

## MULTIVARIATE TIME SERIES ANALYSIS WITH STATE SPACE MODELS†

S. MITTNIK

Department of Economics, State University of New York at Stony Brook,  
Stony Brook, NY 11794-4384, U.S.A.

**Abstract**—The paper proposes a method for estimating linear, time-invariant state space models from multiple time series data. The approach is based on stochastic realization theory. The coefficient matrices of the state space model are derived from the estimated Markov parameters that are associated with the different system inputs, such as lagged endogenous variables, observable exogenous variables, and unobservable noise.

### 1. INTRODUCTION

The use of state space representations for modeling time series has been suggested frequently over recent years (see, for example, Akaike [1], Aoki [2-4], Hannan [5], Kitagawa and Gersch [6], Mittnik [7], Otter [8]). The canonical correlation method proposed by Akaike (for detailed expositions see Akaike [9, 10]) has the disadvantage that it may lead to different, nonequivalent state space representations when changing the order of the variables contained in the vector of endogenous variables. Aoki [3, 4] suggests an alternative method which involves the singular value decomposition of a block Hankel matrix whose entries consist of estimated autocovariances. Employing system-theoretic model reduction techniques, the system matrices are obtained from a lower rank approximation of the Hankel matrix. Aoki's approach has the advantages that lower dimensional state space representations are strictly nested in higher dimensional ones and that they are not affected by the ordering of the variables. Using this method Aoki and Havenner [11, 12] obtain encouraging results in modeling applications.

In this paper we propose an alternative approach to deriving state space models from multiple time series data. The approach is related to Aoki's [4] method and shares the advantages of nestedness and that the resulting models are not affected by the ordering of the endogenous variables, but has, in addition, the attractive feature that it allows for exogenous variables as system inputs. Treating all variables as endogenous, Aoki [4] estimates the system matrices of the state space representation from sample autocovariances. The procedure suggested here is based on the estimated Markov parameters that correspond to the various types of inputs, such as lagged endogenous variables, observable exogenous variables, or unobservable disturbances. It has computational advantages, in particular, when dealing with purely autoregressive processes, autoregressive processes with exogenous variables, or autoregressive approximations of mixed autoregressive moving average processes.

The paper is organized in five sections. The next, Section 2, describes the basic version of the method, assuming only lagged endogenous variables as system inputs. Section 3 outlines the necessary modifications for the approach described in Section 2 when exogenous variables are present. In Section 4 we present extensions of the basic algorithm, enabling us to obtain state space models when moving average terms are present. The final section contains concluding remarks.

---

†Presented at the 1988 Annual Meeting of the Society for Economic Dynamics and Control, Tempe, Arizona. I would like to thank Frank Diebold for his helpful comments.

## 2. ESTIMATION OF SYSTEM MATRICES FOR AUTOREGRESSIVE PROCESSES

Given a sample of  $N$  realizations of a mean-zero, weakly stationary stochastic process  $\{y_t\}$ ,  $y_t \in \mathbb{R}^m$ , our objective is to derive a linear, time-invariant, state space model in innovations representation,

$$z_{t+1} = Az_t + K\epsilon_t, \quad (1a)$$

$$y_t = Cz_t + \epsilon_t, \quad (1b)$$

that approximates the observed sequence. System (1) is such that  $A \in \mathbb{R}^{n \times n}$ ,  $C \in \mathbb{R}^{m \times n}$ ,  $K \in \mathbb{R}^{n \times m}$ ,  $\{\epsilon_t\}$  is a white noise disturbance sequence, and  $\{z_t\}$  represents the sequence of state vectors. Finding the triple  $(A, K, C)$  is known as the stochastic realization problem in the systems literature. To introduce the basic approach to obtaining estimates of  $(A, K, C)$ , we assume in this section the simplest scenario, namely that  $\{y_t\}$  is generated (or can reasonably well be approximated) by the autoregressive (AR) process

$$y_t = \sum_{i=1}^p M_i y_{t-i} + \epsilon_t, \quad (2)$$

where  $\{\epsilon_t\}$  is assumed to be white noise with  $E(\epsilon_t) = 0$  and  $E(\epsilon_t \epsilon_s^T) = \delta_{ts} \Sigma$ . Given the sample  $y_1, y_2, \dots, y_N$ , least squares estimates of the AR coefficient matrices,  $\hat{M}_1, \hat{M}_2, \dots, \hat{M}_k$ , are obtained by

$$\hat{\Theta} = (X^T X)^{-1} X^T Y, \quad (3)$$

where

$$\Theta = [M_1 \quad M_2 \quad \cdots \quad M_k]^T,$$

$$Y = [y_{k+1} \quad y_{k+2} \quad \cdots \quad y_N]^T,$$

$$X^T = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{N-1} \\ y_{k-1} & y_k & & y_{N-2} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ y_1 & y_2 & \cdots & y_{N-k} \end{bmatrix}.$$

Making use of some order selection criterion, integer  $k$  is assumed to be large enough such that  $k \geq p$ .

To derive the state space representation in innovations form (1), we first construct a system of the form

$$z_{t+1} = Fz_t + Ky_t, \quad (4a)$$

$$y_t = Cz_t + \epsilon_t. \quad (4b)$$

Substituting (4b) into (4a) shows that the triple  $(A, K, C)$  associated with the innovations form (1) is obtained from  $(F, K, C)$  by

$$A = F + KC. \quad (5)$$

Representation (4) represents a system whose inputs consist of the system's lagged outputs. The AR coefficients  $M_i$  can then be interpreted as the impulse responses or Markov parameters of system (4). From system theory we know that the Markov parameters of (4) are given by

$$M_i = CF^{i-1}K, \quad i = 1, 2, \dots \quad (6)$$

An estimate of triple  $(F, K, C)$  is derived by constructing the following block Hankel matrix from the estimated parameters  $\hat{M}_t$ ,

$$\hat{H}_k = \begin{bmatrix} \hat{M}_1 & \hat{M}_2 & \cdots & \hat{M}_k \\ \hat{M}_2 & \hat{M}_3 & & \mathbf{0} \\ \vdots & & & \vdots \\ \hat{M}_k & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Matrix  $\hat{H}_k$  can be factorized nonuniquely such that

$$\hat{H}_k = \hat{O}_k \hat{\mathcal{R}}_k. \tag{7}$$

Matrices  $\hat{O}_k$  and  $\hat{\mathcal{R}}_k$  are called observability and reachability matrices, respectively, in the systems literature. It follows from (6) that

$$\hat{O}_k = \begin{bmatrix} C \\ CF \\ \vdots \\ CF^{k-1} \end{bmatrix} \tag{8}$$

and

$$\hat{\mathcal{R}}_k = [K \quad FK \quad \cdots \quad F^{k-1}K]. \tag{9}$$

As is suggested in the context of approximate linear realization theory (see, for example, Zeiger and McEwen [13], Kung [14], Moore [15], Pernebo and Silverman [16]) and used in [4], the singular value decomposition technique can be employed to perform factorization (7), i.e.

$$\hat{H}_k = \hat{U}\hat{Q}\hat{V}^T, \tag{10}$$

where  $\hat{U}^T\hat{U} = \hat{V}^T\hat{V} = I$ , and  $\hat{Q}$  is a diagonal matrix containing the singular values of  $\hat{H}_k$  ordered in a nonincreasing fashion,  $\hat{q}_1 \geq \hat{q}_2 \geq \cdots \geq \hat{q}_{km} \geq 0$ . Defining  $\hat{O}_k$  and  $\hat{\mathcal{R}}_k$  by

$$\hat{O}_k = \hat{U}\hat{Q}^{1/2} \tag{11}$$

$$\hat{\mathcal{R}}_k = \hat{Q}^{1/2}\hat{V}^T \tag{12}$$

yields an ‘‘internally balanced’’ state space representation. The advantages of this particular type of representation will be discussed later.

In a nonstochastic setting, and assuming  $k \geq p$ , the dimension of the system to the realized is given by  $\text{rank}(\hat{H}_k) = n \leq mp$ , i.e.  $Q = \text{diag}(q_1, q_2, \dots, q_n, 0, \dots, 0)$ . In the presence of noise, we expect that  $q_i > 0$  for  $i = n + 1, \dots, mp$ , which prevents us from detecting the exact dimension of the underlying process. When determining the dimension of the state space by the rank of  $\hat{H}_k$ , the resulting model is likely to suffer from overparameterization. Model reduction techniques allow us to separate the overparameterized state space model into a ‘‘dominant’’ and a ‘‘weak’’ subsystem, which can be conceptualized by rewriting the system in terms of partitioned matrices:

$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} = \begin{bmatrix} \hat{F}_{11} & \hat{F}_{12} \\ \hat{F}_{21} & \hat{F}_{22} \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \end{bmatrix} \epsilon_t$$

$$y_t = [\hat{C}_1 \quad \hat{C}_2] \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \epsilon_t.$$

In practice, we may assume that singular values close to zero correspond to the weak subsystem of (4), given by  $(\hat{F}_{22}, \hat{K}_2, \hat{C}_2)$ , which contributes little to the response of the system and may be due

to the presence of noise. The dominant subsystem  $(\hat{\mathbf{F}}_{11}, \hat{\mathbf{K}}_1, \hat{\mathbf{C}}_1)$  that is associated with the  $n$  largest singular values is taken as the model representing the underlying process. Statistical criteria of the types suggested in Otter [8] and Holt and Antill [17] may be employed to estimate the dimension of the system. Other approaches to specifying  $n$  are discussed in [11] and [12]. The lower rank approximation of  $\hat{\mathbf{H}}_k$  that corresponds to the dominant subsystem is derived by eliminating the last  $km - n$  singular values in  $\hat{\mathbf{Q}}$  as well as the corresponding columns of  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$ . Let the modified matrices be denoted by  $\tilde{\mathbf{Q}}$ ,  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$ , respectively. A rank  $n$  approximation of  $\hat{\mathbf{K}}_k$ , denoted by  $\hat{\mathbf{H}}$ , is then obtained by

$$\hat{\mathbf{H}} = \tilde{\mathbf{U}}\tilde{\mathbf{Q}}\tilde{\mathbf{V}}^T. \quad (13)$$

Note that  $\hat{\mathbf{H}}$  will be only approximately Hankel. The lower rank approximations of the reachability and observability matrix are given by

$$\hat{\mathcal{R}} = \tilde{\mathbf{U}}\tilde{\mathbf{Q}}^{1/2}, \quad (14)$$

$$\hat{\mathcal{O}} = \tilde{\mathbf{Q}}^{1/2}\tilde{\mathbf{V}}^T. \quad (15)$$

Using (8), an estimate of matrix  $\mathbf{C}$  is given by the first  $m$  rows of  $\hat{\mathcal{O}}$ , and from (9) it follows that  $\hat{\mathbf{K}}$  consists of the first  $m$  columns of  $\hat{\mathcal{R}}$ . An estimate of the transition matrix  $\mathbf{F}$  of system (4) is given by

$$\hat{\mathbf{F}} = \tilde{\mathbf{Q}}^{-1/2}\tilde{\mathbf{U}}^T\hat{\mathbf{H}}_k^{\dagger}\tilde{\mathbf{V}}\tilde{\mathbf{Q}}^{-1/2}, \quad (16)$$

where  $\hat{\mathbf{H}}_k^{\dagger}$  denotes the shifted Hankel matrix,  $\hat{\mathbf{H}}_k^{\dagger} = L^{-1}\mathbf{H}_k$ , with  $L$  denoting the lag operator. The estimate of the transition matrix of the innovations representation (1) is computed by

$$\hat{\mathbf{A}} = \hat{\mathbf{F}} + \hat{\mathbf{K}}\hat{\mathbf{C}}. \quad (17)$$

The fact that lower dimensional approximations are nested in higher dimensional ones is an attractive feature of Aoki's method. Given an  $n_1$ -dimensional state space model of the form (1), any lower dimensional approximation say of dimension  $n_2 < n_1$ , can be derived by deleting the last  $n_1 - n_2$  rows and columns of the transition matrix  $\hat{\mathbf{A}}$  and the last  $n_1 - n_2$  columns of matrix  $\hat{\mathbf{C}}$ . Since the Hankel matrix factorization in Aoki's method does not really yield the reachability matrix of system (1), an algebraic Riccati equation has to be solved to determine  $\hat{\mathbf{K}}$ . This implies that lower dimensional approximations of  $\hat{\mathbf{K}}$  cannot be obtained by deleting the last  $n_1 - n_2$  rows of  $\hat{\mathbf{K}}$ . Moreover, for each different value of  $n_2$  a different algebraic Riccati equation has to be solved. The nestedness of lower rank realizations is a result of the fact that Aoki's method yields an internally balanced system. A system is said to be internally balanced if its observability and reachability grammians are equal and diagonal (see Moore [15] and Pernebo and Silverman [16] for detailed discussions of balanced realizations). The approach proposed here yields an internally balanced realization for representation (2) and not for the innovations form (1). Given triple  $(\hat{\mathbf{F}}, \hat{\mathbf{K}}, \hat{\mathbf{C}})$  lower dimensional approximations are derived by deleting the corresponding rows and columns of matrices  $\hat{\mathbf{F}}$ ,  $\hat{\mathbf{K}}$  and  $\hat{\mathbf{C}}$ , respectively. But since

$$\begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{F}}_{11} + \hat{\mathbf{K}}_1\hat{\mathbf{C}}_1 & \hat{\mathbf{F}}_{12} + \hat{\mathbf{K}}_1\hat{\mathbf{C}}_2 \\ \hat{\mathbf{F}}_{21} + \hat{\mathbf{K}}_2\hat{\mathbf{C}}_1 & \hat{\mathbf{F}}_{22} + \hat{\mathbf{K}}_2\hat{\mathbf{C}}_2 \end{bmatrix},$$

lower dimensional approximations of system (1) are also obtained by deleting the appropriate columns and rows of matrices  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{K}}$  and  $\hat{\mathbf{C}}$ . Hence, both representation (1) and (4) possess the nestedness property, and the derivation of lower dimensional approximations does not involve an additional computational burden for either representation.

To generate, for example, simulations or predictions from a state space model, it may be necessary to know the initial state vector  $\mathbf{z}_1$ . Given sample  $y_1, y_2, \dots, y_N$  and matrices  $\mathbf{A}$  and  $\mathbf{C}$ ,

a smoothed estimate of  $\mathbf{z}_1$  can be obtained using the following relationship:

$$\mathbf{Y}^* = \mathcal{O}_N \mathbf{z}_1 + \mathbf{S}\mathbf{E}, \tag{18}$$

where

$$\mathbf{Y}^* = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_N^T]^T$$

$$\mathbf{E} = [\boldsymbol{\epsilon}_1^T \quad \boldsymbol{\epsilon}_2^T \quad \cdots \quad \boldsymbol{\epsilon}_N^T]^T,$$

$\mathcal{O}_N$  is the observability matrix as defined in (8), and  $\mathbf{S}$  is the lower triangular Toeplitz matrix

$$\mathbf{S} = \begin{bmatrix} \mathbf{M}_0 & & & & \mathbf{0} \\ \mathbf{M}_1 & \mathbf{M}_0 & & & \\ \cdot & \cdot & \cdot & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \mathbf{M}_{N-1} & \mathbf{M}_{N-2} & \cdots & \mathbf{M}_0 & \end{bmatrix}$$

with  $\mathbf{M}_0 = \mathbf{I}$ . Then, a least squares estimator of  $\mathbf{z}_1$  is given by

$$\hat{\mathbf{z}}_1 = (\hat{\mathcal{O}}_N^T \hat{\mathcal{O}}_N)^{-1} \hat{\mathcal{O}}_N^T \mathbf{Y}^*, \tag{19}$$

where  $\hat{\mathcal{O}}_N$  denotes the observability matrix whose elements are computed from the estimated system matrices.

Rather than using this two-step procedure, i.e. estimating first the system matrices and subsequently the initial state vector, both can be estimated simultaneously. Given an AR process of order  $p$ , the initial state vector represents the effect of the presample realization  $\mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}$ . Let the vector  $\mathbf{s}_k, k = 1, 2, \dots, p$ , capture the effect of the initial state on  $\mathbf{y}_k$ , i.e.

$$\mathbf{s}_k = \sum_{j=k}^p \mathbf{M}_j \mathbf{y}_{k-j} = \mathbf{C}\mathbf{F}^{k-1} \mathbf{z}_1, \quad k = 1, \dots, p. \tag{20}$$

The ‘‘modified Hankel matrix’’

$$\mathbf{H}^e = \begin{bmatrix} \mathbf{s}_1 & \mathbf{M}_1 & \mathbf{M}_2 & \cdots & \mathbf{M}_p \\ \mathbf{s}_2 & \mathbf{M}_2 & \mathbf{M}_3 & & \mathbf{0} \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \mathbf{s}_p & \mathbf{M}_p & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

leads to the following factorization:

$$\mathbf{H}^e = \mathcal{O}_p \mathcal{R}_p^e, \tag{21}$$

where  $\mathcal{O}_p$  is the usual observability matrix and

$$\mathcal{R}_p^e = [\mathbf{z}_1 \quad \mathbf{K} \quad \mathbf{F}\mathbf{K} \quad \cdots \quad \mathbf{F}^{p-1}\mathbf{K}]. \tag{22}$$

Given the estimates  $\hat{\mathcal{O}}_p$  and  $\hat{\mathcal{R}}_p^e$ , the simultaneous derivation of the system matrices and the initial state vector requires only a minor change in the procedure described above. As is clear from (22), the first column of  $\mathcal{R}_p^e$  gives the initial state and columns 2 through  $m + 1$  represent matrix  $\mathbf{K}$ . Matrices  $\hat{\mathbf{C}}$  and  $\hat{\mathbf{F}}$  are derived as before.

To compute least squares estimates of  $\mathbf{s}_i$  and  $\mathbf{M}_i$ , define

$$\Theta^* = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots \quad \mathbf{s}_p \quad \mathbf{M}_1 \quad \mathbf{M}_2 \quad \cdots \quad \mathbf{M}_p]^T,$$

$$\mathbf{Y} = [y_1 \quad y_2 \quad \cdots \quad y_N]^T,$$

$$\mathbf{X}^{*T} = \left[ \begin{array}{c|cccc} \mathbf{I}_p & & & & \mathbf{0} \\ \hline \mathbf{0} & y_1 & y_2 & \cdots & y_p & \cdots & y_{N-1} \\ \mathbf{0} & \mathbf{0} & y_1 & & y_{p-1} & & y_{N-2} \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \mathbf{0} & & & & y_1 & & y_{N-p} \end{array} \right].$$

Then,

$$\hat{\Theta}^* = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^*. \quad (23)$$

Factorizing  $\mathbf{H}^c$  via singular value decomposition and specifying  $\hat{\mathcal{O}}_p$  and  $\hat{\mathcal{R}}_p^c$  by (11) and (12) implies that neither of the representations (1) and (4) will be internally balanced. A balanced realization for system (4) is derived as follows. Construct

$$\hat{\mathbf{H}}_p = \begin{bmatrix} \hat{\mathbf{M}}_1 & \hat{\mathbf{M}}_2 & \cdots & \hat{\mathbf{M}}_p \\ \hat{\mathbf{M}}_2 & \hat{\mathbf{M}}_3 & & \mathbf{0} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \hat{\mathbf{M}}_p & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

from  $\hat{\Theta}^*$  and factorize  $\hat{\mathbf{H}}_p = \hat{\mathcal{O}}_p \hat{\mathcal{R}}_p$ . Defining  $\hat{\mathbf{s}} = [\hat{\mathbf{s}}_1^T \quad \hat{\mathbf{s}}_2^T \quad \cdots \quad \hat{\mathbf{s}}_p^T]^T$ ,

$$\hat{\mathbf{z}}_1 = \tilde{\mathbf{Q}}^{-1/2} \tilde{\mathbf{U}}^T \hat{\mathbf{s}} \quad (24)$$

provides an estimate of the initial state vector with representation (4) being balanced.

### 3. AUTOREGRESSIVE PROCESSES WITH EXOGENOUS VARIABLES

In the presence of exogenous variables, the state space representation in innovations form is given by

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{x}_t + \mathbf{K}\epsilon_t \quad (25a)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{z}_t + \mathbf{D}\mathbf{x}_t + \epsilon_t, \quad (25b)$$

where  $\mathbf{A}$ ,  $\mathbf{K}$ ,  $\mathbf{C}$  are as specified in Section 2,  $\mathbf{B} \in \mathbb{R}^{n \times r}$ ,  $\mathbf{D} \in \mathbb{R}^{m \times r}$ , and  $\{\mathbf{x}_t\}$  is assumed to be an  $r$ -dimensional deterministic sequence of exogenous variables. The procedure described in Section 2 can easily be modified to handle exogenous variables. Given the system matrices of the representation

$$\mathbf{z}_{t+1} = \mathbf{F}\mathbf{z}_t + \mathbf{G}\mathbf{x}_t + \mathbf{K}\mathbf{y}_t \quad (26a)$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{z}_t + \mathbf{D}\mathbf{x}_t + \epsilon_t, \quad (26b)$$

the transition matrix of representation (25) is computed by (17) and matrix  $\mathbf{B}$  is given by

$$\mathbf{B} = \mathbf{G} + \mathbf{K}\mathbf{D}. \quad (27)$$

The estimation of the Markov parameters of system (26) is analogous to (2). Defining

$$\Theta = [M_1 \ M_2 \ \cdots \ M_k \ M_0^x \ M_1^x \ \cdots \ M_{k_x}^x]^T,$$

$$Y = [y_{l+1} \ y_{l+2} \ \cdots \ y_N]^T$$

$$X^T = \begin{bmatrix} y_l & y_{l+1} & \cdots & y_{N-1} \\ y_{l-1} & y_l & & y_{N-2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{l-k+1} & y_{l-k+2} & & y_{N-l+k} \\ x_{l+1} & x_{l+2} & & x_N \\ x_l & x_{l+1} & & x_{N-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{l-k_x+1} & x_{l-k_x+2} & \cdots & x_{N-l+k_x} \end{bmatrix},$$

we obtain the least squares estimates of the coefficient matrices by

$$\hat{\Theta} = (X^T X)^{-1} X^T Y, \tag{28}$$

where  $l = \max(k, k_x)$ . Parameter  $k_x$  represents the lag length considered for the exogenous variables. As with parameter  $k$ , a suitable value for  $k_x$  can be specified using some order selection criterion. Defining  $\hat{N}_i = [\hat{M}_i \ \hat{M}_i^x]$ ,  $i = 1, 2, \dots, l$ , where  $\hat{M}_i = \mathbf{0}$  for  $i > k$  or  $\hat{M}_i^x = \mathbf{0}$  for  $i > k_x$ , allows us to construct the block Hankel matrix

$$\hat{H}_l = \begin{bmatrix} \hat{N}_1 & \hat{N}_2 & \cdots & \hat{N}_l \\ \hat{N}_2 & \hat{N}_3 & & \mathbf{0} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \hat{N}_l & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

The factorization of  $\hat{H}_l$  yields

$$\hat{H}_l = \hat{O}_l \hat{\mathcal{R}}_l. \tag{29}$$

The Markov parameters of system (26) that correspond to "input"  $y$ , are given by (6), and the ones that correspond to  $x$ , are of the form

$$M_i^x = \begin{cases} D, & i = 0 \\ CF^{i-1}G, & i = 1, 2, \dots \end{cases} \tag{30}$$

Therefore, the observability and reachability matrices in (29) can be written as

$$\mathcal{O}_l = \begin{bmatrix} C \\ CF \\ \cdot \\ \cdot \\ \cdot \\ CF^{l-1} \end{bmatrix} \tag{31}$$

and

$$\mathcal{R}_l = [K \ G \ FK \ FG \ \cdots \ F^{l-1}K \ F^{l-1}G], \tag{32}$$

respectively. Analogous to Section 2, given  $\hat{\mathbf{H}}_l$ , the chosen lower rank approximation of  $\hat{\mathbf{H}}_l$ , matrix  $\hat{\mathbf{C}}$  consists of the first  $m$  rows of  $\hat{\mathbf{C}}_l$ ;  $\hat{\mathbf{K}}$  is given by the first  $m$  columns of  $\hat{\mathbf{A}}_l$ ; and  $\hat{\mathbf{G}}$  consists of the columns  $m + 1$  through  $m + r$  in  $\hat{\mathbf{A}}_l$ . The transition matrix  $\hat{\mathbf{F}}$  is computed as in (16). The matrices  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$  of the innovation representation (1) are obtained by (17) and by replacing the matrices on the right-hand side of (27) by their estimates.

#### 4. EXTENSION TO AUTOREGRESSIVE MOVING AVERAGE PROCESSES

In principle, extending the procedure for deriving a state space model for AR processes to autoregressive moving average (ARMA) processes is similar to the modification described in the previous section. Treating the disturbance sequence  $\{\epsilon_t\}$  as a sequence of system inputs, we can include the coefficient matrices associated with the lagged disturbances in the block Hankel matrix as was done with the exogenous variables  $\{x_t\}$ . Therefore, given the estimates of the coefficient matrices, the derivation of the system matrices is as before. The difficulties that arise in the presence of moving average (MA) terms are due to the fact that the disturbances are unobservable. Their coefficients cannot be estimated directly, like those of the observable exogenous variables. Another estimation problem that arises is that the least squares estimation (3) of the AR coefficients will generally be biased. The latter problem can be circumvented by applying instrumental variable (IV) estimators that use lagged  $y_t$  as instruments. The modified Yule-Walker (MYW) method also represents an IV estimator. Among others, Mehra [18] and Cadzow [19, 20] note that the overidentified MYW method usually yields more efficient estimates than the exactly identified version. Stoica *et al.* [21] suggest a multistep MYW procedure that provides estimates whose efficiency approaches the Cramér-Rao lower bound. All these approaches enable us to estimate the AR coefficients without estimating the MA coefficients first. A method for estimating the MA coefficients, given the AR coefficients and the autocovariance function of the process, is described next. It should be pointed out that the method applies also to ARMA processes with deterministic components such as exogenous variables, constant terms and time trends since these components can be removed from the process.

To derive the procedure, we make use of a closed form expression given in Mittnik [22] that relates the parameters of a multivariate ARMA process to its autocovariance function. The exposition is simplified by introducing the following notation. Let  $\mathbf{A} = [\mathbf{A}_1^T \mathbf{A}_2^T \cdots \mathbf{A}_k^T]^T$  be a  $km \times m$  matrix. Then, operator  $H^{\triangleright}[\mathbf{A}]$  generates a  $km \times km$  block Hankel matrix whose first block column is given by  $\mathbf{A}$  and all entries below the main block counterdiagonal are zero. Operator  $H_{\triangleright}[\mathbf{A}]$  generates a  $km \times km$  block Hankel matrix whose last block column is given by  $\mathbf{A}$  and all entries above the main block counterdiagonal are zero. Operator  $T_{\triangleright}[\mathbf{A}]$  generates the  $km \times km$  lower triangular block Toeplitz matrix whose first block column is given by  $\mathbf{A}$ . Finally, operator  $T^{\triangleright}[\mathbf{A}]$  generates a  $km \times km$  upper triangular block Toeplitz matrix whose last block column is given by  $\mathbf{A}$ .

Let an  $m$ -dimensional zero-mean stationary ARMA process of orders  $p$  and  $q$  be given by

$$A(L)y_t = B(L)\epsilon_t, \quad (33)$$

where  $A(L) = \mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p$  and  $B(L) = \mathbf{B}_0 + \mathbf{B}_1 L + \cdots + \mathbf{B}_q L^q$ , with  $\mathbf{A}_i = \mathbf{0}$  for  $i = p + 1, \dots, r$  if  $r = \max(p, q) > p$  or  $\mathbf{B}_i = \mathbf{0}$  for  $i = q + 1, \dots, r$  if  $r > q$ .  $\{\epsilon_t\}$  is a white noise sequence with  $E(\epsilon_t) = 0$  and  $E(\epsilon_t \epsilon_s^T) = \delta_{ts} \mathbf{I}$ . Setting  $\text{var}(\epsilon_t) = \mathbf{I}$  entails no loss in generality since we allow  $\mathbf{B}_0 \neq \mathbf{I}$ . Note that the MA parameters can always be redefined such that  $\mathbf{B}_0 = \mathbf{I}$ . Having  $\mathbf{B}_0 \neq \mathbf{I}$  requires us to premultiply the disturbance terms in the measurement equations of the state space systems by  $\mathbf{B}_0$ . Equations (1b) and (4b), for example, become

$$y_t = \mathbf{C}z_t + \mathbf{B}_0 \epsilon_t.$$



Let  $\Gamma_\tau = E(y_t y_{t-\tau}^T)$ ,  $\tau = 0, 1, \dots$ , denote the autocovariance sequence of process (33) and define the following matrices:

$$\begin{aligned} \Gamma &= [\Gamma_0^T \ \Gamma_1^T \ \dots \ \Gamma_r^T]^T, \\ \Gamma^* &= [\Gamma_0 \ \Gamma_1 \ \dots \ \Gamma_r]^T, \\ \mathbf{A} &= [\mathbf{A}_1^T \ \mathbf{A}_2^T \ \dots \ \mathbf{A}_r^T]^T, \\ \tilde{\mathbf{A}} &= [\mathbf{I} \ -\mathbf{A}_1 \ -\mathbf{A}_2 \ \dots \ -\mathbf{A}_r]^T, \\ \mathbf{B} &= [\mathbf{B}_0^T \ \mathbf{B}_1^T \ \dots \ \mathbf{B}_r^T]^T, \\ \mathbf{B}^* &= [\mathbf{B}_0 \ \mathbf{B}_1 \ \dots \ \mathbf{B}_r]^T, \\ \mathbf{C} &= [\mathbf{C}_0^T \ \mathbf{C}_1^T \ \dots \ \mathbf{C}_r^T]^T, \\ \mathbf{C}^* &= [\mathbf{C}_0 \ \mathbf{C}_1 \ \dots \ \mathbf{C}_r]^T, \end{aligned}$$

where  $\mathbf{C}_i$  denotes the  $i$ th coefficient matrix of the moving average representation of (33), i.e.  $\mathbf{C}(\mathbf{L}) = \mathbf{A}^{-1}(\mathbf{L})\mathbf{B}(\mathbf{L}) = \mathbf{C}_0 + \mathbf{C}_1\mathbf{L} + \dots + \mathbf{A}_s$  as shown in [22], the following relationship can be established:

$$\Gamma = \mathbf{M}_T \Gamma + \mathbf{M}_H \Gamma^* + \mathbf{N} \mathbf{C}^*, \tag{34}$$

where the  $m(r + 1) \times m(r + 1)$  matrices  $\mathbf{M}_T$ ,  $\mathbf{M}_H$ , and  $\mathbf{N}$  are defined by

$$\mathbf{M}_T = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ T_{\triangleright}[\mathbf{A}] & \mathbf{0} \end{bmatrix}, \quad \mathbf{M}_H = \begin{bmatrix} \mathbf{0} & H^{\triangleright}[\mathbf{A}] \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{N} = H^{\triangleright}[\mathbf{B}].$$

Expression (34) relates the autocovariances to the AR coefficients and the coefficients of the moving average representation. By using the fact that

$$\mathbf{C} = (\mathbf{I} - \mathbf{M}_T)^{-1} \mathbf{B} \tag{35}$$

(see [23]) the autocovariances can be related directly to the AR and MA coefficients. By combining (34) and (35), [26] shows that the MA parameters satisfy

$$\mathbf{N} \mathbf{B}^* = H^{\triangleright}[(\mathbf{I} - \mathbf{M}_T)\Gamma - \mathbf{M}_H \Gamma^*] \tilde{\mathbf{A}}. \tag{36}$$

Thus, given the AR coefficients and the autocovariances  $\Gamma_\tau$ ,  $\tau = 0, \dots, r$ , equation (36) provides an implicit expression for the MA coefficients in terms of the AR coefficients and the autocovariances. There is unfortunately no direct way of obtaining matrix  $\mathbf{B}$  given the right-hand side of (36).

The problem of deriving  $\mathbf{B}$  from (36) is related to a procedure for obtaining the MA coefficient estimates for pure MA processes suggested by Jenkins and Alavi [24]. Note that in this case (36) reduces to

$$\mathbf{N} \mathbf{B}^* = \Gamma. \tag{37}$$

Their iterative, linearly convergent procedure for computing  $\mathbf{B}$  from (37) consists of solving recursively the following set of equations:

$$\Sigma = \Gamma_0 - \sum_{i=1}^q \mathbf{B}_i \Sigma \mathbf{B}_i^T \tag{38a}$$

$$\mathbf{B}_k^T = -\Sigma^{-1} \Gamma_k + \Sigma^{-1} \sum_{i=1}^{q-k} \mathbf{B}_i \Sigma \mathbf{B}_{i+k}^T, \quad k = q, q - 1, \dots, 1, \tag{38b}$$

where  $\mathbf{B}_0 = \mathbf{I}$  and  $\Sigma = \text{var}(\epsilon_t)$ . The procedure is initialized by setting  $\mathbf{B}_i = \mathbf{0}$ ,  $i = 1, \dots, q$ . Jenkins and Alavi [24, p. 12] note that “[t]he convergence properties of this method are not satisfactory, especially if  $q > 1$ , and a quadratically convergent method, which applies to mixed models, is now under development.”

The problem of determining  $\mathbf{B}$  from (37) is also related to the spectral factorization problem for which a quadratically convergent procedure is readily available. Tunncliffe Wilson [25] provides

a Newton–Raphson factorization algorithm. It relates successive iterations for  $\mathbf{B}_{\tau+1,k}$  in terms of  $\mathbf{B}_{\tau,k}$  by a set of linear equations given by

$$\sum_{i=0}^{q-k} \mathbf{B}_{\tau+1,j+k} \mathbf{B}_{\tau,j}^T + \mathbf{B}_{\tau,j+k} \mathbf{B}_{\tau+1,j}^T = \Gamma_k + \sum_{j=0}^{q-k} \mathbf{B}_{\tau,j+k} \mathbf{B}_{\tau,j}^T, \quad k = 0, \dots, q, \tag{39}$$

where  $\mathbf{B}_{\tau,k}$  denotes the value for MA coefficient  $\mathbf{B}_k$  in the  $\tau$ th iteration. Using our matrix notation, it can be shown that (39) amounts to the following equation:

$$\mathbf{N}_{\tau+1} \mathbf{B}_{\tau}^* + \mathbf{N}_{\tau} \mathbf{B}_{\tau+1}^* = \Gamma + \mathbf{N}_{\tau} \mathbf{B}_{\tau}^*, \tag{40}$$

with  $\mathbf{N}_{\tau}$  and  $\mathbf{B}_{\tau}^*$  being as defined in (34) and (36), respectively. In case of mixed ARMA processes, equation (40) becomes

$$\mathbf{N}_{\tau+1} \mathbf{B}_{\tau}^* + \mathbf{N}_{\tau} \mathbf{B}_{\tau+1}^* = \Lambda + \mathbf{N}_{\tau} \mathbf{B}_{\tau}^* \tag{41}$$

where

$$\Lambda = (\Lambda_0^T \ \Lambda_1^T \ \dots \ \Lambda_q^T)^T = H^D[(\mathbf{I} - \mathbf{M}_T)\Gamma - \mathbf{M}_H \Gamma^*] \tilde{\Lambda}.$$

Tunncliffe Wilson initializes recursion (39) by setting

$$\mathbf{B}_{0,j} = \begin{cases} \mathbf{T}, & j = 0 \\ \mathbf{0}, & j = 1, \dots, q, \end{cases}$$

where  $\mathbf{T}$  is obtained from the Cholesky decomposition  $\Gamma_0 = \mathbf{T}\mathbf{T}^T$ . Note that one iteration can be saved by initializing (41) with  $\mathbf{B}_{0,\cdot} = \Lambda(\mathbf{H}^T)^{-1}$ , where  $\Lambda_0 = \mathbf{H}\mathbf{H}^T$  and, to avoid confusion,  $\mathbf{B}_{0,\cdot}$  denotes the value of matrix  $\mathbf{B}$  at the initialization. The factorization procedure requires for each iteration the solution of a set of  $m^2(q+1) - m(m-1)/2$  linear equations which can be specified by vectorizing (41). Computationally this can be rather costly, even for moderate values for  $m$  and  $q$ .

An alternative method is proposed in Mittnik [26] and briefly summarized here. It is well known that an  $m$ -dimensional AR process of order  $p$  can be expressed as a first order process of dimension  $mp$ . Using a somewhat different strategy we can transform an  $m$ -dimensional MA process of order  $q$ , given by

$$\mathbf{y}_t = \mathbf{B}(\mathbf{L})\boldsymbol{\epsilon}_t, \tag{42}$$

into a first order MA process of dimension  $mq$ . Defining

$$\begin{aligned} \mathbf{B}_0 &= [\mathbf{B}_0^T \ \mathbf{B}_1^T \ \dots \ \mathbf{B}_{q-1}^T]^T, \\ \mathbf{B}_1 &= [\mathbf{B}_1^T \ \mathbf{B}_2^T \ \dots \ \mathbf{B}_q^T]^T, \\ \mathbf{T}_0 &= T_{\triangleright}[\mathbf{B}_0], \\ \mathbf{T}_1 &= T^{\triangleright}[\mathbf{B}_1], \end{aligned}$$

we can rewrite the MA process (42) as

$$\underline{\mathbf{y}}_{k\Delta} = \mathbf{T}_0 \boldsymbol{\epsilon}_{k\Delta} + \mathbf{T}_1 \boldsymbol{\epsilon}_{k\Delta - \Delta}, \quad k \in \mathbb{Z}, \tag{43}$$

where

$$\begin{aligned} \underline{\mathbf{y}}_{k\Delta} &= [\mathbf{y}_{k-q}^T \ \mathbf{y}_{k-q+1}^T \ \dots \ \mathbf{y}_k^T]^T, \\ \boldsymbol{\epsilon}_{k\Delta} &= [\boldsymbol{\epsilon}_{k-q}^T \ \boldsymbol{\epsilon}_{k-q+1}^T \ \dots \ \boldsymbol{\epsilon}_k^T]^T, \end{aligned}$$

and subscript  $\Delta$  denotes the time interval with which process (43) evolves. For the above definitions of  $\mathbf{B}_0$  and  $\mathbf{B}_1$  and, thus,  $\mathbf{T}_1$  and  $\mathbf{T}_2$ ,  $\Delta$  is equal to  $q$ .

This conversion allows us to obtain an expression for the factorization problem (36) or (37). Let  $\mathbf{D}_0$  and  $\mathbf{D}_1$  denote the variance and first order autocovariance of process (43),  $\mathbf{D}_0 = E(\underline{\mathbf{y}}_{k\Delta} \underline{\mathbf{y}}_{k\Delta}^T)$  and  $\mathbf{D}_1 = E(\underline{\mathbf{y}}_{k\Delta} \underline{\mathbf{y}}_{k\Delta - \Delta}^T)$ , respectively. Note that the autocovariance matrices of process  $\underline{\mathbf{y}}_{k\Delta}$  consist

of the elements of the autocovariance function of the original processes,  $\Gamma_i, i = 0, 1, \dots$ . In particular,

$$\mathbf{D}_0 = \begin{bmatrix} A_0 & A_1^T & \cdots & A_{q-1}^T \\ A_1 & A_0 & & \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ A_{q-1} & \cdots & & A_0 \end{bmatrix},$$

$$\mathbf{D}_1 = \begin{bmatrix} A_q & A_{q-1} & \cdots & A_1 \\ \mathbf{0} & A_q & & A_2 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \mathbf{0} & \cdots & \mathbf{0} & A_q \end{bmatrix}.$$

In case the original process is a pure MA process, the  $A_i$ s in  $\mathbf{D}_0$  and  $\mathbf{D}_1$  are replaced by  $\Gamma_i$ s. Using this notation, the factorization problem can be written as

$$\mathbf{T}_0 \mathbf{T}_0^T + \mathbf{T}_1 \mathbf{T}_1^T = \mathbf{D}_0 \tag{44a}$$

$$\mathbf{T}_1 \mathbf{T}_0^T = \mathbf{D}_1. \tag{44b}$$

Solving (44) amounts to finding the solution to

$$\mathbf{P} + \mathbf{D}_1 \mathbf{P}^{-1} \mathbf{D}_1^T = \mathbf{D}_0, \tag{45}$$

with  $\mathbf{P} = \mathbf{T}_0 \mathbf{T}_0^T$ . Equation (45) is solved iteratively by

$$\mathbf{P}_{\tau+1} = \mathbf{D}_0 - \mathbf{D}_1 \mathbf{P}_\tau^{-1} \mathbf{D}_1^T, \tag{46}$$

with  $\mathbf{P}_0 = \mathbf{D}_0 - \mathbf{D}_1 \mathbf{D}_0^{-1} \mathbf{D}_1^T$ , yielding

$$\mathbf{P} = \lim_{\tau \rightarrow \infty} \mathbf{P}_\tau.$$

Given  $\mathbf{P}$ , we obtain  $\mathbf{T}_0$  via Cholesky decomposition. The first block column of  $\mathbf{T}_0$  represents the MA coefficient matrices  $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{q-1}$  and, rather than solving  $\mathbf{T}_1 = \mathbf{D}_1 (\mathbf{T}_0^T)^{-1}$ ,  $\mathbf{B}_q$  is computed by

$$\mathbf{B}_q = A_q \mathbf{B}_0^{-1}. \tag{47}$$

The procedure just described represents an alternative to the matricial spectral factorization algorithm of Tunnicliffe Wilson [25] which requires in each iteration the solution of an  $m^2(q+1) - m(m-1)/2$ -dimensional system of linear equations. The algorithm suggested here involves the inversion of matrix  $\mathbf{P}_k$  in each iteration. Matrix  $\mathbf{P}_k$  is of dimension  $mq \times mq$  only, and, since it is (block) symmetric positive definite, its inverse can be computed rather efficiently. Note also that  $\mathbf{D}_0$  has a block Toeplitz structure and can be efficiently inverted by applying the block Levinson algorithm suggested by Akaike [27], involving only  $O(q^2)$  operations.

The proposed factorization procedure can be applied to the standard spectral factorization problem, providing us directly with estimates of the coefficients of the moving average representation,  $\mathbf{C}_0, \mathbf{C}_1, \dots$ , without first having to estimate the AR coefficients. The latter can be estimated in a subsequent step by applying one of the IV-type methods mentioned earlier. Then, using (35), the MA coefficients can be determined by

$$\mathbf{B} = (\mathbf{I} - \mathbf{M}_\tau) \mathbf{C}. \tag{48}$$

An advantage of proceeding in this fashion is that the convergence of the iterative factorization is not affected by poorly estimated AR coefficients.

## 5. CONCLUDING REMARKS

A method for estimating linear, time-invariant state space models from multiple time series data has been suggested. The system matrices are derived from estimated Markov parameters associated with the system's inputs. The inputs may consist of lagged outputs, observable exogenous variables, unobservable noise, or any combination of these three types. For pure AR processes, ARMA processes that can be approximated reasonably well by AR processes, and AR processes with exogenous variables, the approach is computationally simpler than the one proposed by Aoki, since it does not require the solution of an algebraic Riccati equation. In the presence of MA terms, the proposed algorithm involves a step that is related to the spectral factorization problem. In fact, we accomplish the factorization by solving (45) which can be regarded as a special and rather simple version of a Riccati equation. The equivalence between Tunnicliffe Wilson's [25] spectral factorization algorithm and finding the solution to an algebraic Riccati equation has been pointed out by Anderson [28]. Hence, the approach by Aoki and the one described here are more closely related than it may first appear. A major computational advantage of the proposed method is that the factorization has to be solved only once, not re-solved each time considering a different lower dimensional approximation. Also, the separate estimation of the AR coefficients allows us to employ more efficient estimators for this step.

It remains to be seen how both approaches compare in practical applications. A comparison of alternative strategies to modeling multivariate time series (such as unrestricted vector autoregressions, Bayesian vector autoregressions, multiple ARMA models obtained via maximum likelihood estimation, and the ones mentioned here) which investigates their performance with respect to various objectives is the subject of future research.

## REFERENCES

1. H. Akaike, Markovian representation of stochastic processes and its applications to the analysis of autoregressive moving average processes. *Ann. Inst. statist. Math.* **20**, 363–388 (1974).
2. M. Aoki, *Optimal Control and System Theory in Dynamic Analysis*. North-Holland, Amsterdam (1976).
3. M. Aoki, *Notes on Economic Time Series Analysis: System Theoretic Perspectives*. Springer, Berlin (1983).
4. M. Aoki, *State Space Modeling of Time Series*. Springer, Berlin (1987).
5. E. J. Hannan, The identification and parameterization of ARMAX and state space forms. *Econometrica* **44**, 713–723 (1976).
6. G. Kitagawa and W. Gersch, A smoothness priors-state space modeling of time series with trend and seasonality. *J. Am. statist. Assoc.* **79**, 378–389 (1984).
7. S. Mittnik, Time series analysis via approximate realization theory. *1984 Winter Meeting Econometric Society*, Dallas, Texas (1984).
8. P. W. Otter, *Dynamic Feature Space Modelling, Filtering and Self-Tuning Control of Stochastic Systems*. Springer, Berlin (1985).
9. H. Akaike, Markovian representation of stochastic processes by canonical variables. *SIAM J. Control Optim.* **13**, 162–173 (1975).
10. H. Akaike, Canonical correlation analysis of time series and the use of an information criterion. In *System Identification and Case Studies* (Eds R. Mehra and D. Lainiotis). Academic Press, New York (1976).
11. M. Aoki and A. Havenner, A method for approximate representation of vector-valued time series, Unpublished manuscript (1985).
12. M. Aoki and A. Havenner, Approximate state space models of some vector-valued macroeconomic time series for cross-country comparisons. *J. econ. Dynam. Control* **10**, 149–155 (1986).
13. H. P. Zeiger and A. J. McEwen, Approximate linear realization of given dimension via Ho's algorithm. *IEEE Trans. autom. Control* **AC19**, 153 (1974).
14. S. Y. Kung, A new identification and model reduction algorithm via singular value decompositions. *Proc. 12th Ann. Asilomar Conf. Circuits, Systems and Computers*, pp. 705–714 (1978).
15. B. C. Moore, Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. autom. Control* **AC26**, 17–32 (1981).
16. L. Pernebo and L. M. Silverman, Model reduction via state space representation. *IEEE Trans. autom. Control* **AC27**, 382–387 (1982).
17. J. N. Holt and R. J. Antill, Determining the number of terms in a Prony algorithm exponential fit. *Math. Biosci.* **36**, 319–332 (1977).
18. R. K. Mehra, On-line identification of linear dynamic systems with applications to Kalman filtering. *IEEE Trans. autom. Control* **AC16**, 12–22 (1971).
19. J. A. Cadzow, High performance spectral estimation—A new ARMA method. *IEEE Trans. Acoust. Speech Signal Processing ASSP28*, 524–529 (1980).
20. J. A. Cadzow, Spectral estimation: an overdetermined rational equation approach. *Proc. IEEE* **70**, 907–939 (1982).
21. P. Stoica, B. Friedlander and T. Söderström, An approximate maximum likelihood approach to ARMA spectral estimation. *Proc. 24th Conf. Decision and Control*, 1276–1281 (1985).

22. S. Mittnik, Alternative methods for computing the theoretical autocovariance function of multivariate ARMA processes: a comparison. In *Computer Science and Statistics: Proc. 20th Symp. The Interface*. American Statistical Association, Alexandria. In press.
23. S. Mittnik, Nonrecursive methods for computing the coefficients of the autoregressive and the moving-average representation of mixed ARMA processes. *Econ. Lett.* **23**, 279–284 (1987).
24. G. M. Jenkins and A. S. Alavi, Some aspects of modelling and forecasting multivariate time series. *J. Time Ser. Analysis* **2**, 1–47 (1981).
25. G. Tunnicliffe Wilson, the factorization of matricial spectral densities. *SIAM J. appl. Math.* **23**, 420–426, (1972).
26. S. Mittnik, Estimating moving average coefficients of multivariate ARMA processes. Unpublished manuscript (1988).
27. H. Akaike, Block Toeplitz matrix inversion. *SIAM J. appl. Math.* **24**, 234–241 (1973).
28. B. D. O. Anderson, Second order convergent algorithms for the steady-state Riccati equation. *Int. J. Control* **28**, 295–306 (1978).