Signal Processing 91 (2011) 955–969

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Order selection criteria for vector autoregressive models

Mahmood Karimi *

Department of Communications and Electronics, School of Electrical and Computer Engineering, Shiraz University, Zand Street, Shiraz, Iran

ARTICLE INFO

Article history:
Received 28 December 2009
Received in revised form
20 July 2010
Accepted 25 September 2010
Available online 1 October 2010

Keywords:
AIC
FPE
Kullback–Leibler index
Autoregressive process
Information criterion

ABSTRACT

The least-squares method for estimating the parameters of the vector autoregressive (VAR) model is considered and new estimates for the covariance matrix of the VAR model input noise and the prediction error covariance matrix are derived. Based on these new estimates, the criteria FPEF and AICF for VAR model order selection are proposed. FPEF can replace the final prediction error (FPE) criterion, and AICF, which is an estimate of the Kullback–Leibler index, can replace the Akaike information criterion (AIC) and its corrected version AICC. A simulation study shows that FPEF is less biased than FPE, and AICF is less biased than AIC and AICC. In addition, the performance of the proposed criteria is compared with that of other well-known criteria and the results show that AICF has the best performance and gives the smallest average prediction error.

1. Introduction

An important but difficult step in vector autoregressive (VAR) modeling is the selection of the order. Several order selection criteria for VAR models have been proposed to facilitate the model building process. These criteria are usually the generalized versions of the order selection criteria developed for univariate (one-dimensional) autoregressive models. Selecting an appropriate VAR model for a sequence of observed data requires two steps, selecting the model order and estimating the coefficients of the model. The most common estimators for the coefficients of a VAR model are Yule–Walker, least-squares, and Nuttal–Strand estimators [1]. All VAR coefficient estimation methods seem to be equivalent in the asymptotic case, while practical experience shows important differences in finite samples [2]. The performance of a number of VAR coefficient estimators is examined and compared with each other in Ref. [3]. In the present paper, we assume that the coefficients are estimated by the least-squares method.

In addition to the above-mentioned VAR parameter estimators, there are estimators for the parameters of VAR signals corrupted with noise. Two such estimators are proposed in Refs. [4,5]. The estimator proposed in Ref. [4] removes the noise-induced estimation bias in the least-squares parameter estimation method to yield an unbiased estimate of the VAR coefficients. The estimator proposed in Ref. [5] is an inverse filtering based method for noisy VAR estimation.

The VAR model which we consider in this paper is a linear model, but there are nonlinear VAR models that are used in various applications. A well-known nonlinear VAR model is the vector or multivariate autoregressive conditional heteroscedasticity (ARCH) model. In a recent paper [6], the performance and asymptotic properties of the two-stage least-squares parameter estimation method for multivariate ARCH model are discussed.

There are several well-known order selection criteria for VAR models. One of the earliest criteria is the final prediction error (FPE) [7], which is an estimator of the trace or determinant of the covariance matrix of the prediction error. Another criterion is the Akaike information criterion (AIC) [8] that was designed to be an asymptotically unbiased estimator of the Kullback–Leibler

* Tel.: +98 711 6133043; fax: +98 711 2303081.
E-mail address: karimi@shirazu.ac.ir

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.09.021

index of the fitted model relative to the true model. It is well-known that FPE and AIC are quite biased in the finite sample case, i.e., in the case where the data sample size N is not large relative to the maximum candidate order. Hurvich and Tsai [9] proposed a corrected Akaike information criterion (AICC) that is less biased in the finite sample case. Cavanaugh [10] proposed the Kullback information criterion (KIC), which was designed to be an asymptotically unbiased estimator of the Kullback's symmetric divergence. Seghouane [11] proposed a corrected KIC (KICC) criterion that is less biased in the finite sample case. Schwarz [12] proposed the Bayesian information criterion (BIC) which is a consistent criterion. All of the mentioned criteria use the order selection approach to VAR model selection. Another approach to VAR model selection is the subset selection approach that selects a subset of all possible parameters to model the multivariate process. Two recently proposed subset selection methods for VAR model selection are given in Refs. [13,14].

The bias in FPE and AIC often leads to severe overfitting in the finite sample case. As will be shown in this paper, even AICC can be quite biased in the finite sample case, leading to overfitting. For univariate autoregressive (AR) model selection based on finite samples, Karimi [15–17] has proposed finite sample FPE (FPEF) and finite sample AIC (AICF) criteria, which are less biased, and as a consequence, tend to select much better models than FPE, AIC, and AICC. Derivation of FPEF and AICF for univariate case is based on the theoretical results derived by Karimi [18]. For vector AR models the overfitting problem is of greater importance than in the univariate case. The reason is that when the model order increases, the parameter count of the VAR model increases more rapidly than the univariate AR model. In this paper, we derive and propose FPEF and AICF criteria for vector AR model selection. These two criteria contain the univariate versions as special cases. The bias and order selection performance of FPEF and AICF are studied and it is shown that these new criteria are less biased and have better order selection performance than FPE, AIC, and AICC. In addition, the order selection performance of the proposed criteria is compared with that of FPE, AIC, AICC, KIC, KICC, and BIC criteria, and the results show that AICF has the best performance and gives the smallest average prediction error.

The remainder of this paper is organized as follows. In Section 2, some sets of vector variables are defined and some relations that will be used in the subsequent sections are derived. In Section 3, formulas for the residual variance and the covariance matrix of the VAR model input noise are derived. In Section 4, formulas for the prediction error and the prediction error covariance matrix are derived. Statistical properties of the parameters used in the formulas derived in Sections 3 and 4 are investigated in Section 5. The steps leading to the derivation of FPEF and AICF criteria are discussed in Sections 6 and 7, respectively. In Section 8, simulation results are presented that evaluate the performance of FPEF and AICF and compare it with that of other existing criteria.

2. Vector autoregressive model and the least-squares estimation

A real m-dimensional vector autoregressive process $\mathbf{x}_n$ with zero mean is given by

$$\mathbf{x}_n = \mathbf{\Phi}_1\mathbf{x}_{n-1} + \ldots + \mathbf{\Phi}_p\mathbf{x}_{n-p} + \mathbf{w}_n \tag{1}$$

where the $\mathbf{\Phi}_i$ are $m \times m$ coefficient matrices and p is the model order. The process $\mathbf{w}_n$ is a zero mean independent identically distributed (i.i.d.) vector process with covariance matrix Σ. It is assumed that $\mathbf{w}_n$ and $\mathbf{x}_{n-k}$ are independent for each $k>0$. In addition, it is assumed that $\mathbf{x}_n$ is mean and covariance ergodic.

Suppose that the observations $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_N$ are generated by the VAR process given by Eq. (1). The least-squares method can be used to fit a VAR model with order q to the observations. Suppose that the estimated model is as follows:

$$\mathbf{x}_n = \hat{\mathbf{\Phi}}_1\mathbf{x}_{n-1} + \ldots + \hat{\mathbf{\Phi}}_q\mathbf{x}_{n-q} + \hat{\mathbf{w}}_n \tag{2}$$

The coefficient matrices of this estimated model are obtained by solving the following set of equations:

$$\mathbf{X}\mathbf{A}_i = \mathbf{x}_{q+1,i}; \quad i = 1,2,\ldots,m \tag{3}$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_q^T & \mathbf{x}_{q-1}^T & \cdots & \mathbf{x}_1^T \\ \mathbf{x}_{q+1}^T & \mathbf{x}_q^T & \cdots & \mathbf{x}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N-1}^T & \mathbf{x}_{N-2}^T & \cdots & \mathbf{x}_{N-q}^T \end{bmatrix}, \quad \mathbf{A}_i = \begin{bmatrix} \mathbf{\Phi}_1^T(i,\ :) \\ \mathbf{\Phi}_2^T(i,\ :) \\ \vdots \\ \mathbf{\Phi}_q^T(i,\ :) \end{bmatrix},$$

$$\mathbf{x}_{q+1,i} = \begin{bmatrix} \mathbf{x}_{q+1}(i) \\ \mathbf{x}_{q+2}(i) \\ \vdots \\ \mathbf{x}_N(i) \end{bmatrix} \tag{4}$$

where $\mathbf{\Phi}_j(i,:)$ is the ith row of $\mathbf{\Phi}_j$. The least-squares solutions of Eq. (3) are given by

$$\hat{\mathbf{A}}_i = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}_{q+1,i}; \quad i = 1,2,\ldots,m \tag{5}$$

We define the m-dimensional vector variables $\mathbf{z}_{n,i}$ as follows:

$$\mathbf{z}_{n,i}(j) = \begin{cases} \dfrac{\mathbf{x}_{n-1}(1)}{\sigma_{1,1}}; & i=1, \quad j=1 \\ \dfrac{\mathbf{x}_{n-i}(j) - \hat{\mathbf{x}}_{MS,n,i,j}}{\sigma_{i,j}}; & (i=1, \quad j=2,\ldots,m) \text{ or } (i>1, \quad j=1,2,\ldots,m) \end{cases} \tag{6}$$

where $\hat{\mathbf{x}}_{MS,n,i,j}$ is the linear minimum mean square error (LMMSE) estimate of $\mathbf{x}_{n-i}(j)$ given $\mathbf{x}_{n-1}$, $\mathbf{x}_{n-2}, \ldots, \mathbf{x}_{n-i+1}$ and $\mathbf{x}_{n-i}(1)$, ..., $\mathbf{x}_{n-i}(j-1)$, and $\sigma_{i,j}$ is the root mean square error of this estimate

$$\sigma_{i,j} = \begin{cases} \{E[\mathbf{x}_{n-1}^2(1)]\}^{1/2}; & i=1, \quad j=1 \\ \{E[(\mathbf{x}_{n-i}(j) - \hat{\mathbf{x}}_{MS,n,i,j})^2]\}^{1/2}; & i \geq 1, \quad j=2,\ldots,m \end{cases} \tag{7}$$

Note that $\sigma_{i,j}$ is independent of n because the process $\mathbf{x}_n$ is stationary. It is obvious that the random variables $\mathbf{z}_{n,i}(j)$ are the results of a Gram–Schmidt orthogonalization. So, $\hat{\mathbf{x}}_{MS,n,i,j}$ is also the LMMSE estimate of $\mathbf{x}_{n-i}(j)$ given $\mathbf{z}_{n,1}$, ..., $\mathbf{z}_{n,i-1}$ and $\mathbf{z}_{n,i}(1)$, ..., $\mathbf{z}_{n,i}(j-1)$. It can be seen from the orthogonality principle that

$$E[\mathbf{z}_{n,i_1}(j_1)\mathbf{z}_{n,i_2}(j_2)]=0;\quad i_1\neq i_2\quad\text{or}\quad j_1\neq j_2 \tag{8}$$

From Eqs. (6) and (7) we can see that

$$E[\mathbf{z}^2_{n,i}(j)]=1;\quad i=1,2,\ldots,\quad j=1,2,\ldots,m \tag{9}$$

It can be seen from Eqs. (8) and (9) that

$$E[\mathbf{z}_{n,i_1}\mathbf{z}^T_{n,i_2}]=\begin{cases}\mathbf{0}; & i_1\neq i_2\\ \mathbf{I}_m; & i_1=i_2\end{cases} \tag{10}$$

where $\mathbf{0}$ is the $m\times m$ zero matrix and $\mathbf{I}_m$ the $m\times m$ identity matrix.

According to the definition of $\hat{\mathbf{x}}_{MS,n,i,j}$, this parameter can be written as the following linear combination:

$$\hat{\mathbf{x}}_{MS,n,i,j}=\mathbf{F}_{i,1}(j,:)\mathbf{x}_{n-1}+\mathbf{F}_{i,2}(j,:)\mathbf{x}_{n-2}+\ldots+\mathbf{F}_{i,i-1}(j,:)\mathbf{x}_{n-i+1}+g_{i,1}\mathbf{x}_{n-i}(1)+\ldots+g_{i,j-1}\mathbf{x}_{n-i}(j-1) \tag{11}$$

where $\mathbf{F}_{i,k}$ is an $m\times m$ matrix and $g_{i,k}$ is a scalar. Using Eqs. (6) and (11), we can also write $\mathbf{z}_{n,i}(j)$ as the following linear combination:

$$\mathbf{z}_{n,i}(j)=\frac{-1}{\sigma_{i,j}}\{\mathbf{F}_{i,1}(j,:)\mathbf{x}_{n-1}+\ldots+\mathbf{F}_{i,i-1}(j,:)\mathbf{x}_{n-i+1}+g_{i,1}\mathbf{x}_{n-i}(1)+\ldots+g_{i,j-1}\mathbf{x}_{n-i}(j-1)-\mathbf{x}_{n-i}(j)\};\quad i\geq 1,\quad j=1,\ldots,m \tag{12}$$

Note that if $i=1$, the terms containing the coefficients $\mathbf{F}_{i,k}(j,:)$ are omitted from Eqs. (11) and (12). On the other hand, if $j=1$, the terms containing the coefficients $g_{i,k}$ are omitted from these equations. Since the process $\mathbf{x}_n$ is stationary, the coefficients used in Eq. (11) for deriving the estimate $\hat{\mathbf{x}}_{MS,n,i,j}$ from $\mathbf{x}_{n-1},\ \mathbf{x}_{n-2},\ \ldots,\ \mathbf{x}_{n-i+1}$ and $\mathbf{x}_{n-i}(1),\ \ldots,\ \mathbf{x}_{n-i}(j-1)$ are constant and do not depend on n. Thus, it results from Eq. (12) and the ergodicity of the process $\mathbf{x}_n$ that for each i and j the sequence of random variables $\mathbf{z}_{n,i}(j)$ is an ergodic process with time variable n. In addition, it can be seen that for each i the sequence of random vector variables $\mathbf{z}_{n,i}$ is also an ergodic vector process with time variable n. Note that $\mathbf{z}_{n,i}$ is a zero mean vector process since $\mathbf{x}_n$ is zero mean.

Now, we consider the Euclidean space $\Re^{N-q}$ and assume that $mq<N-q$ (i.e., we assume that the number of unknown variables in Eq. (3) is less than the number of equations). The following vectors are members of $\Re^{N-q}$:

$$\mathbf{x}_{j,i}=\begin{bmatrix}\mathbf{x}_j(i) & \mathbf{x}_{j+1}(i) & \ldots & \mathbf{x}_{N-q+j-1}(i)\end{bmatrix}^T;\quad i=1,2,\ldots,m,$$

$$j=\ldots,1,2,\ldots,q+1,\ldots \tag{13}$$

Note that the set of vectors defined in Eq. (13) contain the vectors $\mathbf{x}_{q+1,i}$ which were used in Eq. (3). $\Re^{N-q}$ is a Hilbert space with the inner product $\langle\mathbf{r},\mathbf{s}\rangle=\sum_{i=1}^{N-q}r(i)s(i)$, where $\mathbf{r}$ and $\mathbf{s}$ are arbitrary members of $\Re^{N-q}$. The following set, which contains mq vectors, is a subset of $\Re^{N-q}$:

$$S=\{\mathbf{x}_{q,1},\mathbf{x}_{q,2},\ldots,\mathbf{x}_{q,m},\mathbf{x}_{q-1,1},\ldots,\mathbf{x}_{q-1,m},\ldots,\mathbf{x}_{1,1},\mathbf{x}_{1,2},\ldots,\mathbf{x}_{1,m}\} \tag{14}$$

We denote the closed span of S as M

$$M=\overline{SP}\{S\} \tag{15}$$

Note that M is the set of all linear combinations of the members of S. Now, we define

$$\mathbf{z}_{1,i,j}=[\mathbf{z}_{q+1,i}(j)\quad \mathbf{z}_{q+2,i}(j)\quad\ldots\quad\mathbf{z}_{N,i}(j)]^T;\quad i=1,2,\ldots,q,\quad j=1,2,\ldots,m \tag{16}$$

It is shown in Appendix A that each vector $\mathbf{z}_{1,i,j}$ can be written as a linear combination (with constant coefficients) of the members of S. So, all vectors $\mathbf{z}_{1,i,j}$ defined by Eq. (16) are members of M.

We defined the matrix $\mathbf{X}$ earlier in this section. It follows from this definition and Eq. (13) that

$$\mathbf{X}=\begin{bmatrix}\mathbf{x}_q(1) & \cdots & \mathbf{x}_q(m) & \mathbf{x}_{q-1}(1) & \cdots & \mathbf{x}_{q-1}(m) & \cdots & \mathbf{x}_1(1) & \cdots & \mathbf{x}_1(m)\\ \mathbf{x}_{q+1}(1) & \cdots & \mathbf{x}_{q+1}(m) & \mathbf{x}_q(1) & \cdots & \mathbf{x}_q(m) & \cdots & \mathbf{x}_2(1) & \cdots & \mathbf{x}_2(m)\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathbf{x}_{N-1}(1) & \cdots & \mathbf{x}_{N-1}(m) & \mathbf{x}_{N-2}(1) & \cdots & \mathbf{x}_{N-2}(m) & \cdots & \mathbf{x}_{N-q}(1) & \cdots & \mathbf{x}_{N-q}(m)\end{bmatrix}$$
$$=\begin{bmatrix}\mathbf{x}_{q,1} & \cdots & \mathbf{x}_{q,m} & \mathbf{x}_{q-1,1} & \cdots & \mathbf{x}_{q-1,m} & \cdots & \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,m}\end{bmatrix} \tag{17}$$

It can be seen from Eqs. (14) and (17) that $\mathbf{X}$ has mq columns and the set of columns of $\mathbf{X}$ is identical with S. Now, we define the $(N-q)\times mq$ matrix $\mathbf{Z}$ as follows:

$$\mathbf{Z}=\begin{bmatrix}\mathbf{z}_{1,1,1} & \mathbf{z}_{1,1,2} & \cdots & \mathbf{z}_{1,1,m} & \mathbf{z}_{1,2,1} & \cdots & \mathbf{z}_{1,2,m} & \cdots & \mathbf{z}_{1,q,1} & \cdots & \mathbf{z}_{1,q,m}\end{bmatrix} \tag{18}$$

Note that the dimensions of $\mathbf{X}$ and $\mathbf{Z}$ are identical. The matrix $\mathbf{Z}$ contains all vectors $\mathbf{z}_{1,i,j}$ defined by Eq. (16) as its columns. Thus, all mq columns of $\mathbf{Z}$ are members of M.

It can be seen from the definition of the VAR process $\mathbf{x}_n$ that each column of $\mathbf{X}$ can be written as a summation of a noise vector and a linear combination of the vectors $\mathbf{x}_{j,i}$. This noise vector cannot be written as a linear combination of other columns of $\mathbf{X}$ (we ignore very exceptional cases that may occur for the VAR data). So, no column of $\mathbf{X}$ can be written as a linear combination of other columns of $\mathbf{X}$. This shows that all mq columns of $\mathbf{X}$ are linearly independent and the rank of $\mathbf{X}$ is equal to mq. (Note that number of rows of $\mathbf{X}$ is more than the number of its columns, i.e., $mq<N-q$.) It is shown in Appendix B that $\mathbf{Z}$ can be written as the product of $\mathbf{X}$ and a nonsingular matrix $\mathbf{F}$. It is known that multiplication by a nonsingular matrix does not change the rank [19]. Thus, the rank of $\mathbf{Z}$ is equal to mq. This shows that the columns of $\mathbf{Z}$ form a set of mq linearly independent vectors in M, and M is equal to their closed span

$$M=\overline{SP}\{S\}=\overline{SP}\{\mathbf{z}_{1,1,1},\mathbf{z}_{1,1,2},\ldots,\mathbf{z}_{1,1,m},\mathbf{z}_{1,2,1},\ldots,\mathbf{z}_{1,2,m},\ldots,\mathbf{z}_{1,q,1},\ldots,\mathbf{z}_{1,q,m}\} \tag{19}$$

The matrices $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Z}^T\mathbf{Z}$ are $mq\times mq$, and the rank of $\mathbf{X}$ and $\mathbf{Z}$ is equal to mq. Thus, $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Z}^T\mathbf{Z}$ are full rank. So, using the projection theorem for Euclidean spaces [20], it can be seen that the unique orthogonal projection of $\mathbf{x}_{q+1,i}$ onto M, denoted by $\hat{\mathbf{x}}_{q+1,i}$ or $P_M(\mathbf{x}_{q+1,i})$, can be obtained in the following ways:

$$\hat{\mathbf{x}}_{q+1,i}=P_M(\mathbf{x}_{q+1,i})=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}_{q+1,i};\quad i=1,2,\ldots,m \tag{20}$$

$$\hat{\mathbf{x}}_{q+1,i}=P_M(\mathbf{x}_{q+1,i})=\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{x}_{q+1,i};\quad i=1,2,\ldots,m \tag{21}$$

Note that combining Eqs. (5) and (20), we have

$$\hat{\mathbf{x}}_{q+1,i}=\mathbf{X}\hat{\mathbf{A}}_i;\quad i=1,2,\ldots,m \tag{22}$$

Using Eq. (1), we can write

$$\mathbf{x}_n(i)=\mathbf{\Phi}_1(i,:)\mathbf{x}_{n-1}+\ldots+\mathbf{\Phi}_q(i,:)\mathbf{x}_{n-q}+\mathbf{w}_n(i);$$
$$i=1,2,\ldots,m,\quad q\geq p \tag{23}$$

where

$$\mathbf{\Phi}_i=\mathbf{0};\quad i>p \tag{24}$$

Writing Eq. (23) for $n=q+1,\ldots,N$ and collecting these equations in a matrix equation, we obtain

$$\begin{bmatrix}\mathbf{x}_{q+1}(i)\\ \mathbf{x}_{q+2}(i)\\ \vdots\\ \mathbf{x}_N(i)\end{bmatrix}=\begin{bmatrix}\mathbf{x}_q^T & \mathbf{x}_{q-1}^T & \cdots & \mathbf{x}_1^T\\ \mathbf{x}_{q+1}^T & \mathbf{x}_q^T & \cdots & \mathbf{x}_2^T\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{x}_{N-1}^T & \mathbf{x}_{N-2}^T & \cdots & \mathbf{x}_{N-q}^T\end{bmatrix}\begin{bmatrix}\mathbf{\Phi}_1^T(i,:)\\ \mathbf{\Phi}_2^T(i,:)\\ \vdots\\ \mathbf{\Phi}_q^T(i,:)\end{bmatrix}+\begin{bmatrix}\mathbf{w}_{q+1}(i)\\ \mathbf{w}_{q+2}(i)\\ \vdots\\ \mathbf{w}_N(i)\end{bmatrix};\quad i=1,2,\ldots,m,\quad q\geq p \tag{25}$$

Using Eq. (4), we can rewrite Eq. (25) in the following way:

$$\mathbf{x}_{q+1,i}=\mathbf{X}\mathbf{A}_i+\mathbf{w}_{q+1,i};\quad i=1,2,\ldots,m,\quad q\geq p \tag{26}$$

where

$$\mathbf{w}_{q+1,i}=\begin{bmatrix}\mathbf{w}_{q+1}(i) & \mathbf{w}_{q+2}(i) & \cdots & \mathbf{w}_N(i)\end{bmatrix}^T;\quad i=1,2,\ldots,m \tag{27}$$

The vector $\mathbf{X}\mathbf{A}_i$ is a linear combination of the columns of $\mathbf{X}$. So, $\mathbf{X}\mathbf{A}_i$ is a member of M because the columns of $\mathbf{X}$ are members of M. This shows that the projection of $\mathbf{X}\mathbf{A}_i$ onto M is equal to $\mathbf{X}\mathbf{A}_i$ itself. Since the operator $P_M(.)$ is linear, we have

$$\begin{aligned}P_M(\mathbf{x}_{q+1,i})&=P_M(\mathbf{X}\mathbf{A}_i+\mathbf{w}_{q+1,i})=P_M(\mathbf{X}\mathbf{A}_i)+P_M(\mathbf{w}_{q+1,i})\\ &=\mathbf{X}\mathbf{A}_i+\hat{\mathbf{w}}_{q+1,i};\quad i=1,2,\ldots,m,\quad q\geq p\end{aligned} \tag{28}$$

where $\hat{\mathbf{w}}_{q+1,i}$ is the projection of $\mathbf{w}_{q+1,i}$ onto M. According to the projection theorem, we have

$$\hat{\mathbf{w}}_{q+1,i}=\mathbf{Z}\hat{\mathbf{c}}_i,\quad \hat{\mathbf{c}}_i=(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{w}_{q+1,i};\quad i=1,2,\ldots,m \tag{29}$$

From Eqs. (26) and (28), the following equality for the projection errors of $\mathbf{x}_{q+1,i}$ and $\mathbf{w}_{q+1,i}$ onto M is obtained:

$$\mathbf{x}_{q+1,i}-\hat{\mathbf{x}}_{q+1,i}=\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i};\quad i=1,2,\ldots,m,\quad q\geq p \tag{30}$$

3. Residual variance and covariance matrix of the input noise

The covariance matrix of the input noise of the VAR model can be estimated as follows:

$$\hat{\mathbf{\Sigma}}=\frac{1}{N-q}\sum_{k=q+1}^{N}(\mathbf{x}_k-\hat{\mathbf{x}}_k)(\mathbf{x}_k-\hat{\mathbf{x}}_k)^{\mathrm{T}} \tag{31}$$

where

$$\hat{\mathbf{x}}_k=\hat{\mathbf{\Phi}}_1\mathbf{x}_{k-1}+\ldots+\hat{\mathbf{\Phi}}_q\mathbf{x}_{k-q} \tag{32}$$

and the $\hat{\mathbf{\Phi}}_i$ are determined by Eqs. (4) and (5). It follows from Eq. (32) that

$$\hat{\mathbf{x}}_{q+j}(i)=\sum_{k=1}^{q}\hat{\mathbf{\Phi}}_k(i,:)\mathbf{x}_{q+j-k};\quad i=1,2,\ldots,m,\quad j=1,2,\ldots,N-q \tag{33}$$

Using Eqs. (4), (22), and (33), we have

$$\hat{\mathbf{x}}_{q+1,i}(j)=\sum_{k=1}^{q}\mathbf{x}_{q+j-k}^T\hat{\mathbf{\Phi}}_k^T(i,:)=\sum_{k=1}^{q}\hat{\mathbf{\Phi}}_k(i,:)\mathbf{x}_{q+j-k}=\hat{\mathbf{x}}_{q+j}(i);$$
$$i=1,2,\ldots,m,\quad j=1,2,\ldots,N-q \tag{34}$$

It can be seen from Eqs. (4), (31), and (34) that

$$\begin{aligned}\hat{\mathbf{\Sigma}}(i,j)&=\frac{1}{N-q}\sum_{k=q+1}^{N}[\mathbf{x}_k(i)-\hat{\mathbf{x}}_k(i)][\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)]\\ &=\frac{1}{N-q}(\mathbf{x}_{q+1,i}-\hat{\mathbf{x}}_{q+1,i})^T(\mathbf{x}_{q+1,j}-\hat{\mathbf{x}}_{q+1,j});\quad i,j=1,\ldots,m\end{aligned} \tag{35}$$

It follows from Eqs. (35) and (30) that

$$\hat{\mathbf{\Sigma}}(i,j)=\frac{1}{N-q}(\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i})^T(\mathbf{w}_{q+1,j}-\hat{\mathbf{w}}_{q+1,j});\quad q\geq p. \tag{36}$$

Using Eq. (36), it is shown in Appendix C that

$$\hat{\mathbf{\Sigma}}=\frac{1}{N-q}\left\{\left[\sum_{k=1}^{N-q}\mathbf{w}_{q+k}\mathbf{w}_{q+k}^T\right]-\hat{\mathbf{C}}^T\mathbf{Z}^T\mathbf{Z}\hat{\mathbf{C}}\right\};\quad q\geq p. \tag{37}$$

where the matrix $\hat{\mathbf{C}}$ is defined as

$$\hat{\mathbf{C}}=\begin{bmatrix}\hat{\mathbf{c}}_1 & \cdots & \hat{\mathbf{c}}_m\end{bmatrix} \tag{38}$$

In addition, it is shown in Appendix C that

$$E[\mathbf{Z}^T\mathbf{Z}]=(N-q)\mathbf{I}_{mq} \tag{39}$$

and for large $N-q$ it is shown that

$$\mathbf{Z}^T\mathbf{Z}\approx(N-q)\mathbf{I}_{mq} \tag{40}$$

Henceforth, we always assume that $N-q$ is large (i.e., $N-q\gg 1$).

Substituting Eq. (40) into Eq. (37), we obtain

$$\hat{\mathbf{\Sigma}}\approx\frac{1}{N-q}\left[\sum_{k=1}^{N-q}\mathbf{w}_{q+k}\mathbf{w}_{q+k}^T\right]-\hat{\mathbf{C}}^T\hat{\mathbf{C}};\quad q\geq p \tag{41}$$

It follows:

$$\begin{aligned}E\left\{\hat{\mathbf{\Sigma}}\right\}&\approx\frac{1}{N-q}\left[\sum_{k=1}^{N-q}E\{\mathbf{w}_{q+k}\mathbf{w}_{q+k}^T\}\right]-E\{\hat{\mathbf{C}}^T\hat{\mathbf{C}}\}\\ &=\mathbf{\Sigma}-E\{\hat{\mathbf{C}}^T\hat{\mathbf{C}}\};\quad q\geq p\end{aligned} \tag{42}$$

For the estimated model given by Eq. (2), we define the residual variance as

$$S^2(q) = \frac{1}{N-q}\sum_{i=q+1}^{N} (\mathbf{x}_i - \hat{\mathbf{x}}_i)^T(\mathbf{x}_i - \hat{\mathbf{x}}_i) \tag{43}$$

where $\hat{\mathbf{x}}_i$ is given by Eq. (32). Residual variance is a measure of fitness of the estimated model to the data used for estimating model parameters. It can be seen from Eqs. (35) and (43) that

$$\begin{aligned} S^2(q) &= \frac{1}{N-q}\sum_{i=q+1}^{N}\sum_{j=1}^{m} (\mathbf{x}_i(j) - \hat{\mathbf{x}}_i(j))^2 \\ &= \frac{1}{N-q}\sum_{j=1}^{m}\sum_{i=q+1}^{N} (\mathbf{x}_i(j) - \hat{\mathbf{x}}_i(j))^2 = \sum_{j=1}^{m} \hat{\boldsymbol{\Sigma}}(j,j) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}) \end{aligned} \tag{44}$$

where tr(.) indicates the trace. It follows from Eqs. (42) and (43) that

$$E\{S^2(q)\} \approx \mathrm{tr}(\boldsymbol{\Sigma}) - E\{\mathrm{tr}(\hat{\mathbf{C}}^T\hat{\mathbf{C}})\} = \mathrm{tr}(\boldsymbol{\Sigma}) - \sum_{i=1}^{m} E\{\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_i\}; \quad q \ge p \tag{45}$$

Eqs. (42) and (45) will be used in Section 6 for deriving new approximations for $E\{\hat{\boldsymbol{\Sigma}}\}$ and $E\{S^2(q)\}$.

4. Prediction error and the prediction error covariance matrix

We assumed that the observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ are samples of a sample function of the VAR process given by Eq. (1). Suppose that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ are samples of another arbitrary sample function of this process. We define the matrices and vectors $\mathbf{Y}$, $\mathbf{Z}'$, $\mathbf{y}_{q+1,i}$, $\mathbf{w}'_{q+1,i}$, and $\hat{\mathbf{c}}'_i$ for $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$, exactly in the same way that we defined $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{x}_{q+1,i}$, $\mathbf{w}_{q+1,i}$, and $\hat{\mathbf{c}}_i$, respectively, for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. We define the prediction error covariance matrix as follows:

$$\mathbf{P}_\eta = E_y\{(\mathbf{y}_i - \hat{\mathbf{y}}_i)(\mathbf{y}_i - \hat{\mathbf{y}}_i)^T\} \tag{46}$$

where

$$\hat{\mathbf{y}}_i = \hat{\boldsymbol{\Phi}}_1\mathbf{y}_{i-1} + \ldots + \hat{\boldsymbol{\Phi}}_q\mathbf{y}_{i-q} \tag{47}$$

and the $\hat{\boldsymbol{\Phi}}_j$'s are given by Eq. (5). Note that, in Eq. (46), the $\hat{\boldsymbol{\Phi}}_j$'s are supposed to be deterministic and the expectation is taken assuming that $\mathbf{y}_i$'s are stochastic. Now, we define

$$\begin{aligned} &\mathbf{U} = \begin{bmatrix} \mathbf{y}_{q+1} & \mathbf{y}_{q+2} & \cdots & \mathbf{y}_N \end{bmatrix}^T, \\ &\hat{\mathbf{U}} = \begin{bmatrix} \hat{\mathbf{y}}_{q+1} & \hat{\mathbf{y}}_{q+2} & \cdots & \hat{\mathbf{y}}_N \end{bmatrix}^T, \quad \mathbf{H} = E_y\{(\mathbf{U} - \hat{\mathbf{U}})^T(\mathbf{U} - \hat{\mathbf{U}})\} \end{aligned} \tag{48}$$

Using Eqs. (46) and (48), we get

$$\begin{aligned} \frac{1}{N-q}\mathbf{H} &= \frac{1}{N-q}E_y\left\{(\mathbf{U} - \hat{\mathbf{U}})^T(\mathbf{U} - \hat{\mathbf{U}})\right\} \\ &= \frac{1}{N-q}E_y\left\{\sum_{k=1}^{N-q}[(\mathbf{y}_{q+k} - \hat{\mathbf{y}}_{q+k})(\mathbf{y}_{q+k} - \hat{\mathbf{y}}_{q+k})^T]\right\} \\ &= \frac{1}{N-q}\sum_{k=1}^{N-q} E_y\{(\mathbf{y}_{q+k} - \hat{\mathbf{y}}_{q+k})(\mathbf{y}_{q+k} - \hat{\mathbf{y}}_{q+k})^T\} = \mathbf{P}_\eta \end{aligned} \tag{49}$$

It is shown in Appendix D that

$$\begin{aligned} \mathbf{U} - \hat{\mathbf{U}} &= \begin{bmatrix} \mathbf{y}_{q+1,1} - \mathbf{Y}\hat{\mathbf{A}}_1 \cdots & \mathbf{y}_{q+1,m} - \mathbf{Y}\hat{\mathbf{A}}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{w}'_{q+1,1} & \cdots & \mathbf{w}'_{q+1,m} \end{bmatrix} - \mathbf{Z}'\begin{bmatrix} \hat{\mathbf{c}}_1 \cdots & \hat{\mathbf{c}}_m \end{bmatrix}; \quad q \ge p \end{aligned} \tag{50}$$

From Eqs. (38), (50), and the definition of $\mathbf{w}'_{q+1,i}$, we obtain

$$\mathbf{U} - \hat{\mathbf{U}} = \begin{bmatrix} \mathbf{w}'_{q+1} & \cdots & \mathbf{w}'_N \end{bmatrix}^T - \mathbf{Z}'\hat{\mathbf{C}}; \quad q \ge p \tag{51}$$

The matrix $\mathbf{H}$ is defined by Eq. (48) and it can be seen from Eq. (50) that the element in the ith row and jth column of $\mathbf{H}$ is equal to

$$\begin{aligned} (\mathbf{H})_{i,j} &= E_y\{(\mathbf{w}'_{q+1,i} - \mathbf{Z}'\hat{\mathbf{c}}_i)^T(\mathbf{w}'_{q+1,j} - \mathbf{Z}'\hat{\mathbf{c}}_j)\} \\ &= E_y\{\mathbf{w}'^T_{q+1,i}\mathbf{w}'_{q+1,j}\} - \hat{\mathbf{c}}_i^T E_y\{\mathbf{Z}'^T\mathbf{w}'_{q+1,j}\} \\ &\quad - E_y\{\mathbf{w}'^T_{q+1,i}\mathbf{Z}'\}\hat{\mathbf{c}}_j + \hat{\mathbf{c}}_i^T E_y\{\mathbf{Z}'^T\mathbf{Z}'\}\hat{\mathbf{c}}_j; \quad q \ge p \end{aligned} \tag{52}$$

For each i and each $m \ge n$, $\mathbf{w}_m(i)$ is independent of $\mathbf{x}_{n-1}, \mathbf{x}_{n-2}, \ldots$. On the other hand, $\mathbf{z}_{n,k}(j)$ is equal to a linear combination of the elements of $\mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-k}$. So if $m \ge n$, then $\mathbf{w}_m(i)$ is independent of $\mathbf{z}_{n,k}(j)$for each i, j, and k. Thus, for each i, each element of $\mathbf{w}_{q+1,i}$ is independent of the corresponding element in each column of $\mathbf{Z}$. It can be shown in the same way that, for each i, each element of $\mathbf{w}'_{q+1,i}$ is independent of the corresponding element in each column of $\mathbf{Z}'$. It follows that

$$E_y\{\mathbf{w}'^T_{q+1,i}\mathbf{Z}'\} = E_y\{\mathbf{w}'^T_{q+1,i}\}E_y\{\mathbf{Z}'\} = \mathbf{0} \tag{53}$$

It is obvious that a relation similar to Eq. (39) is correct for $\mathbf{Z}'$

$$E[\mathbf{Z}'^T\mathbf{Z}'] = (N-q)\mathbf{I}_{mq} \tag{54}$$

It can be seen from Eqs. (27), (52)–(54), and the definition of $\boldsymbol{\Sigma}$ that

$$(\mathbf{H})_{i,j} = (N-q)\boldsymbol{\Sigma}(i,j) + (N-q)\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_j; \quad q \ge p \tag{55}$$

It follows that

$$\mathbf{H} = E_y\{(\mathbf{U} - \hat{\mathbf{U}})^T(\mathbf{U} - \hat{\mathbf{U}})\} = (N-q)\boldsymbol{\Sigma} + (N-q)\hat{\mathbf{C}}^T\hat{\mathbf{C}}; \quad q \ge p \tag{56}$$

Substituting Eq. (56) into Eq. (49), we obtain

$$\mathbf{P}_\eta = \boldsymbol{\Sigma} + \hat{\mathbf{C}}^T\hat{\mathbf{C}}; \quad q \ge p \tag{57}$$

and thus

$$E\{\mathbf{P}_\eta\} = \boldsymbol{\Sigma} + E\{\hat{\mathbf{C}}^T\hat{\mathbf{C}}\}; \quad q \ge p \tag{58}$$

For the estimated model given by Eq. (2), we define the prediction error as

$$\mathrm{PE}(q) = E_y\left\{\frac{1}{N-q}\sum_{i=q+1}^{N} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^T(\mathbf{y}_i - \hat{\mathbf{y}}_i)\right\} \tag{59}$$

Prediction error is a measure of fitness of the estimated model to the samples of a realization of the AR process independent of the observed one. It can be seen from Eqs. (46) and (59) that

$$\begin{aligned} \mathrm{PE}(q) &= \frac{1}{N-q}\sum_{i=q+1}^{N} E_y\{(\mathbf{y}_i - \hat{\mathbf{y}}_i)^T(\mathbf{y}_i - \hat{\mathbf{y}}_i)\} \\ &= E_y\{(\mathbf{y}_i - \hat{\mathbf{y}}_i)^T(\mathbf{y}_i - \hat{\mathbf{y}}_i)\} = \mathrm{tr}(\mathbf{P}_\eta) \end{aligned} \tag{60}$$

Substituting Eq. (57) into Eq. (60), we get

$$\mathrm{PE}(q) = \mathrm{tr}(\boldsymbol{\Sigma}) + \mathrm{tr}(\hat{\mathbf{C}}^T\hat{\mathbf{C}}) = \mathrm{tr}(\boldsymbol{\Sigma}) + \sum_{i=1}^{m} \hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_i; \quad q \ge p \tag{61}$$

So the expectation of the prediction error is equal to

$$E\{\mathrm{PE(q)}\} = \mathrm{tr}(\boldsymbol{\Sigma}) + \mathrm{tr}(E\{\hat{\mathbf{C}}^T\hat{\mathbf{C}}\}) = \mathrm{tr}(\boldsymbol{\Sigma}) + \sum_{i=1}^{m} E\{\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_i\}; \quad q \ge p \tag{62}$$

Combining Eqs. (45) and (62), we obtain

$$\frac{E\{\mathrm{PE(q)}\}}{E\{S^2(q)\}} \approx \frac{\mathrm{tr}(\boldsymbol{\Sigma}) + \sum_{i=1}^{m} E\{\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_i\}}{\mathrm{tr}(\boldsymbol{\Sigma}) - \sum_{i=1}^{m} E\{\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_i\}}; \quad q \ge p \tag{63}$$

5. Statistical properties of $\hat{\mathbf{c}}_i$'s

In this section some statistical properties of the random vectors $\hat{\mathbf{c}}_i$ are obtained. These properties will be used in the following sections to obtain new AR order selection criteria.

Substituting Eq. (40) into Eq. (29), we get

$$\hat{\mathbf{c}}_i = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{w}_{q+1,i} \approx \frac{1}{N-q}\mathbf{Z}^T\mathbf{w}_{q+1,i}; \quad i = 1,2,\ldots,m \tag{64}$$

It follows from Eqs. (16), (18), (27), and (64) that

$$\hat{\mathbf{c}}_i(j) \approx \frac{1}{N-q}\mathbf{z}_{1,l,h}^T\mathbf{w}_{q+1,i} = \frac{1}{N-q}\sum_{k=1}^{N-q}\mathbf{z}_{q+k,l}(h)\mathbf{w}_{q+k}(i);$$
$$i = 1,\ldots,m, \quad j = 1,\ldots,mq \tag{65}$$

where l and h are defined as follows:

$$l = \left\lceil \frac{j-1}{m} \right\rceil + 1, \quad h = j-(l-1)m \tag{66}$$

where $\lceil\,\rceil$ indicates the integer part. The discussions following Eq. (52) indicate that $\mathbf{w}_{q+k}(i)$ is independent of $\mathbf{z}_{q+k,l}(h)$ for each i and j. So, we have

$$E\{\hat{\mathbf{c}}_i(j)\} \approx \frac{1}{N-q}\sum_{k=1}^{N-q} E\{\mathbf{z}_{q+k,l}(h)\}E\{\mathbf{w}_{q+k}(i)\} = 0; \quad i = 1,\ldots,m, \quad j = 1,\ldots,mq \tag{67}$$

and thus

$$E\{\hat{\mathbf{c}}_i\} \approx \mathbf{0}, \quad E\{\hat{\mathbf{C}}\} \approx \mathbf{0} \tag{68}$$

Using Eq. (65), it is shown in Appendix E that

$$E\{\hat{\mathbf{c}}_{i_1}(j_1)\hat{\mathbf{c}}_{i_2}(j_2)\} \approx \begin{cases} 0; & j_1 \ne j_2 \\ \dfrac{1}{(N-q)^2}\displaystyle\sum_{k=1}^{N-q}\boldsymbol{\Sigma}(i_1,i_2) = \dfrac{1}{(N-q)}\boldsymbol{\Sigma}(i_1,i_2); & j_1 = j_2 \end{cases} \tag{69}$$

Thus we conclude that

$$E\left\{\hat{\mathbf{c}}_i\hat{\mathbf{c}}_j^T\right\} \approx \frac{\boldsymbol{\Sigma}(i,j)}{N-q}\mathbf{I}_{mq} \tag{70}$$

$$E\{\hat{\mathbf{c}}_i^T\hat{\mathbf{c}}_j\} \approx \frac{mq}{N-q}\boldsymbol{\Sigma}(i,j) \tag{71}$$

It follows that

$$E\left\{\hat{\mathbf{C}}^T\hat{\mathbf{C}}\right\} \approx \frac{mq}{N-q}\boldsymbol{\Sigma} \tag{72}$$

6. The finite sample FPE

The results obtained in the previous sections make it possible to derive new approximations for the expectations of residual variance and prediction error. Substituting Eq. (71) into Eqs.(45) and (62), we obtain

$$E\{S^2(q)\} \approx \mathrm{tr}(\boldsymbol{\Sigma}) - \sum_{i=1}^{m}\frac{mq}{N-q}\boldsymbol{\Sigma}(i,i) = \left(1-\frac{mq}{N-q}\right)\mathrm{tr}(\boldsymbol{\Sigma}); \quad q \ge p \tag{73}$$

$$E\{\mathrm{PE}(q)\} \approx \mathrm{tr}(\boldsymbol{\Sigma}) + \sum_{i=1}^{m}\frac{mq}{N-q}\boldsymbol{\Sigma}(i,i) = \left(1+\frac{mq}{N-q}\right)\mathrm{tr}(\boldsymbol{\Sigma}); \quad q \ge p \tag{74}$$

It can be seen from Eqs. (73) and (74) that the expectations of $S^2(q)$ and $\mathrm{PE}(q)$ are (approximately) independent of the true process order p if q is greater than or equal to p.

Substituting Eq. (71) into Eq. (63), we obtain

$$\frac{E\{\mathrm{PE(q)}\}}{E\{S^2(q)\}} \approx \left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right); \quad q \ge p \tag{75}$$

In practice, when we have N samples of the VAR process we can compute $S^2(q)$ by using Eq. (43), but $\mathrm{PE}(q)$ cannot be computed. We can use Eq. (75) to obtain the following estimate for prediction error:

$$\hat{\mathrm{PE}}(q) = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)S^2(q); \quad q \ge p \tag{76}$$

It is reasonable to choose the integer q that minimizes the estimated prediction error as the appropriate order for the VAR process. So, we propose the first version of finite sample FPE (FPEF1) criterion as follows:

$$\mathrm{FPEF1}(q) = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)S^2(q) = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) \tag{77}$$

One version of the Akaike's FPE criterion [7] is defined as follows:

$$\mathrm{FPE1}(q) = \left(\left(1+\frac{mq}{N}\right)\Big/\left(1-\frac{mq}{N}\right)\right)\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) \tag{78}$$

This version of FPE gives an estimate of prediction error. Another version of FPE gives an estimate of the determinant of the prediction error covariance matrix $\mathbf{P}_\eta$. It is defined as follows:

$$\mathrm{FPE2}(q) = \left(\left(1+\frac{mq}{N}\right)\Big/\left(1-\frac{mq}{N}\right)\right)^m\det(\hat{\boldsymbol{\Sigma}}) \tag{79}$$

where det(.) indicates the determinant. The results obtained in previous sections enable us to derive an estimate of the determinant of $\mathbf{P}_\eta$. Substituting Eq. (72) into Eq. (42), we get

$$E\left\{\hat{\boldsymbol{\Sigma}}\right\} \approx \boldsymbol{\Sigma} - \frac{mq}{N-q}\boldsymbol{\Sigma} = \left(1-\frac{mq}{N-q}\right)\boldsymbol{\Sigma}; \quad q \ge p \tag{80}$$

In addition, substituting Eq. (72) into Eq. (58), we get

$$E\{\mathbf{P}_\eta\} \approx \boldsymbol{\Sigma} + \frac{mq}{N-q}\boldsymbol{\Sigma} = \left(1+\frac{mq}{N-q}\right)\boldsymbol{\Sigma}; \quad q \ge p \tag{81}$$

We can combine Eqs. (80) and (81) to obtain

$$E\{\mathbf{P}_\eta\} \approx \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)E\left\{\hat{\boldsymbol{\Sigma}}\right\}; \quad q \geq p \tag{82}$$

So the prediction error covariance matrix can be estimated as follows:

$$\hat{\mathbf{P}}_{\boldsymbol{\eta}} = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)\hat{\boldsymbol{\Sigma}}; \quad q \geq p \tag{83}$$

It follows from Eq. (83) that

$$\det(\hat{\mathbf{P}}_{\boldsymbol{\eta}}) = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)^m \det(\hat{\boldsymbol{\Sigma}}); \quad q \geq p \tag{84}$$

So we propose the second version of finite sample FPE (FPEF2) criterion as follows:

$$\mathrm{FPEF2}(q) = \left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)^m \det(\hat{\boldsymbol{\Sigma}}) \tag{85}$$

Comparing Eqs. (77) with (78) and (85) with (79), we see that the increase in FPEF1(q) and FPEF2(q) with q is more rapid than FPE1(q) and FPE2(q), respectively. As a result, the probability of overfit for FPEF1 and FPEF2 is less than FPE1 and FPE2, respectively. However, it can be seen that in the asymptotic case ($N \to \infty$), the new criteria FPEF1 and FPEF2 become identical with the criteria FPE1 and FPE2, respectively.

In the univariate (scalar) case, the two versions of FPE and FPEF become identical and we have

$$\mathrm{FPE}(q) = \left(\left(1+\frac{q}{N}\right)\Big/\left(1-\frac{q}{N}\right)\right)S^2(q) = \left(\frac{N+q}{N-q}\right)S^2(q) \tag{86}$$

$$\mathrm{FPEF}(q) = \left(\left(1+\frac{q}{N-q}\right)\Big/\left(1-\frac{q}{N-q}\right)\right)S^2(q) = \left(\frac{N}{N-2q}\right)S^2(q) \tag{87}$$

which is identical to the FPEF criterion obtained in Refs. [15,17].

It should be noted that, in deriving all results in this section and the previous sections, we have not assumed a particular probability density function (pdf) for the $\mathbf{w}_n$ process and these results are valid for any pdf.

7. The finite sample AIC

In this section we use the results obtained in the previous sections to derive a corrected version of the AIC criterion. It is assumed in this section that the VAR process given by Eq. (1) is a Gaussian process.

A useful measure of the discrepancy between the true and fitted models is the Kullback–Leibler index

$$\Delta(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}\{-2\ln[f(\mathbf{y}_1,\ldots,\mathbf{y}_N;\ \hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\Phi}}_1,\ldots,\hat{\boldsymbol{\Phi}}_q)]\} \tag{88}$$

where the samples $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ are defined in Section 4, $\hat{\boldsymbol{\theta}}$ is a matrix containing the parameters of the estimated (fitted) VAR model (i.e., containing $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Phi}}_1, \ldots,$ and $\hat{\boldsymbol{\Phi}}_q$), $\boldsymbol{\theta}_0$ is a matrix containing the true parameters of the VAR model, and $f(\mathbf{y}_1,\ldots,\mathbf{y}_N; \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\Phi}}_1, \ldots, \hat{\boldsymbol{\Phi}}_q)$ is the joint pdf of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ under the estimated model. $E_{\boldsymbol{\theta}_0}$ denotes the expectation under the true model, taken assuming that $\hat{\boldsymbol{\theta}}$ is deterministic. We have

$$f(\mathbf{y}_1,\ldots,\mathbf{y}_N;\ \hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\Phi}}_1,\ldots,\hat{\boldsymbol{\Phi}}_q) = \prod_{i=1}^{N}\left\{\frac{1}{\sqrt{(2\pi)^m\det(\hat{\boldsymbol{\Sigma}})}}\exp\left[\frac{-1}{2}(\mathbf{y}_i-\hat{\mathbf{y}}_i)^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i-\hat{\mathbf{y}}_i)\right]\right\} \tag{89}$$

where $\hat{\mathbf{y}}_i$ is given by Eq. (47). It follows that

$$\begin{aligned}-2\ln\left[f(\mathbf{y}_1,\ldots,\mathbf{y}_N;\ \hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\Phi}}_1,\ldots,\hat{\boldsymbol{\Phi}}_q)\right] &= \sum_{i=1}^{N}\left\{m\ln(2\pi)+\ln(\det(\hat{\boldsymbol{\Sigma}}))+(\mathbf{y}_i-\hat{\mathbf{y}}_i)^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i-\hat{\mathbf{y}}_i)\right\}\\ &= mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+\sum_{i=1}^{N}(\mathbf{y}_i-\hat{\mathbf{y}}_i)^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i-\hat{\mathbf{y}}_i)\end{aligned} \tag{90}$$

Substituting Eq. (90) into Eq. (88), we obtain

$$\begin{aligned}\Delta(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0) &= mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+\sum_{i=1}^{N}E_{\boldsymbol{\theta}_0}\{(\mathbf{y}_i-\hat{\mathbf{y}}_i)^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i-\hat{\mathbf{y}}_i)\}\\ &= mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+\left(\frac{N}{N-q}\right)\sum_{i=q+1}^{N}E_{\boldsymbol{\theta}_0}\{(\mathbf{y}_i-\hat{\mathbf{y}}_i)^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i-\hat{\mathbf{y}}_i)\}\\ &= mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+\left(\frac{N}{N-q}\right)E_{\boldsymbol{\theta}_0}\left\{\mathrm{tr}\left[(\mathbf{U}-\hat{\mathbf{U}})\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{U}-\hat{\mathbf{U}})^T\right]\right\}\end{aligned} \tag{91}$$

We can rewrite the last term in Eq. (91) in the following way:

$$\begin{aligned}E_{\boldsymbol{\theta}_0}\{\mathrm{tr}[(\mathbf{U}-\hat{\mathbf{U}})\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{U}-\hat{\mathbf{U}})^T]\} &= E_{\boldsymbol{\theta}_0}\{\mathrm{tr}[\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{U}-\hat{\mathbf{U}})^T(\mathbf{U}-\hat{\mathbf{U}})]\}\\ &= \mathrm{tr}\{\hat{\boldsymbol{\Sigma}}^{-1}E_{\boldsymbol{\theta}_0}[(\mathbf{U}-\hat{\mathbf{U}})^T(\mathbf{U}-\hat{\mathbf{U}})]\} = \mathrm{tr}\{\hat{\boldsymbol{\Sigma}}^{-1}[(N-q)\mathbf{P}_\eta]\}\end{aligned} \tag{92}$$

where the last equality follows from Eq. (49). Substituting Eq. (92) into Eq. (91), we get

$$\Delta(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0) = mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+N\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{P}_\eta\} \tag{93}$$

The prediction error covariance matrix $\mathbf{P}_\eta$ is not known, so we substitute its estimate, given by Eq. (83) into Eq. (93), to obtain an estimate of the Kullback–Leibler index as follows:

$$\begin{aligned}\Delta(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0) &\approx mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+N\left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right)\mathrm{tr}\left\{\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}\right\}\\ &= mN\ln(2\pi)+N\ln(\det(\hat{\boldsymbol{\Sigma}}))+mN\left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right); \quad q \geq p\end{aligned} \tag{94}$$

Subtracting the constant term $mN\ln(2\pi)$ from Eq. (94), we obtain the finite sample AIC (AICF) criterion as follows:

$$\mathrm{AICF}(q) = N\ln(\det(\hat{\boldsymbol{\Sigma}}))+mN\left(\left(1+\frac{mq}{N-q}\right)\Big/\left(1-\frac{mq}{N-q}\right)\right) \tag{95}$$

Subtracting the constant term mN from Eq. (95), we can rewrite the AICF criterion in the following form:

$$\mathrm{AICF}(q)=N\ln(\det(\hat{\boldsymbol{\Sigma}}))+\frac{2m^2qN}{N-(m+1)q} \tag{96}$$

The AIC criterion [8] is defined as follows:

$$\mathrm{AIC}(q)=N\ln(\det(\hat{\boldsymbol{\Sigma}}))+2m^2q \tag{97}$$

The AICC criterion [9] is given as

$$\mathrm{AICC}(q)=N\ln(\det(\hat{\boldsymbol{\Sigma}}))+mN\left(\left(1+\frac{mq}{N}\right)\Big/\left(1-\frac{mq+m+1}{N}\right)\right) \tag{98}$$

Subtracting the constant term mN from Eq. (98), we can rewrite the AICC criterion in the following form:

$$\mathrm{AICC}(q)=N\ln(\det(\hat{\boldsymbol{\Sigma}}))+N\left(\frac{2m^2q+m^2+m}{N-mq-m-1}\right) \tag{99}$$

Comparing Eqs. (96) with (97) and (95) with (98), we see that the increase in the penalty factor of AICF with q is more rapid than AIC and AICC. As a result, the probability of overfit for AICF is less than AIC and AICC.

In the univariate (scalar) case, it follows from Eq. (96) that

$$\mathrm{AICF}(q)=N\ln(S^2(q))+\frac{2qN}{N-2q} \tag{100}$$

which is identical to the AICF criterion obtained in Refs. [16,17].

8. Simulations

The criterion FPEF1 gives an estimate of the prediction error for orders greater than or equal to the true process order. So it is compared with the true prediction error and the estimate that FPE1 gives for the prediction error.

Consider the two-dimensional second-order Gaussian VAR process defined by

$$\boldsymbol{\Sigma}=\begin{bmatrix}1 & -0.08\\ -0.08 & 1\end{bmatrix},\quad \boldsymbol{\Phi}_1=\begin{bmatrix}0.5 & -0.3\\ 0.2 & 0.65\end{bmatrix},$$
$$\boldsymbol{\Phi}_2=\begin{bmatrix}-0.5 & 0.3\\ 0 & -0.4\end{bmatrix} \tag{101}$$

which was used in a simulation study on the performance of several order selection criteria in Ref. [9]. Fig. 1 gives plots of the mean value of the true prediction error, together with the mean value of the estimates given by FPE1 and FPEF1 for the process defined by Eq. (101). The plotted mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

The criterion FPEF2 gives an estimate of the determinant of the prediction error covariance matrix $\mathbf{P}_\eta$ for orders greater than or equal to the true process order. It is compared with the determinant of the true prediction error covariance matrix and the estimate given by FPE2 for it. Fig. 2 gives plots of the mean value of the determinant of the true prediction error covariance matrix, together with the mean value of the estimates that FPE2 and FPEF2 give for it, for the process defined by Eq. (101). The plotted mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

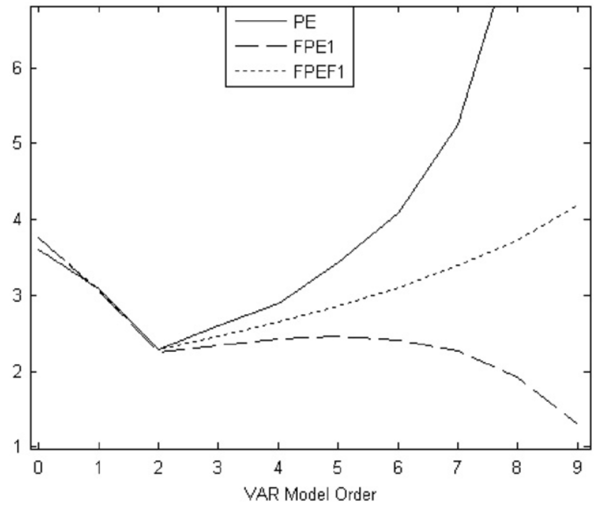

Fig. 1. Mean values of the true prediction error (PE), FPE1, and FPEF1, for the VAR process defined by Eq. (101). The mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

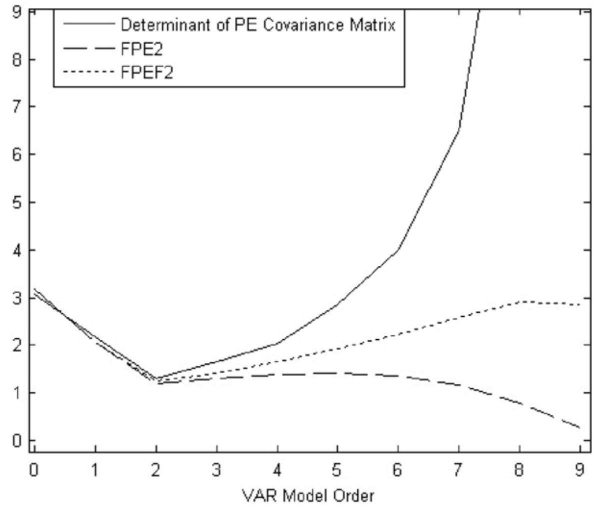

Fig. 2. Mean values of the determinant of the true prediction error (PE) covariance matrix, FPE2, and FPEF2, for the VAR process defined by Eq. (101). The mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

As the criterion AICF is an estimator of the Kullback–Leibler index for orders greater than or equal to the true process order, it is compared with the true Kullback–Leibler index and the estimates that AIC and AICC give for this index. Fig. 3 gives plots of the mean value of the Kullback–Leibler index $\Delta(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0)$ as given by Eq. (93), together with the mean values of the estimates that AIC, AICC, and AICF give for this index for the process defined by Eq. (101). The plotted mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

As another example, consider the one-dimensional fourth-order Gaussian AR process defined by

$$\boldsymbol{\Sigma}=1,\quad \boldsymbol{\Phi}_1=2.6978,\quad \boldsymbol{\Phi}_2=-3.3081,$$
$$\boldsymbol{\Phi}_3=2.1852,\quad \boldsymbol{\Phi}_4=-0.6561 \tag{102}$$

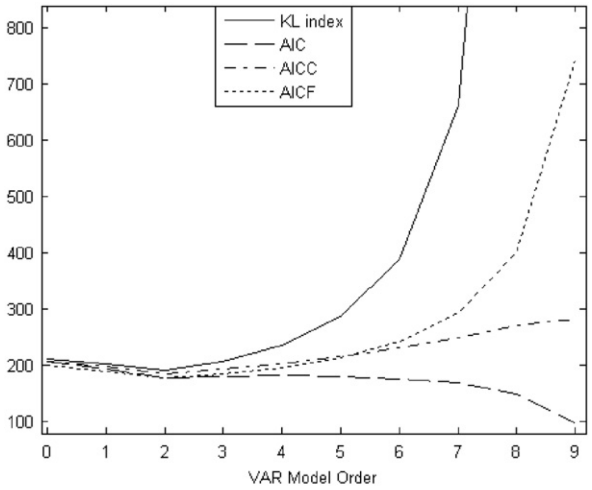

Fig. 3. Mean values of the Kullback–Leibler (KL) index, AIC, AICC, and AICF, for the VAR process defined by (101). The mean values are obtained using 2000 simulation runs of N=30 samples of the VAR process.

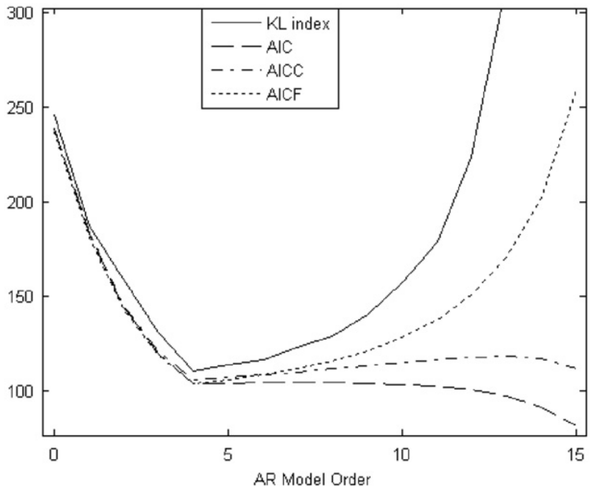

Fig. 4. Mean values of the Kullback–Leibler (KL) index, AIC, AICC, and AICF, for the AR process defined by Eq. (102). The mean values are obtained using 2000 simulation runs of N=35 samples of the AR process.

Fig. 4 gives plots of the mean value of the Kullback–Leibler index and its estimates for the process defined by Eq. (102). The plotted mean values are obtained using 2000 simulation runs of N=35 samples of the AR process.

Fig. 5 gives plots of the mean value of the Kullback–Leibler index and its estimates for another one-dimensional seventh-order Gaussian AR process defined by

$$\Sigma = 1, \quad \Phi_1 = 3.3400, \quad \Phi_2 = -5.9726, \quad \Phi_3 = 7.6320,$$
$$\Phi_4 = -7.3231, \quad \Phi_5 = 5.2747, \quad \Phi_6 = -2.7254,$$
$$\Phi_7 = 0.7661 \tag{103}$$

The plotted mean values are obtained using 2000 simulation runs of N=35 samples of the AR process.

It can be seen from Figs. 1 and 2 that FPEF1 and FPEF2 are less biased than FPE1 and FPE2, respectively. In addition, it can be seen from Figs. 3–5 that AICF is less biased than AIC and AICC. It is obvious from Figs. 1–5 that,

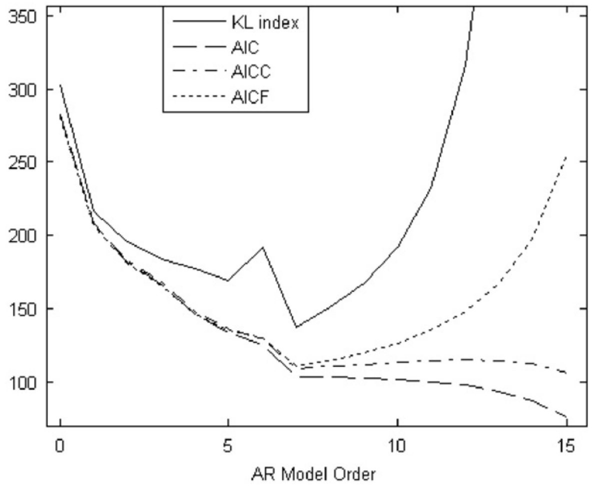

Fig. 5. Mean values of the Kullback–Leibler (KL) index, AIC, AICC, and AICF, for the AR process defined by Eq. (103). The mean values are obtained using 2000 simulation runs of N=35 samples of the AR process.

in the finite sample case, the criteria FPE1, FPE2, and AIC suffer from the serious drawback that they are often minimized at the maximum possible order. This drawback often leads to severe overfitting.

It is useful to compare the performance of various criteria in order selection for VAR processes. In Table 1, the performances of the criteria proposed in this paper are compared with other well-known criteria. This table gives the frequency of the orders selected by the criteria for the VAR process defined by Eq. (101). In addition, this table gives the average of prediction errors that result from the orders selected by each of the criteria. The criteria BIC [12], KIC [10], and KICC [11], which are used in this table, are defined as follows:

$$\mathrm{BIC}(q) = N\ln(\det(\hat{\Sigma})) + m^2 q \ln(N) \tag{104}$$

$$\mathrm{KIC}(q) = N\ln(\det(\hat{\Sigma})) + 3m^2 q \tag{105}$$

$$\mathrm{KICC}(q) = N\ln(\det(\hat{\Sigma})) + \frac{Nm(2mq+m+1)}{N-mq-m-1} + \frac{Nm}{N-mq-0.5(m-1)} + m^2 q \tag{106}$$

Another performance comparison is given in Table 2, where the average of prediction errors that result from the orders selected by each of the criteria is given. The process used in this Table is the AR process defined by Eq. (102).

It can be seen from Tables 1 and 2 that the FPEF criteria give lower average prediction errors than the FPE criteria. In addition, it is seen that the AICF criterion has the best performance and gives the minimum average prediction error among all criteria. Table 1 shows that many criteria suffer from the overfitting problem in the finite sample case. In fact, AICF is the only criterion that has never selected the maximum possible order.

An assumption we used in our approximations was that $N-q$ is large ($N-q \gg 1$). So, we expect that when

Table 1
Performance of various criteria in order selection for N=30 samples of the VAR process defined by Eq. (101). The entries of the first 10 rows represent the number of times a particular order was selected out of 2000 trials. The entries of the last row represent the average of prediction errors that result from the orders selected by each of the criteria.

Order	FPE1	FPEF1	FPE2	FPEF2	AIC	AICC	AICF	KIC	KICC	BIC
0	**1**	**8**	**0**	**6**	**0**	**41**	**25**	**8**	**213**	**37**
1	**9**	**71**	**5**	**59**	**0**	**221**	**198**	**22**	**373**	**57**
2	**117**	**1105**	**36**	**702**	**15**	**1528**	**1650**	**144**	**1349**	**268**
3	**17**	**191**	**3**	**138**	**2**	**102**	**105**	**8**	**39**	**12**
4	**8**	**81**	**2**	**66**	**0**	**29**	**20**	**4**	**4**	**2**
5	**10**	**56**	**4**	**52**	**1**	**7**	**2**	**1**	**2**	**3**
6	**13**	**36**	**3**	**44**	**1**	**0**	**0**	**1**	**0**	**1**
7	**22**	**56**	**5**	**57**	**1**	**3**	**0**	**5**	**0**	**4**
8	**81**	**81**	**48**	**131**	**25**	**5**	**0**	**38**	**1**	**27**
9	**1722**	**315**	**1894**	**745**	**1955**	**64**	**0**	**1769**	**19**	**1589**
Average PE	**8.521**	**3.840**	**8.914**	**5.406**	**9.001**	**1.722**	**1.251**	**8.607**	**1.443**	**8.115**

Table 2
Performance of various criteria in order selection for 2000 trials of N=35 samples of the AR process defined by Eq. (102). The entries represent the average of prediction errors that result from the orders selected by each of the criteria. The minimum and maximum candidate orders are 0 and 15, respectively.

	FPE	FPEF	AIC	AICC	AICF	KIC	KICC	BIC
Average PE	**5.561**	**2.972**	**5.644**	**3.666**	**1.285**	**5.033**	**2.773**	**4.620**

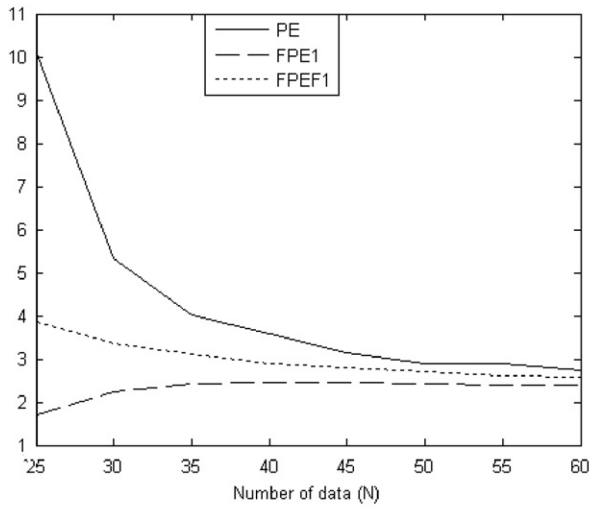

Fig. 6. Mean values of the true prediction error (PE), FPE1, and FPEF1, for the VAR process defined by Eq. (101). The mean values are obtained using 2000 simulation runs of N samples of the VAR process and the order q=7 is assumed for the data generated by the VAR process.

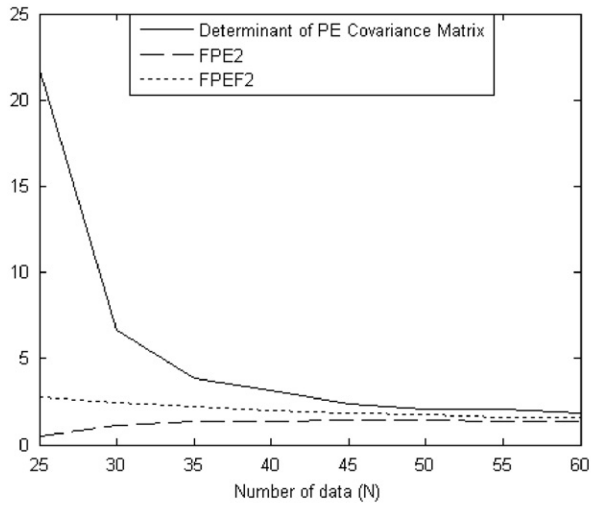

Fig. 7. Mean values of the determinant of the true prediction error (PE) covariance matrix, FPE2, and FPEF2, for the VAR process defined by Eq. (101). The mean values are obtained using 2000 simulation runs of N samples of the VAR process and the order q=7 is assumed for the data generated by the VAR process.

$N-q$ grows and other parameters are fixed, FPEF1(q) and FPEF2(q) become more successful in estimating the prediction error and the determinant of the prediction error covariance matrix, respectively. Fig. 6 gives plots of mean values of FPEF1(q), FPE1(q) and true prediction error for the process defined by Eq. (101). In addition, Fig. 7 gives plots of mean values of FPEF2(q), FPE2(q) and the determinant of the prediction error covariance matrix for this VAR process. In these figures the number of data N is varied from 25 to 60 but the order q considered for the process is fixed and is equal to 7 in all cases. The plotted mean values are obtained using 2000 simulation runs of the VAR process. These two figures show that when $N-q$ grows, the performance of the estimators of the prediction error and the determinant of the prediction error covariance matrix becomes better. In these two figures the performance of FPEF1 and FPEF2 criteria is always better than FPE1 and FPE2 criteria, respectively.

9. Concluding remarks

The results presented in the previous sections suggest that, when the least-squares method is used for parameter estimation, FPEF should replace FPE and AICF should replace AIC and AICC. Simulation results indicate that these replacements cause a great performance improvement in the finite sample case. In the large sample case, FPEF is equivalent to FPE, and AICF is equivalent to AIC and AICC.

It should be noted that FPEF and AICF are asymptotically efficient since they are equivalent to the asymptotically efficient criteria FPE and AIC, respectively, in the large sample case. FPEF and AICF are not consistent as the consistency of an order selection criterion can be obtained only at the cost of efficiency. Usually, of the two properties, efficiency is preferred. The order selection criteria FPE, AIC, AICC, KIC, and KICC are all asymptotically efficient, while BIC is consistent.

Appendix A

In this appendix it is shown that each vector $\mathbf{z}_{1,i,j}$ defined by Eq. (16) can be written as a linear combination of the members of S.

It follows from Eq. (12) that

$$\begin{cases} \mathbf{z}_{q+1,i}(j) = \dfrac{-1}{\sigma_{i,j}}\{\mathbf{F}_{i,1}(j,:)\mathbf{x}_q + \ldots + \mathbf{F}_{i,i-1}(j,:)\mathbf{x}_{q+2-i} + g_{i,1}\mathbf{x}_{q+1-i}(1) + \ldots + g_{i,j-1}\mathbf{x}_{q+1-i}(j-1) - \mathbf{x}_{q+1-i}(j)\} \\ \mathbf{z}_{q+2,i}(j) = \dfrac{-1}{\sigma_{i,j}}\{\mathbf{F}_{i,1}(j,:)\mathbf{x}_{q+1} + \ldots + \mathbf{F}_{i,i-1}(j,:)\mathbf{x}_{q+3-i} + g_{i,1}\mathbf{x}_{q+2-i}(1) + \ldots + g_{i,j-1}\mathbf{x}_{q+2-i}(j-1) - \mathbf{x}_{q+2-i}(j)\} \\ \vdots \\ \mathbf{z}_{N,i}(j) = \dfrac{-1}{\sigma_{i,j}}\{\mathbf{F}_{i,1}(j,:)\mathbf{x}_{N-1} + \ldots + \mathbf{F}_{i,i-1}(j,:)\mathbf{x}_{N-i+1} + g_{i,1}\mathbf{x}_{N-i}(1) + \ldots + g_{i,j-1}\mathbf{x}_{N-i}(j-1) - \mathbf{x}_{N-i}(j)\} \end{cases} ; \quad i=1,\ldots,q, \quad j=1,\ldots,m \tag{A1}$$

It can be seen from Eqs. (13), (16), and (A1) that

$$\begin{aligned} \mathbf{z}_{1,i,j} = [\mathbf{z}_{q+1,i}(j)\ \mathbf{z}_{q+2,i}(j) \ldots \mathbf{z}_{N,i}(j)]^T = \frac{-1}{\sigma_{i,j}} \Bigg\{ & \mathbf{F}_{i,1}(j,1)\begin{bmatrix}\mathbf{x}_q(1)\\ \mathbf{x}_{q+1}(1)\\ \vdots\\ \mathbf{x}_{N-1}(1)\end{bmatrix} + \ldots + \mathbf{F}_{i,1}(j,m)\begin{bmatrix}\mathbf{x}_q(m)\\ \mathbf{x}_{q+1}(m)\\ \vdots\\ \mathbf{x}_{N-1}(m)\end{bmatrix} + \ldots + \mathbf{F}_{i,i-1}(j,1)\begin{bmatrix}\mathbf{x}_{q+2-i}(1)\\ \mathbf{x}_{q+3-i}(1)\\ \vdots\\ \mathbf{x}_{N-i+1}(1)\end{bmatrix} \\ & + \ldots + \mathbf{F}_{i,i-1}(j,m)\begin{bmatrix}\mathbf{x}_{q+2-i}(m)\\ \mathbf{x}_{q+3-i}(m)\\ \vdots\\ \mathbf{x}_{N-i+1}(m)\end{bmatrix} + g_{i,1}\begin{bmatrix}\mathbf{x}_{q+1-i}(1)\\ \mathbf{x}_{q+2-i}(1)\\ \vdots\\ \mathbf{x}_{N-i}(1)\end{bmatrix} + \ldots + g_{i,j-1}\begin{bmatrix}\mathbf{x}_{q+1-i}(j-1)\\ \mathbf{x}_{q+2-i}(j-1)\\ \vdots\\ \mathbf{x}_{N-i}(j-1)\end{bmatrix} - \begin{bmatrix}\mathbf{x}_{q+1-i}(j)\\ \mathbf{x}_{q+2-i}(j)\\ \vdots\\ \mathbf{x}_{N-i}(j)\end{bmatrix} \Bigg\} \\ = \frac{-1}{\sigma_{i,j}} \{ & \mathbf{F}_{i,1}(j,1)\mathbf{x}_{q,1} + \ldots + \mathbf{F}_{i,1}(j,m)\mathbf{x}_{q,m} + \ldots + \mathbf{F}_{i,i-1}(j,1)\mathbf{x}_{q+2-i,1} + \ldots + \mathbf{F}_{i,i-1}(j,m)\mathbf{x}_{q+2-i,m} \\ & + g_{i,1}\mathbf{x}_{q+1-i,1} + \ldots + g_{i,j-1}\mathbf{x}_{q+1-i,j-1} - \mathbf{x}_{q+1-i,j}\}; \quad i=1,\ldots,q, \quad j=1,\ldots,m \end{aligned} \tag{A2}$$

Note that if $i=1$ the terms containing the coefficients $\mathbf{F}_{i,k}(j,l)$ are omitted from Eqs. (A1) and (A2). On the other hand, if $j=1$ the terms containing the coefficients $g_{i,k}$ are omitted from these equations. It can be seen from Eqs. (A2) and (14) that each vector $\mathbf{z}_{1,i,j}$ defined by Eq. (16) can be written as a linear combination (with constant coefficients) of the members of S.

Appendix B

In this appendix, it is shown that $\mathbf{Z}$ can be written as the product of $\mathbf{X}$ and a nonsingular matrix.

If we write Eq. (A2) for $i=1, 2, \ldots, q$ and $j=1, 2, \ldots, m$, we obtain mq equations that can be collected in the following matrix equation:

$$\begin{aligned} &\begin{bmatrix}\mathbf{z}_{1,1,1} & \mathbf{z}_{1,1,2} & \cdots & \mathbf{z}_{1,1,m} & \mathbf{z}_{1,2,1} & \cdots & \mathbf{z}_{1,2,m} & \cdots & \mathbf{z}_{1,q,1} & \cdots & \mathbf{z}_{1,q,m}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{x}_{q,1} & \mathbf{x}_{q,2} & \cdots & \mathbf{x}_{q,m} & \mathbf{x}_{q-1,1} & \cdots & \mathbf{x}_{q-1,m} & \cdots & \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,m}\end{bmatrix} \end{aligned}$$

$$
\times \begin{bmatrix}
\frac{1}{\sigma_{1,1}} & \frac{-g_{1,1}}{\sigma_{1,2}} & \cdots & \frac{-g_{1,1}}{\sigma_{1,m}} & \frac{-\mathbf{F}_{2,1}(1,1)}{\sigma_{2,1}} & \frac{-\mathbf{F}_{2,1}(2,1)}{\sigma_{2,2}} & \cdots & \frac{-\mathbf{F}_{2,1}(m,1)}{\sigma_{2,m}} & \cdots & \frac{-\mathbf{F}_{q,1}(1,1)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,1}(m,1)}{\sigma_{q,m}} \\
0 & \frac{1}{\sigma_{1,2}} & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\vdots & 0 & \ddots & \frac{-g_{1,m-1}}{\sigma_{1,m}} & & & & & & & & \\
 & \vdots & \ddots & \frac{1}{\sigma_{1,m}} & \frac{-\mathbf{F}_{2,1}(1,m)}{\sigma_{2,1}} & \frac{-\mathbf{F}_{2,1}(2,m)}{\sigma_{2,2}} & \cdots & \frac{-\mathbf{F}_{2,1}(m,m)}{\sigma_{2,m}} & \cdots & \frac{-\mathbf{F}_{q,1}(1,m)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,1}(m,m)}{\sigma_{q,m}} \\
 & & & 0 & \frac{1}{\sigma_{2,1}} & \frac{-g_{2,1}}{\sigma_{2,2}} & \cdots & \frac{-g_{2,1}}{\sigma_{2,m}} & \cdots & \frac{-\mathbf{F}_{q,2}(1,1)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,2}(m,1)}{\sigma_{q,m}} \\
 & & & \vdots & 0 & \frac{1}{\sigma_{2,1}} & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
 & & & & \vdots & 0 & \ddots & \frac{-g_{2,m-1}}{\sigma_{2,m}} & \ddots & & & \\
 & & & & & \vdots & \ddots & \frac{1}{\sigma_{2,m}} & \ddots & \frac{-\mathbf{F}_{q,2}(1,m)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,2}(m,m)}{\sigma_{q,m}} \\
 & & & & & & & 0 & & \vdots & & \vdots \\
 & & & & & & & \vdots & \ddots & \frac{-\mathbf{F}_{q,q-1}(1,1)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,q-1}(m,1)}{\sigma_{q,m}} \\
 & & & & & & & & \ddots & \vdots & & \vdots \\
 & & & & & & & & \ddots & \frac{-\mathbf{F}_{q,q-1}(1,m)}{\sigma_{q,1}} & \cdots & \frac{-\mathbf{F}_{q,q-1}(m,m)}{\sigma_{q,m}} \\
 & & & & & & & & & \frac{1}{\sigma_{q,1}} & \cdots & \frac{-g_{q,1}}{\sigma_{q,m}} \\
 & & & & & & & & & & & \vdots \\
 & & & & & & & & & 0 & \ddots & \frac{-g_{q,m-1}}{\sigma_{q,m}} \\
 & & & & & & & & & \vdots & \ddots & \frac{1}{\sigma_{q,m}} \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots &
\end{bmatrix} \tag{B1}
$$

The first and second matrices in Eq. (B1) are **Z** and **X**, respectively. The third matrix is an upper triangular $mq \times mq$ matrix we have named **F**. So Eq. (B1) can be written as follows:

$$\mathbf{Z} = \mathbf{XF} \tag{B2}$$

It is obvious that

$$\det(\mathbf{F}) = 1/(\sigma_{1,1}\sigma_{1,2}\ldots\sigma_{1,m}\sigma_{2,1}\ldots\sigma_{2,m}\ldots\sigma_{q,1}\ldots\sigma_{q,m}) = 1\Bigg/\left(\prod_{i=1}^{q}\left(\prod_{j=1}^{m}\sigma_{i,j}\right)\right) > 0 \tag{B3}$$

So, **F** is nonsingular.

Appendix C

In this appendix, the relations (37), (39), and (40) are derived.

Relation (36) states that

$$\hat{\Sigma}(i,j) = \frac{1}{N-q}(\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i})^T(\mathbf{w}_{q+1,j}-\hat{\mathbf{w}}_{q+1,j}); \quad q \geq p \tag{C1}$$

The term $\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i}$ is the error in projecting $\mathbf{w}_{q+1,i}$ onto M. This error is orthogonal to the members of M, and $\hat{\mathbf{w}}_{q+1,j}$ is a member of M. Therefore, we have

$$\langle(\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i}),\hat{\mathbf{w}}_{q+1,j}\rangle = (\mathbf{w}_{q+1,i}-\hat{\mathbf{w}}_{q+1,i})^T\hat{\mathbf{w}}_{q+1,j} = 0; \quad i,j = 1,\ldots,m \tag{C2}$$

It follows that

$$\mathbf{w}_{q+1,i}^T\hat{\mathbf{w}}_{q+1,j} = \hat{\mathbf{w}}_{q+1,i}^T\hat{\mathbf{w}}_{q+1,j}; \quad i,j = 1,\ldots,m \tag{C3}$$

Substituting Eqs. (C2) and (C3) into Eq. (C1), and using Eq. (29), we get

$$\begin{aligned}\hat{\Sigma}(i,j) &= \frac{1}{N-q}(\mathbf{w}_{q+1,i}^T\mathbf{w}_{q+1,j}-\hat{\mathbf{w}}_{q+1,i}^T\mathbf{w}_{q+1,j}) = \frac{1}{N-q}(\mathbf{w}_{q+1,i}^T\mathbf{w}_{q+1,j}-\hat{\mathbf{w}}_{q+1,i}^T\hat{\mathbf{w}}_{q+1,j}) \\ &= \frac{1}{N-q}\left\{\left[\sum_{k=1}^{N-q}\mathbf{w}_{q+k}(i)\mathbf{w}_{q+k}(j)\right]-\hat{\mathbf{c}}_i^T\mathbf{Z}^T\mathbf{Z}\hat{\mathbf{c}}_j\right\}; \quad q \geq p\end{aligned} \tag{C4}$$

We define the matrix $\hat{\mathbf{C}}$ as follows:

$$\hat{\mathbf{C}} = \left[\, \hat{\mathbf{c}}_1 \cdots \quad \hat{\mathbf{c}}_m \,\right] \tag{C5}$$

Using Eqs. (C4) and (C5), we have

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-q}\left\{\left[\sum_{k=1}^{N-q} \mathbf{w}_{q+k}\mathbf{w}_{q+k}^T\right] - \hat{\mathbf{C}}^T\mathbf{Z}^T\mathbf{Z}\hat{\mathbf{C}}\right\}; \quad q \geq p \tag{C6}$$

Now, our goal is to approximate the term $\mathbf{Z}^T\mathbf{Z}$ in Eq. (C4). It follows from Eq. (18) that the element in the ith row and jth column of $\boldsymbol{Z}^T\boldsymbol{Z}$ is equal to

$$(\mathbf{Z}^T\mathbf{Z})_{i,j} = \mathbf{z}_{1,k_1,l_1}^T \mathbf{z}_{1,k_2,l_2} = \sum_{k=q+1}^{N} \mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2) \tag{C7}$$

where k_1, k_2, l_1, and l_2 are defined as follows:

$$k_1 = \left\lceil \frac{i-1}{m} \right\rceil + 1, \quad l_1 = i-(k_1-1)m, \quad k_2 = \left\lceil \frac{j-1}{m} \right\rceil + 1, \quad l_2 = j-(k_2-1)m \tag{C8}$$

where $\lceil\,\rceil$ indicates the integer part. It follows from Eqs. (8) and (9) that

$$E[\mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)] = \begin{cases} 0; & i \neq j\,(k_1 \neq k_2 \quad \text{or} \quad l_1 \neq l_2) \\ 1; & i = j\,(k_1 = k_2 \quad \text{and} \quad l_1 = l_2) \end{cases} \tag{C9}$$

Using Eqs. (C7) and (C9), we obtain

$$E[(\mathbf{Z}^T\mathbf{Z})_{i,j}] = \sum_{k=q+1}^{N} E[\mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)] = \begin{cases} 0; & i \neq j \\ N-q; & i = j \end{cases} \tag{C10}$$

It follows that

$$E[\mathbf{Z}^T\mathbf{Z}] = (N-q)\mathbf{I}_{mq} \tag{C11}$$

As we saw in Section 2, for each i and j the sequence of random variables $\mathbf{z}_{n,i}(j)$ is an ergodic process with time variable n. Thus, we have

$$E\left[\mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)\right] = \lim_{L\to\infty}\left\{\frac{1}{2L+1}\sum_{k=-L}^{L} \mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)\right\} \tag{C12}$$

It follows that for large $N-q$ (i.e., $N-q \gg 1$)

$$E\left[\mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)\right] \approx \frac{1}{N-q}\sum_{k=q+1}^{N} \mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2) \tag{C13}$$

Henceforth, we will always assume in this paper that $N-q$ is large enough and Eq. (C13) is valid. It follows from Eqs. (C7), (C9), and (C13) that

$$(\mathbf{Z}^T\mathbf{Z})_{i,j} = \sum_{k=q+1}^{N} \mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2) \approx (N-q)E[\mathbf{z}_{k,k_1}(l_1)\mathbf{z}_{k,k_2}(l_2)] = \begin{cases} 0; & i \neq j \\ N-q; & i = j \end{cases} \tag{C14}$$

So, we have

$$\mathbf{Z}^T\mathbf{Z} \approx (N-q)\mathbf{I}_{mq} \tag{C15}$$

Appendix D

In this appendix, the relation (50) is derived.
It can be seen from Eqs. (22), (29), and (30) that

$$\mathbf{x}_{q+1,i} - \mathbf{X}\hat{\mathbf{A}}_i = \mathbf{x}_{q+1,i} - \hat{\mathbf{x}}_{q+1,i} = \mathbf{w}_{q+1,i} - \hat{\mathbf{w}}_{q+1,i} = \mathbf{w}_{q+1,i} - \mathbf{Z}\hat{\mathbf{c}}_i; \quad q \geq p \tag{D1}$$

We want to derive a relation similar to Eq. (D1) for $\mathbf{y}_i$'s. Relation (21) can be rewritten as follows:

$$\hat{\mathbf{x}}_{q+1,i} = \mathbf{Z}\hat{\mathbf{b}}_i, \quad \hat{\mathbf{b}}_i = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{x}_{q+1,i}; \quad i = 1,2,\ldots,m \tag{D2}$$

We define $\hat{\mathbf{y}}_{q+1,i}$ as the orthogonal projection of $\mathbf{y}_{q+1,i}$ onto the closed span of columns of $\boldsymbol{Z}'$. Thus, similar to Eq. (D2), it can be written as follows:

$$\hat{\mathbf{y}}_{q+1,i} = \mathbf{Z}'\hat{\mathbf{b}}'_i, \quad \hat{\mathbf{b}}'_i = (\mathbf{Z}'^T\mathbf{Z}')^{-1}\mathbf{Z}'^T\mathbf{y}_{q+1,i}; \quad i = 1,2,\ldots,m \tag{D3}$$

Similar to Eqs. (B2), (28), and (29), we have

$$\mathbf{Z}' = \mathbf{YF} \tag{D4}$$

$$\hat{\mathbf{y}}_{q+1,i} = \mathbf{YA}_i + \hat{\mathbf{w}}'_{q+1,i}; \quad i = 1,2,\ldots,m, \quad q \geq p \tag{D5}$$

$$\hat{\mathbf{w}}'_{q+1,i} = \mathbf{Z}'\hat{\mathbf{c}}'_i\,; \qquad i = 1,\,2,\,\ldots,\ m \tag{D6}$$

The matrix $\mathbf{F}$ is nonsingular and it follows from Eq. (D4) that

$$\mathbf{Y} = \mathbf{Z}'\mathbf{F}^{-1} \tag{D7}$$

It follows from Eqs. (D5)–(D7) that

$$\hat{\mathbf{y}}_{q+1,i} = \mathbf{YA}_i + \hat{\mathbf{w}}'_{q+1,i} = \mathbf{Z}'\mathbf{F}^{-1}\mathbf{A}_i + \mathbf{Z}'\hat{\mathbf{c}}'_i = \mathbf{Z}'(\mathbf{F}^{-1}\mathbf{A}_i + \hat{\mathbf{c}}'_i)\,; \qquad q \geq p \tag{D8}$$

Similar to $\boldsymbol{Z}^T\boldsymbol{Z}$, it can be seen that $\mathbf{Z}'^T\mathbf{Z}'$ is nonsingular. So, it follows from the projection theorem for Euclidean spaces [20] that the vector $\hat{\mathbf{b}}'_i$ in Eq. (D3) is unique. Thus, it can be seen from Eqs. (D3) and (D8) that

$$\hat{\mathbf{b}}'_i = \mathbf{F}^{-1}\mathbf{A}_i + \hat{\mathbf{c}}'_i; \quad q \geq p \tag{D9}$$

It can be shown in a quite similar way that

$$\hat{\mathbf{b}}_i = \mathbf{F}^{-1}\mathbf{A}_i + \hat{\mathbf{c}}_i\,; \qquad q \geq p \tag{D10}$$

It follows from Eqs. (D4), (D2), (B2), and from the nonsingularity of $\mathbf{F}$ and $\mathbf{X}^T\mathbf{X}$ that

$$\mathbf{Z}'\hat{\mathbf{b}}_i = \mathbf{YF}[(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{x}_{q+1,i}] = \mathbf{YF}[(\mathbf{F}^T\mathbf{X}^T\mathbf{XF})^{-1}\mathbf{F}^T\mathbf{X}^T\mathbf{x}_{q+1,i}] = \mathbf{Y}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}_{q+1,i} = \mathbf{Y}\hat{\mathbf{A}}_i \tag{D11}$$

Similar to Eq. (26), it is obvious that we can write

$$\mathbf{y}_{q+1,i} = \mathbf{YA}_i + \mathbf{w}'_{q+1,i}; \quad q \geq p \tag{D12}$$

Using Eqs. (D11), (D12), (D7), and (D10), we obtain

$$\mathbf{y}_{q+1,i} - \mathbf{Y}\hat{\mathbf{A}}_i = \mathbf{y}_{q+1,i} - \mathbf{Z}'\hat{\mathbf{b}}_i = (\mathbf{YA}_i + \mathbf{w}'_{q+1,i}) - \mathbf{Z}'\hat{\mathbf{b}}_i = (\mathbf{Z}'\mathbf{F}^{-1}\mathbf{A}_i + \mathbf{w}'_{q+1,i}) - \mathbf{Z}'(\mathbf{F}^{-1}\mathbf{A}_i + \hat{\mathbf{c}}_i) = \mathbf{w}'_{q+1,i} - \mathbf{Z}'\hat{\mathbf{c}}_i; \quad q \geq p \tag{D13}$$

Using the definition of $\boldsymbol{y}_{q+1,i}$, it can be seen from Eq. (48) that

$$\mathbf{U} = \begin{bmatrix} \mathbf{y}_{q+1,1} & \mathbf{y}_{q+1,2} & \cdots & \mathbf{y}_{q+1,m} \end{bmatrix} \tag{D14}$$

In addition, we can use Eqs. (48), (47), the definition of $\mathbf{Y}$, and the relation between $\hat{\boldsymbol{\Phi}}_i$'s and $\hat{\mathbf{A}}_i$'s to obtain

$$\hat{\mathbf{U}} = \begin{bmatrix} \mathbf{Y}\hat{\mathbf{A}}_1 & \mathbf{Y}\hat{\mathbf{A}}_2 & \cdots & \mathbf{Y}\hat{\mathbf{A}}_m \end{bmatrix} \tag{D15}$$

It follows from Eqs. (D13)-(D15) that

$$\mathbf{U} - \hat{\mathbf{U}} = \begin{bmatrix} \mathbf{y}_{q+1,1} - \mathbf{Y}\hat{\mathbf{A}}_1 \cdots & \mathbf{y}_{q+1,m} - \mathbf{Y}\hat{\mathbf{A}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{w}'_{q+1,1} & \cdots & \mathbf{w}'_{q+1,m} \end{bmatrix} - \mathbf{Z}'\begin{bmatrix} \hat{\mathbf{c}}_1 \cdots & \hat{\mathbf{c}}_m \end{bmatrix}; \quad q \geq p \tag{D16}$$

Appendix E

In this appendix, the relation (69) is derived.
It follows from Eq. (65) that

$$E\{\hat{\mathbf{c}}_{i_1}(j_1)\hat{\mathbf{c}}_{i_2}(j_2)\} \approx \frac{1}{(N-q)^2}\sum_{k=1}^{N-q}\sum_{s=1}^{N-q} E\{\mathbf{z}_{q+k,l_1}(h_1)\mathbf{w}_{q+k}(i_1)\mathbf{z}_{q+s,l_2}(h_2)\mathbf{w}_{q+s}(i_2)\} = \frac{1}{(N-q)^2}\sum_{k=1}^{N-q}\sum_{s=1}^{N-q} E\{d_{k,s,j_1,j_2}\} \tag{E1}$$

where

$$l_1 = \left\lceil \frac{j_1-1}{m} \right\rceil + 1, \quad h_1 = j_1-(l_1-1)m, \quad l_2 = \left\lceil \frac{j_2-1}{m} \right\rceil + 1, \quad h_2 = j_2-(l_2-1)m \tag{E2}$$

and

$$d_{k,s,j_1,j_2} = \mathbf{z}_{q+k,l_1}(h_1)\mathbf{w}_{q+k}(i_1)\mathbf{z}_{q+s,l_2}(h_2)\mathbf{w}_{q+s}(i_2) \tag{E3}$$

In the case that $k>s$, the discussions following Eq. (52) indicate that $\mathbf{w}_{q+k}(i_1)$ is independent of $\mathbf{z}_{q+k,l1}(h_1)$ and $\mathbf{z}_{q+s,l_2}(h_2)$. In addition, $\mathbf{w}_{q+k}(i_1)$ is independent of $\mathbf{w}_{q+s}(i_2)$ in this case. Thus, we have

$$E\{d_{k,s,j_1,j_2}\} = E\{\mathbf{w}_{q+k}(i_1)\}E\{\mathbf{z}_{q+k,l_1}(h_1)\mathbf{z}_{q+s,l_2}(h_2)\mathbf{w}_{q+s}(i_2)\} = 0; \quad k>s \tag{E4}$$

In the case that $s>k$, it can be shown, similarly, that $\mathbf{w}_{q+s}(i_2)$ is independent of the other three variables in the right-hand side of Eq. (E4). Thus, we have

$$E\{d_{k,s,j_1,j_2}\} = E\{\mathbf{w}_{q+s}(i_2)\}E\{\mathbf{z}_{q+k,l_1}(h_1)\mathbf{z}_{q+s,l_2}(h_2)\mathbf{w}_{q+k}(i_1)\} = 0; \quad s>k \tag{E5}$$

In the case that k=s, $\mathbf{w}_{q+k}(i_1)$ and $\mathbf{w}_{q+k}(i_2)$ are independent of $\mathbf{z}_{q+k,l_1}(h_1)$ and $\mathbf{z}_{q+k,l_2}(h_2)$. So it follows from Eqs. (8), (9), and (E2) that

$$E\{d_{k,s,j_1,j_2}\} = E\{\mathbf{w}_{q+k}(i_1)\mathbf{w}_{q+k}(i_2)\}E\{\mathbf{z}_{q+k,l_1}(h_1)\mathbf{z}_{q+k,l_2}(h_2)\} = \boldsymbol{\Sigma}(i_1,i_2)E\{\mathbf{z}_{q+k,l_1}(h_1)\mathbf{z}_{q+k,l_2}(h_2)\} = \begin{cases} 0; & k=s, \quad j_1 \neq j_2 \\ \boldsymbol{\Sigma}(i_1,i_2); & k=s, \quad j_1 = j_2 \end{cases} \tag{E6}$$

Substituting Eqs. (E4)–(E6) into (E1), we obtain

$$E\{\hat{\mathbf{c}}_{i_1}(j_1)\hat{\mathbf{c}}_{i_2}(j_2)\} \approx \begin{cases} 0; & j_1 \neq j_2 \\ \dfrac{1}{(N-q)^2}\displaystyle\sum_{k=1}^{N-q}\boldsymbol{\Sigma}(i_1,i_2) = \dfrac{1}{(N-q)}\boldsymbol{\Sigma}(i_1,i_2); & j_1 = j_2 \end{cases} \tag{E7}$$

References

[1] S.L. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[2] P.M.T. Broersen, Vector autoregressive order selection in practice, IEEE Trans. Instrum. Meas. 58 (8) (2009) 2565–2573.
[3] A. Schogl, A comparison of multivariate autoregressive estimators, Signal Process. 86 (2006) 2426–2429.
[4] A. Mahmoudi, M. Karimi, Estimation of the parameters of multichannel autoregressive signals from noisy observations, Signal Process 88 (2008) 2777–2783.
[5] Md.K. Hasan, Md.J. Hossain, Md.A. Haque, Parameter estimation of multichannel autoregressive processes in noise, Signal Process. 83 (2003) 603–610.
[6] S. Mousazadeh, M. Karimi, Estimating multivariate ARCH parameters by two-stage least squares method, Signal Process 89 (2009) 921–932.
[7] H. Akaike, Autoregressive model fitting for control, Ann. Inst. Stat. Math 23 (1971) 163–180.
[8] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Proceedings of the Second International Symposium on Information Theory, Budapest, Hungary, pp. 267–281, 1973.
[9] C.M. Hurvich, C.L. Tsai, A corrected Akaike information criterion for vector autoregressive model selection, J. Time Series Anal. 14 (3) (1993) 271–279.
[10] J.E. Cavanaugh, A large-sample model selection criterion based on Kullback's symmetric divergence, Stat. Probab. Lett. 42 (1999) 333–343.
[11] A.K. Seghouane, Vector autoregressive model-order selection from finite samples using Kullback's symmetric divergence, IEEE Trans. Circuits Syst. I 53 (10) (2006) 2335–2337.
[12] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[13] N.J. Hsu, H.L. Hung, Y.M. Chang, Subset selection for vector autoregressive processes using Lasso, Comput. Stat. Data Anal. 52 (2008) 3645–3657.
[14] C. Gatu, E.J. Kontoghiorghes, M. Gilli, P. Winker, An efficient branch-and-bound strategy for subset vector autoregressive model selection, J. Econ. Dyn. Control 32 (2008) 1949–1963.
[15] M. Karimi, A corrected FPE criterion for autoregressive processes, in: Proceedings of the 15th European Signal Processing Conference (EUSIPCO), Poznan, Poland, 3–7 September, 2007.
[16] M. Karimi, Finite sample AIC for autoregressive model order selection, in: Proceedings of the IEEE International Conference on Signal Processing and Communications (ICSPC), Dubai, United Arab Emirates, 24–27 November, 2007.
[17] M. Karimi, Finite sample criteria for autoregressive model order selection, Iran. J. Sci. Technol. Trans. B: Eng. 31 (B3) (2007) 329–344.
[18] M. Karimi, On the residual variance and the prediction error for the LSF estimation method and new modified finite sample criteria for autoregressive model order selection, IEEE Trans. Signal Process. 53 (7) (2005) 2432–2441.
[19] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Univ. Press, 1990.
[20] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd ed., Springer-Verlag, NY, 1991.