the model fitted is an approximation to the model class which generated the data. The results obtained were that in both cases the SPMCAR method of spectrum estimation was superior in the relative mean square error sense than the conventional Yule-Walker AIC method of MCAR modeling. The SPMCAR was also superior to the conventional MCAR modeling in the sense that the results were more stable. That is, the variances of each of the estimated quantities were smaller than those for the MCAR model. Also and quite importantly, the SPMCAR model exhibited a significant insensitivity to choice of model order.

6.1.3 SPMCTVAR Modeling

The acronym SPMCTVAR refers to our smoothness priors multichannel time varying autoregressive modeling approach to the modeling of multichannel nonstationary covariance time series.

A discussion of the SPMCTVAR algorithm for the realization of multichannel autoregressive modeling of nonstationary covariance time series data and several worked examples were shown. The fitting of a multichannel time varying autoregressive model to data requires the estimation of many more parameters than there are data. Specifically if there are d channels of N data points each, or a total of $N \times d$ observations and an order P multichannel time varying autoregressive model is fitted to that data, then $N \times P \times d \times d$ autoregressive parameters and $N \times d \times (d + 1)/2$ instantaneous innovations variance parameters describe that model. The SPMCTVAR modeling algorithm requires that only $2 \times d$ hyperparameters be explicitly estimated for each AR model order considered. The final MCTVAR

model parameters of the best fitted model are actually expressed implicitly in terms of the hyperparameters of the likelihood best model.

The SPMCAR modeling was achieved doing multichannel things one channel at-a-time using only scalar computations. Each scalar AR time series is modeled using a Levinson type lattice structure. The additional "trick" in our procedure was to estimate the instantaneous evolution of the partial correlation coefficients (PARCORS), in a stochastic regression using smoothness priors. The the instantaneous PARCORS were modeled as a non-stationary trend time series using the state-space Kalman filter algorithm with which the likelihood of the hyperparameters was computed. In each scalar time varying autoregression, the best scalar AR model order was determined via the sum of squares of the forward prediction errors.

Worked examples of the SPMCTVAR algorithm, using bivariate and trivariate multichannel time varying autoregressive simulation models were demonstrated. In addition applications of the SPMCTVAR modeling method to the analysis of nonstationary covariance bivariate interval timing data from an experiment conducted with a percussion studio performer and trivariate nonstationary human EEG data were shown.

6.2 FURTHER RESEARCH

In each of the multivariate time series modeling problem considered we assumed that the data was multivariate normally distributed. In several of the problem areas that assumption is not necessary. In this section, where we understand how to do it, we show how that assumption can be relaxed. In addition there are significant number of other research areas that are well worth considering in order to extend the applicability of the do multichannel things one channel at-a-time paradigm. The suggested research activities are treated here in chapter order.

6.2.1 Research in Parsimonious Modeling

Several research issues remain to be explored in comparing the prediction performance of the conventional Whittle-AIC MCAR modeling and our one channel at-a-time subset selection MCAR

modeling. Perhaps the most intriguing analytical and empirical research problem is to explore the potential usefulness of the relationship of the decomposition of information between vectors of time series (Geweke 1982,1984), to the expected one-step ahead prediction variance improvement. Geweke's result is that the information between vectors of time series can be decomposed into the (feedforward) information from one set of time series to the other set of time series, a second (feedforward) information from the second set of time series to the first set of time series and an instantaneous mutual information between the sets of time series. A reasonable conjecture is that the enhancement in one-step-ahead prediction variance performance in a particular channel will increase as the relative feedforward information from a set of time series to that particular channel increases. This research requires analytic analysis as well as empirical computational study.

Secondly, it is known that even for normally distributed data, the one-step-ahead prediction is itself not normally distributed. A research topic of interest would be a study of the distribution of prediction errors. It is not clear how to do that analytically. That may be accomplished empirically via simulation studies with a bootstrap analysis.

Model order determination is necessary both in fitting the conventional MCAR models and the one channel at-a-time AR models to data. In each situation of model order determination the assumption that the data was normally distributed was the motivation for the use of the AIC criterion. It is still an open problem to determine a model order selection criterion for non-normal data. A rational but very lengthy computational research study would be to study other penalized least squares AR model order determination schemes when the generating data was not normal. A first start might be to simulate data from multichannel AR models with say t distributed data and examine the influence of the choice of the parameter penalty term, (for the AIC, it is 2 times the number of fitted parameters), on prediction performance.

The situation is a little simpler when we consider subset selection in scalar autoregressive models. The AIC and AIC_C criteria are based on the assumption of normality. Instead we could use Allen's PRESS, (Allen 1974), a cross-validation method of computing the predictions of successively deleted data points, as a distribution free method for achieving subset selection.

6.2.2 Possible Research in SPMCAR Modeling

There are several research issues to explore in order to examine the relative improvement achieved by SPMCAR modeling over conventional Yule-Walker AIC MCAR modeling. The criterion we used in a Monte Carlo simulation study to compare the spectrum estimation performance of the two different modeling methods was a heuristic, mean of relative square errors criterion. That criterion was used in simulation studies of bivariate time series. The spectrum estimation in each of two channels, transfer function estimation and coherence estimation were evaluated for both methods of MCAR modeling. As an alternative, a rational global criterion based on the Kullback-Leibler number (Kullback 1959), might be researched in order to more rigorously quantify the difference between the modeling performance of the alternative forms of MCAR modeling. The Kullback-Leibler (K-L) number is a measure of dissimilarity between distributions. Gersch (1981) showed a theorem for computing the K-L number between a theoretical MCAR model and an empirical MCAR model for normally distributed data. The K-L number result can be computed either from the theoretical and fitted MCAR models as a kind of discrepancy in template matching or from the theoretical and estimated power spectral densities. Computing such a measure would yield a more theoretically based measure of the difference in the performance of the alternative forms of MCAR modeling.

Additional comparative computational studies of the difference in performance between the alternative modeling approaches for spectrum estimation might be done for data sets of different lengths. The Yule-Walker AIC MCAR modeling method is based on covariance computations. Those computational results and the MCAR model order tend be increasingly unstable with decreasing duration data spans. The question of interest to be examined in a simulation study of involving longer duration data is how the increased accuracy in covariance function estimation would be balanced by the increased order MCAR models that would be fitted to the increased data spans. An increased MCAR model would tend to be fitted with an increased data span is based on an assumption in the development of the AIC. That assumption is that the true model which generated the data is of infinite order and that the finite order model fit is only an approximation to that infinite order model.

Finally, inasmuch as there are essentially no available analytic results on the statistical performance of multichannel autoregressive modeled spectrum estimation it would desirable to do simulation studies to attempt to codify that performance.

6.2.3 Further Research in SPMCTVAR Modeling

The multichannel time varying AR model has potential applications in the analysis of a variety of multichannel nonstationary time series physical and socio-economic data. In addition, the model fitted to a single event occurrence can be used for simulation purposes. Such an application might for example be used to study the responses of building structure models to the three dimensional vibrations of simulated earthquakes which replicate the covariance structure of a recorded earthquake. A potentially important application of this algorithm in the analysis of nonstationary EEGs include the study of the evolution of brain structure participation in epilepsy and in cognition.

The SPMCTVAR algorithm could be applied to the spectral analysis of multichannel stationary time series data. Application of the SPMCTVAR algorithm to the stationary data case implicates very small values of the hyperparameters. One consequence of using the SPMCTVAR approach for the modeling of stationary time series is that such stationary multichannel AR modeling could be realized with relatively short data spans. The statistical performance of the SPMCTVAR and SPMCAR models for multichannel spectrum estimation could be compared in a simulation study.

With some modifications, the SPMCTVAR algorithm could be applied to time series whose covariance structure changes sharply with time. This could be accomplished by exploiting a version of the state space nonstationary- nongaussian model for the instantaneous PARCORS exhibited in Kitagawa (1988).

Also, it would be desirable to conduct Monte Carlo studies to get some sense of the confidence interval estimates that might be expected from spectrum estimates achieved using the SPMCTVAR algorithm.

6.3 CONCLUSIONS

The basic objective of the thesis, to verify the potential usefulness and seeming generality of the "do multichannel things one channel at-a-time" approach appears to have been successfully concluded. The crucial new idea was to systematically exploit an instantaneous response-orthogonal innovations multichannel autoregressive model in addressing several important problems in the modeling of multichannel stationary and nonstationary covariance data. That approach yielded a set of orthogonal scalar time series which we could then model using scalar time series modeling techniques on instantaneous data. In each of the problems addressed we appear to have demonstrated improvement over existing approaches.

REFERENCES

Akaike, H. (1969), Fitting autoregressive models for prediction, Ann. Inst. Stat. Math., 21, 243-247,

Akaike, H. (1970), On a semiautomatic power spectrum estimation procedure, in Proc. 3rd Hawaii Intl. Conf. System Sciences, 974-977.

Akaike, H. (1970), Statistical predictor identification, Ann. Inst. Stat. Math., 22, 203-217,

Akaike H. (1971), Autoregressive model fitting for control, Ann. Inst. Stat. Math., 23, 163-180,

Akaike, H. (1972), Use of an information theoretic quantity for statistical model identification, in Proc. 5th Hawaii Intl. Conf. System Sciences, 99-101.

Akaike H. (1973), Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory, (eds. B.N. Petrov and F. Csaki), Akademia Kiado, Budapest, pp 267-281.

Akaike H. (1974), A new look at the statistical model identification, IEEE Transactions on Automatic Control, AC-19, 716-723.

Akiake H. (1987a) Factor analysis and AIC, Psychometrika, 52, 345-370.

Akaike H. (1987b) Some reflections on the modeling of time series, In Time Series and Econometric Modeling, editors I. B. MacNeil, G.J. Umphrey, Reidel, Holland, 13-28.

Akaike H., Kitagawa G, Arahata E. and Tada F. (1979), *TIMSAC-78, Computer Science Monograph No. 10*, Institute of Statistical Mathematics, Tokyo

Akaike, H. (1980a), Likelihood and the Bayes procedure, *Bayesian Statistics*, (eds. J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), University Press, Valencia , Spain, 143-166.

Akaike, H. (1980b), Seasonal Adjustment by a Bayesian Modeling, Journal of Time Series Analysis, 1,1-13.

Allen D.M. (1971), Mean square error of prediction as a criterion for selecting variables, Technometrics, 13, 469-475.

Anderson T.W.A. (1971), *The Statistical Analysis of Time Series*, Wiley, New York.

Bartlett M.S. (1955), *An Introduction to Stochastic Processes*, Cambridge University Press, London.

Becker, R.A. , Chambers J.M. and Wilks A.R. (1988), *The New S Language, A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole, Pacific Grove.

Berger J.O. (1980), *Statistical Decision Theory*, Springer-Verlag, New York.

Beveridge W. H. (1921) Weather and harvest cycles, Econ. J. 31, 429-452.

Blackman RB and Tukey J W (1959), *The measurement of Power Spectra from the Viewpoint of Communications Engineering*, Dover, New York.

Bozdogan H and Sclove S. L. (1987), Multi-sample cluster analysis using Akaike's information criterion, Psychometrika 52, 345-370.

Brillinger D R (1975), *Time Series: Data Analysis and Theory*, Holt, Rhinehart and Winston , New York.

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

Brotherton T. and Gersch W. (1981), A Data Analytic Approach to the Smoothing Problem and Some of its Variations, Proceedings of the 20th IEEE Conference on Decision and Control, 1061-1069.

P.H. Crandall, (1987), *Surgical Treatment of the Epilepsies*, J. Engel Jr. ed., Raven Press, New York, 377-404.

Davidon W.C. (1959), Variable metric methods for minimization, AEC Research and Development Report ANL-5990, Argonne National Laboratory, Illinois.

Durbin J (1960), The fitting of time-series models, Rev. Int. Inst. Stat. 28, 233-244.

Einstein A. (1905), On the theory of Brownian motion, Annalen der Physik 17, 371-381.

Findley D. F. (1981), Geometrical and lattice versions of Levinson's general algorithm, in: *Applied Analysis II.* Ed. D.F.Findley, Academic Press , New York.

Findley D. F. (1985), On the unbiasedness property of AIC for exact or approximating linear stochastic time series models, Jo. Time Series Analysis 6, 229-252.

Fletcher R. (1981), *Practical Methods in Optimization*, Wiley, Chichester.

Fornell and Rust (1989), Incorporating prior theory in covariance structure analysis: A Bayesian approach, Psychometrika, 54, 249-259.

Friedlander B. (1982), Lattice filters for adaptive processing, Proceedings IEEE, 70, 829-867.

Furnival , G. M. and Wilson R. W. Jr. (1974), Regressions by leaps and bounds, Technometrics, 16, 499-511.

Gersch W. (1981), Nearest neighbor rule classification of stationary and nonstationary time series, in: *Applied Time Series Analysis II*, D. F. Findley editor, Academic Press, 221-270.

Gersch W. (1985), Modeling nonstationary time series and inferring instantaneous dependency, feedback and causality: An application to human epileptic seizure event data, Identification and System Parameter Estimation 1985, *7 th IFAC/IFORS Symposium*, York, UK, 737-742.

Gersch W. (1987a), The linear dependency structure of covariance nonstationary time series, Technical Report NPS55-87-006, Naval Postgraduate School, Monterey California.

Gersch W. (1987b), Non-Stationary Multichannel Time Series Analysis, Chapter 10, *EEG Handbook, Revised Series*, vol 1, (ed. A. Gevins), Academic Press, 261-296, 1987.

Gersch W. (1991), Nonstationary covariance time series: Modeling by smoothness priors methods, *Systems and Control*, (eds. T.Ono and F. Kozin), Mita Press, Tokyo, 73-84.

Gersch W. (1992), Smoothness priors, in: *New Directions in Time Series Analysis Part II*, D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu editors, The IMA Volumes in Mathematics and its Applications, Vol. 46, Springer-Verlag, 113-146.

Gersch W. and Goddard G.V. (1970), Epileptic focus location, spectral analysis method, Science, 150, 701-702.

Gersch, W. and Kitagawa, G. (1983), The Prediction of Time Series with Trends and Seasonalities, Journal of Business and Economic Statistics, 1, 253-264.

Gersch W. and Kitagawa G. (1983), A multivariate time varying autoregressive modeling of nonstationary econometric time series, Proceedings of the Business and Economics Statistics Section, American Statistical Association, Toronto Canada, August 1983.

Gersch W. and Kitagawa G (1985), A Time Varying AR Coefficient Model for Modeling and Simulating Earthquake Ground Motion, Earthquake Engineering and Structural Dynamics, 13, 243-254.

Gersch W. and Kitagawa G. (1988), Smoothness Priors in Time Series, in: *Bayesian Analysis of Time Series and Dynamic Systems*, J.C. Spall ed, Marcel Dekker,431-476.

Gersch W. and Sharpe D R, (1973) Estimation of power spectra with finite-order autoregressive models, IEEE Trans. Automatic Control, AC-18, 367-369.

Good I.J. (1965), *The Estimation of Probabilities*, M.I.T. Press, Cambridge, Mass.

Good I.J. and Gaskins J.R. (1980), Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data, 75, 42-73 Journal of the American Statistical Association, 75,42-73.

Geweke J. (1982), The measurement of linear dependence and feedback between multiple time series, (with discussion), Journal of the American Statistical Association, 77, 304-324.

Goff J.A. and Jordan T.H. (1988) Stochastic modeling of sea floor morphology: Inversion of Sea Beam data for second-order statistics, Journal of Geophysical Research, 93, 13, 589-13, 608.

Guo L. (1990), Estimating time-varying parameters by the Kalman filter based algorithm: Stability and convergence, IEEE Transactions on Automatic Control, AC-35, 141-147.

Guo L. , L. Xia and J.B. Moore (1991), Tracking randomly varying parameters: Analysis of a standard algorithm, MCSS, Mathematics of Control, Signals and Systems.

Granger CWJ and Hatanaka M (1964), Spectral Analysis of Economic Time Series, Princeton University Press, New Jersey.

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

Hsiao C. (1979), Autoregressive modeling of Canadian Money and Income Data, Journal of the American Statistical Association, 74, 553-560.

Hurvich C. M. (1992), Selection of time series models and spectrum estimates using a bias-corrected generalization of the AIC, in: *New Directions in Time Series Analysis Part I*, D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu editors, The IMA Volumes in Mathematics and its Applications, Vol. 46, Springer-Verlag, 155-168.

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

Hurvich C. M. and Tsai C. L. (1992), A corrected Akaike information criterion for vector autoregressive model selection, Report SOR-92-10 New York University L. N. Stern School of Business.

Ishiguro M. and Sakamoto Y. (1983), A Bayesian Approach To Binary Response Curve Estimation, Annals of the Institute of Statistical Mathematics,35,B,115-137.

Ishiguro M, Akaike H., Doe M. and Nakai S.(1981) A Bayesian Approach to the Analysis of Earth Tides, Proc. 9th Inst. Conference on Earth Tides.

Jones RH (1964), Prediction of multivariate time series, J. Appl. Meteorology, 3, 285-289.

Jones R H (1973), Autoregressive spectrum estimation, in 3rd Conf. Probability and Statistics in Atmospheric Sciences.

Jones RH (1974), Identification and autoregressive spectrum estimation, IEEE Trans. Autom. Control, AC-19, 894-897.

Jones RH (1976), Autoregressive order selection, Geophy. 41, 771-773.

Joreskog K. G. (1992), Applications of some new fit measures for structural equation models, Paper presented at the Intnl. Conf. on Social Science Methodology in Trento. Italy, June 22-26, 1992.

Kalman R.E.K. (1960), A New Approach to Linear Filtering and Prediction Problems, Transactions of ASME, Journal of Basic Engineering, 82D, 35-45.

Kay(1988), *Modern Spectral Estimation: Theory and Applications*, Prentice Hall.

Kitagawa, G. (1981), A Nonstationary Time Series Model and its Fitting by a Recursive Filter, Journal of Time Series Analysis, 2, 103-116.

Kitagawa G. (1983), Changing spectrum estimation, Journal of Sound and Vibration, 89, 433-445.

Kitagawa G. (1988), Numerical Approach to non-Gaussian smoothing and its applications, 20th Interface Symposium Computer Science and Statistics.

Kitagawa G. and Akaike H. (1981), On TIMSAC-78 in D. F. Findley ed. *Applied Time Series Analysis II*, Academic Press, 499-547.

Kitagawa G. and Gersch W. (1984), A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality, Journal of the American Statistical Association, 79, 378-389.

Kitagawa G. and Gersch W. (1985a), A Smoothness Priors Long AR Model Method for Spectral Estimation, IEEE Transactions on Automatic Control, AC-30, 57-65.

Kitagawa G. and Gersch W. (1985b), A Smoothness Priors Time Varying AR Coefficient Modeling of Nonstationary Time Series, IEEE Transactions on Automatic Control, AC-30, 48-56.

Koopmans L.H. (1974), *The Spectral Analysis of Time Series*, Academic Press, New York.

Kotz S and Johnson N. L. (1992) editors, *Breakthroughs in Statistics, Vol. 1, Foundations and Basic Theory*, Springer-Verlag, New York.

Kullback S. and Leibler R. A. (1951), On information and sufficiency, Ann. Math. Stat., 22, 79-86.

Kullback S. (1959), *Information Theory and Statistics*, Wiley, New York.

Leamer E. E. (1983), Model choice and specification analysis, in: *Handbook of Econometrics Vol I*, Editors Z. Griliches and M. D. Intriligator, North Holland, 286-330.

Lee D.T.L., Morf M. and Friedlander B. (1981), Recursive least squares ladder estimation algorithms, IEEE Transactions Acoust., Speech, Signal Processing, ASSP-29, 627-641.

Lewis R and Reinsel G. C. (1985), Prediction of multivariate time series by autoregressive model fitting, Journal of Multivariate Analysis 16, 393-411.

Lev-Ari H. (1987), Modular architectures for adaptive multichannel lattice algorithms, IEEE Trans. Acoust. Speech & Signal Process., ASSP-35, 543-552.

Levinson N. (1947), The Wiener RMS error criterion in filter design and prediction, J. Math and Physics, 24, 261-278.

Lindley D.V. and Smith A.F.M. (1972), Bayes estimate for the linear model, Journal of the Royal Statistical Society, B, 34, 1-41.

Ling F and Proakis, (1984), A generalized multichannel least squares lattice algorithm based on sequential processing stages, IEEE Trans. Acoust. Speech & Signal Process., ASSP-32, 381-390.

Linhart H. and Zucchini W. (1986), *Model Selection*, Wiley, New York.

Ljung L. (1991), An exact formula for the tracking ability of adaptive algorithms: Implications for the general case, *Proc. 9th IFAC/IFORS Symp. on Identification and System Parameter estimation*, Budapest, Hungary.

Lutkepohl H. (1991), *Introduction To Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Mallows C.L. (1973) Some comments on the C_p, Technometrics 15, 661-675.

Mann, H. B. and Wald, A. (1943), On the statistical treatment of linear stochastic differential equations, Econometrica, 11, 173-179

Marple (1987), *Digital Spectral Analysis*, Prentice Hall.

Martin W. and Flandrin P. (1985), Wigner-Ville spectral analysis of nonstationary processes, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-33, 1461-1476.

Morf M., Viera A., and Kailath T. (1978a), Covariance characterization by partial autocorrelation matrices, Annals of Statistics, 6, 643-648.

Morf M., Viera A., Lee D.T.L. and Kailath T. (1978b), Recursive multichannel maximum entropy spectral estimation, IEEE Trans. Geosci. Electron., GE-16, 85-94.

Nakamura T.(1986), Bayesian Cohort Models for General Cohort Table Analysis, Annals of the Institute of Statistical Mathematics, 38, Part B, 353-370.

Newton J.J. (1982), Using periodic autoregressions for multiple spectral estimation, Technometrics, 24, 109-116.

O'Sullivan F., Yandell B.S. and Raynor W.J. Jr. (1986), Automatic Smoothing of Regression Functions in Generalized Linear Models, Journal of the American Statistical Association, 81,96-103.

Otomo T. , Nakagawa T and Akaike H (1972), Statistical approach to computer control of cement rotary kilns, Automatica, 8, 35-48.

Pagano M. (1978), Periodic and multiple autoregressions, Annals of Statistics 6, 1310-1317.

Page C H (1952), Instantaneous power spectra, J. Appl. Phys., 23, 103-106.

Porat B., Friedlander B. and Morf M. (1982), Square-root covariance ladder estimation algorithms, IEEE Transactions Automat. Control, AC-17, 813-829.

Priestley M.B. (1965), Evolutionary spectra and nonstationary processes, J. Royal Stat.Soc. b, 27, 204-237.

Priestley MB (1981) Spectral Analysis and Time Series, 2 vols. Academic Press, London.

Quenouille, M. H. (1957), The Analysis of Multiple Time Series, Charles Griffin and Co, Lmtd., London.

Quenouille, M. H. (1947), A large sample test for the goodness of fit of autoregressive schemes, J. Royal Statist. Society B, 110,123-129.

Sakai H. (1982), Circular lattice filtering using Pagano's method; IEEE Trans. Autom. Control, AC-19, 716-723.

Sakamoto T., Ishiguro M. and Kitagawa G.(1986), *Akaike Information Criterion Statistics*, D. Reidel, Holland.

Sambamboorthi N (1989), Information theoretic criterion approach to dimensionality reduction in multinomial regression models, Commun. in Stat. Theory and Methodology 18, 2087-2106.

Schwartz G. (1978), Estimating the dimension of a model, Ann. of Statistics 6, 461-464.

Shibata R (1976), Selection of the order of an autoregressive model by Akaike's information criterion, Biometrika, 63, 117-126.

Shibata R (1980), Asymptotically efficient selection of the order of the model for estimating the parameters of a linear process, Annals of Statistics, 8 147-164.

Shibata R (1981), An optimal selection of regression variables, Biometrika, 68, 47-54.

Shibata R (1982), Asymptotic mean efficiency of aselecation procedure for the number of regression variables, Biometrika, 71, 43-49.

Shiller R. (1973), A distributed lag estimator derived from smoothness priors, Econometrica, 41,775-778.

Sugira N. (1978) Further analysis of the data by Akaike's information criterion and the finite corrections, Commun. in Stat. Theory A7 13-26.

Shuster, A. (1998), On the investigation of hidden periodicities with an application to a supposed 26-day period of meteorological phenomena, Terr. Magn. 3, 13-41.1

Takane Y., Bozdigan H. and Shibayama T. (1987), Ideal point discriminant analysis, Psychometrika 52, 371-392.

Tanabe K. , Sagae M. and Ueda S. (1988), BNDE, Fortran subroutines for computing Bayesian nonparametric univariate and bivariate density estimator, Computer Science Monograph #24, Institute of Statistical Mathematics, Tokyo, Japan.

Terasvirta T. (1987), Smoothness in regression: Asymptotic considerations, In Time Series and Econometric Modeling, editors I. B. MacNeil, G.J. Umphrey, Reidel, Holland, 47-64.

Tikhonov A. N. (1965), Incorrect Problems of Linear Algebra and a Stable Method for their Solution. Soviet Math. Dokl. 6, 988-991.

Taio G.C. and Tsay R.S. (1985), A canonical correlation approach to modeling multivariate time series, Proceedings Econ. Bus. Stat. Sec. Amer. Stat. Assoc., 112-120.

Taio G.C. and Tsay R.S. (1989), Model specification in multivariate time series, Jour. Royal Statistical Society, (with discussion) 157-213

Thisted R. A. (1988), *Elements of Statistical Computing*, Chapman and Hall, New York.

Ville J. (1948), Therie et applications de la notion de signal analytique, Cables et Transmission, 2, 61-74.

Wahba G. (1977), A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them, in: Applications of Statistics, ed P.R. Krishnaih, North Holland, 507-524.

Wahba G. (1990), *Spline Models for Observational Data*, Society of Industrial and Applied Mathematics, Philadelphia.

Waldmeier M (1961), The sunspot activity in the years 1610-1960, Schultess & Co. Zurich.

Walker, A.M. (1952), Some properties of the asymptotic power functions of goodness-of-fit tests for linear autoregressive schemes, J. Royal Statist. Society B, 14, 117-

Weiner N. (1949), *The Extrapolation, Interpolation and Smoothing of Stationary Time-Series with Engineering Applications*, M.I.T. Press, Cambridge.

Whittaker E.T. (1923), On a new method of graduation, Proceedings of the Edinborough Mathematical Association, 78, 81-89.

Whittaker E.T. and Robinson G. (1924), *Calculus of Observations, A Treasure on Numerical Calculations*, Blackie and Son, Lmtd., London. pp 303-306.

Whittle P. (1963), On the fitting of multivariable autoregressions and the approximate canonical factorization of a spectral density matrix, Biometrika 50, 129-134.

Wiggins R.A. and Robinson E.A. (1965), Recursive solution to the multichannel filtering problem, Journal of Geophy. Research, 70, 1885-1891.

Yule g. U. (1927), On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, Philos. Trans. Roy. Soc. London A, 226, 267-298.