PERGAMON

Automatica 36 (2000) 1605–1615

automatica

www.elsevier.com/locate/automatica

Bootstrap-based estimates of uncertainty in subspace identification methods☆

Sergio Bittanti, Marco Lovera*

Dipartimento di Elettronica e Informazione, Politecnico di Milano, 32, Piazza Leonardo da Vinci, 20133 Milano, Italy

Received 6 May 1998; revised 9 June 1999; received in final form 14 February 2000

The application of computational statistics provides simple algorithms for assessing model uncertainty in subspace identification.

Abstract

The problem of evaluating model uncertainty in the framework of subspace identification methods is considered and the use of the methods of computational statistics is proposed. In particular, a procedure for the estimation of the standard error of the eigenvalues and of the frequency response of the identified models is presented. Simuation results demonstrate that accurate estimates of the standard error can be obtained by this approach. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Identification algorithms; Subspace methods; Statistical analysis; Uncertainty

1. Introduction

The recently developed class of subspace identification methods has attracted an increasing attention in the last few years (Viberg, 1995; Verhaegen, 1994; Van Overschee & De Moor, 1996; Lovera, 1998). Among the main advantages of such methods we mention the numerical robustness and efficiency, the ability to deal with MIMO problems in a straightforward way and the ease of use for the non-specialist due to the small number of parameters which have to be chosen by the user.

Work remains to be done, however, in the statistical analysis of such algorithms. In particular, while it is possible to prove consistency for most of the existing methods (see Bauer, Deistler & Scherrer, 1995), the optimality of some algorithms has been only demonstrated by example (Larimore, 1994) and no proof of efficiency has been established so far (at least to the best knowledge of the authors). In addition, very little work has been done as far as the evaluation of model uncertainty is concerned, mainly because of the complexity of the statistical theory underlying SMI. At present, results concerning the asymptotic distribution of the pole estimates (Viberg, Ottersten, Wahlberg & Ljung, 1991, 1993; Bauer & Jansson, 2000) are available, and a recent analysis of the limit distribution for the estimates of the system matrices (Bauer, Deistler & Scherrer, 1999; Bauer & Jansson, 2000) has been carried out.

The aim of this paper is to present the results obtained in the application of the methods of computational statistics to the problem of evaluating the uncertainty of the identified models. The main contribution consists in a procedure for the evaluation of the standard deviation of the frequency response of the estimated models. The statistical tool we will resort to is the bootstrap method (Efron & Tibshirani, 1993; Shao & Tu, 1995).

The paper is organized as follows: first some background on subspace identification methods (with particular reference to the MOESP class of algorithms, Verhaegen, 1994; Verhaegen and Dewilde, 1992) is provided; then the bootstrap method for the estimation of the standard error is introduced and its existing variants for time-series analysis are presented. Finally,

☆This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor B. Ninness under the direction of Editor T. Söderström.

* Corresponding author. Tel.: +39-02-23993592; fax +39-02-23993412.

E-mail address: lovera@elet.polimi.it (M. Lovera).

PII: S0005-1098(00)00081-9

the proposed estimation procedure is outlined and the simulation results are described and discussed.

2. Subspace model identification

2.1. The output error problem

Consider the MIMO discrete-time LTI system

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t, \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t, \end{aligned} \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^n$, $\mathbf{u}_t \in \mathbb{R}^m$, $\mathbf{y}_t \in \mathbb{R}^l$ and $\mathbf{v}_t \in \mathbb{R}^l$ is a zero-mean white noise process with covariance matrix $V = \sigma^2 I$.

The problem of estimating a quadruple $[\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$ from N samples of $\mathbf{u}_t$ and $\mathbf{y}_t$ can be faced as follows.

With the given data, form the block Hankel matrices $\mathbf{Y}_{t,i,j}$ and $\mathbf{U}_{t,i,j}$:

$$\mathbf{Y}_{t,i,j}(li, j) = \begin{bmatrix} \mathbf{y}_t & \mathbf{y}_{t+1} & \cdots & \mathbf{y}_{t+j-1} \\ \mathbf{y}_{t+1} & \mathbf{y}_{t+2} & \cdots & \mathbf{y}_{t+j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t+i-1} & \mathbf{y}_{t+i} & \cdots & \mathbf{y}_{t+i+j-2} \end{bmatrix} \tag{2}$$

and similarly for $\mathbf{U}_{t,i,j}$ and $\mathbf{V}_{t,i,j}$.

Then it is easy to show that system (1) can be written as

$$\mathbf{Y}_{t,i,j} = \mathbf{\Gamma}_i \mathbf{X}_{t,j} + \mathbf{H}_i \mathbf{U}_{t,i,j} + \mathbf{V}_{t,i,j}, \tag{3}$$

where Γ_i is the extended observability matrix for the system, $\mathbf{X}_{t,j}$ is formed by consecutive state vectors $\mathbf{x}$:

$$\mathbf{X}_{t,j}(n, j) = [\mathbf{x}_t \quad \mathbf{x}_{t+1} \quad \cdots \quad \mathbf{x}_{t+j-1}],$$

$\mathbf{H}_i$ is the block Toeplitz matrix:

$$\mathbf{H}_i(li, mi) = \begin{bmatrix} \mathbf{D} & 0 & 0 & \cdots & 0 \\ \mathbf{CB} & \mathbf{D} & 0 & \cdots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{i-2}\mathbf{B} & \mathbf{CA}^{i-3}\mathbf{B} & \mathbf{CA}^{i-4}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix}. \tag{4}$$

It can be proved that starting from Eq. (3), under the stated assumption for $\mathbf{v}_t$ and appropriate persistency of excitation assumptions on $\mathbf{u}_t$, consistent estimates of $\mathbf{\Gamma}_i$ can be computed.

The algorithm described herein (ordinary MOESP, see Verhaegen & Dewilde, 1992) estimates the state space matrices of system (1) in two distinct stages. In the first stage, the orthogonal projection of the rows of $\mathbf{Y}$ onto the subspace $[\mathrm{span}_{\mathrm{row}}(\mathbf{U})]^{\perp}$ is computed:

$$\mathbf{Y}_{t,i,j}\mathbf{\Pi}^{\perp} = \mathbf{\Gamma}_i \mathbf{X}_{t,j}\mathbf{\Pi}^{\perp} + \mathbf{V}_{t,i,j}\mathbf{\Pi}^{\perp}, \tag{5}$$

where

$$\mathbf{\Pi}^{\perp} = \mathbf{I} - \mathbf{U}_{t,i,j}^{\dagger \mathrm{T}} \mathbf{U}_{t,i,j}. \tag{6}$$

Then by taking the SVD of the projected data matrix $\mathbf{Y}_{t,i,j}\mathbf{\Pi}^{\perp}$ and provided that $\mathbf{U}_{t,i,j}$ is full rank:

$$\mathbf{Y}_{t,i,j}\mathbf{\Pi}^{\perp} = [\hat{\mathbf{U}}_n \quad \hat{\mathbf{U}}_n^{\perp}] \begin{bmatrix} \hat{\mathbf{S}}_n & 0 \\ 0 & \hat{\mathbf{S}}_n^{\perp} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_n^{\mathrm{T}} \\ (\hat{\mathbf{V}}_n^{\perp})^{\mathrm{T}} \end{bmatrix}, \tag{7}$$

one has that the subspace estimate is then given by $\hat{\mathbf{U}}_n$, i.e., by the n principal left singular vectors of $\mathbf{Y}_{t,i,j}\mathbf{\Pi}^{\perp}$. The estimates of the $\mathbf{A}$ and $\mathbf{C}$ matrices are then given by

$$\hat{\mathbf{A}} = (\mathbf{J}_1 \hat{\mathbf{U}}_n)^{\dagger} \mathbf{J}_2 \hat{\mathbf{U}}_n, \quad \hat{\mathbf{C}} = \text{first 1 rows of } \hat{\mathbf{U}}_n, \tag{8}$$

where

$$\mathbf{J}_1 = [\mathbf{I}_{(l-1)i} \quad \mathbf{0}_{(l-1)i,i}], \qquad \mathbf{J}_2 = [\mathbf{0}_{(l-1)i,i} \quad \mathbf{I}_{(l-1)i}]. \tag{9}$$

The results of the first stage make it possible to set up a least-squares problem from which $\mathbf{B}$ and $\mathbf{D}$ can be estimated.

2.2. The general problem

We now turn to the general case, which is obtained by adding in the state equation in (1) an additional white process noise term $\mathbf{w}_t$, possibly correlated with $\mathbf{v}_t$. Then the new data equation is given by

$$\mathbf{Y}_{t,i,j} = \mathbf{\Gamma}\mathbf{X}_{t,j} + \mathbf{H}_i \mathbf{U}_{t,i,j} + \mathbf{E}_i \mathbf{W}_{t,i,j} + \mathbf{V}_{t,i,j},$$

where matrix $\mathbf{E}_i$ has a structure similar to $\mathbf{H}_i$ (see Verhaegen, 1994 for details).

The idea in this case (see, e.g., Van Overschee, 1994; Verhaegen, 1994; Cho & Kailath, 1995), is to split the data equation in two parts, a "past" one and a "future" one.

The algorithm described herein (MOESP PO, see Verhaegen, 1994) is based on the use of both the past inputs and past outputs as instrumental variables. Again we start by forming block Hankel matrices from the input/output data: $\mathbf{Y}_p$ and $\mathbf{U}_p$ defined as in (2) with j columns and i block rows; $\mathbf{Y}_f$ and $\mathbf{U}_f$, which are structurally identical to (2) but formed from the elements of sequences $\mathbf{u}_t$ and $\mathbf{y}_t$ shifted i samples ahead in time.

Given these matrices, the following relationships hold:

$$\begin{aligned} \mathbf{Y}_p &= \mathbf{\Gamma}_i \mathbf{X}_p + \mathbf{H}_i \mathbf{U}_p + \mathbf{E}_i \mathbf{W}_p + \mathbf{V}_p, \\ \mathbf{Y}_f &= \mathbf{\Gamma}_i \mathbf{X}_f + \mathbf{H}_i \mathbf{U}_f + \mathbf{E}_i \mathbf{W}_f + \mathbf{V}_f, \end{aligned} \tag{10}$$

where $\mathbf{W}_p$ and $\mathbf{V}_p$ are matrices formed from the noise sequences $\mathbf{w}_t$ and $\mathbf{v}_t$ starting from time t, while $\mathbf{W}_f$ and $\mathbf{V}_f$ are in turn Hankel matrices formed from sequences $\mathbf{w}_t$ and $\mathbf{v}_t$ starting from time $t + i$ and $\mathbf{X}_p$ and $\mathbf{X}_f$ are defined accordingly.

As a first step in the algorithm, the orthogonal projection of the rows of $\mathbf{Y}_f$ onto the subspace spanned by the rows of the IV matrix is computed, by means of the following RQ factorization:

$$\begin{bmatrix}\mathbf{U}_f\\ \mathbf{U}_p\\ \mathbf{Y}_p\\ \mathbf{Y}_f\end{bmatrix}=\begin{bmatrix}\mathbf{R}_{11}^{(mi,mi)} & 0 & 0 & 0\\ \mathbf{R}_{21}^{(mi,mi)} & \mathbf{R}_{22}^{(mi,mi)} & 0 & 0\\ \mathbf{R}_{31}^{(li,mi)} & \mathbf{R}_{32}^{(li,mi)} & \mathbf{R}_{33}^{(li,li)} & 0\\ \mathbf{R}_{41}^{(li,mi)} & \mathbf{R}_{42}^{(li,mi)} & \mathbf{R}_{43}^{(li,li)} & \mathbf{R}_{44}^{(li,li)}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_1^{(mi,j)}\\ \mathbf{Q}_2^{(mi,j)}\\ \mathbf{Q}_3^{(li,j)}\\ \mathbf{Q}_4^{(li,j)}\end{bmatrix}. \quad (11)$$

Finally, by taking a SVD of $[\mathbf{R}_{42}\ \mathbf{R}_{43}]$, a consistent estimate of the (rank n) column space of $\mathbf{\Gamma}_i$ can be retrieved. From the estimate of $\mathrm{span}_{\mathrm{col}}(\mathbf{\Gamma}_i)$ it is finally possible to compute $\mathbf{A}$ and $\mathbf{C}$, just as in the orthogonal projection algorithm.

3. Previous work on the uncertainty analysis of subspace methods

With reference to the problem of assessing model uncertainty in SMI, only two studies have been performed so far: the first one (see Viberg et al., 1991, 1993) aimed specifically at deriving closed-form expressions for the uncertainty of the eigenvalues of the estimated state-space models, both for the ordinary and for the IV version of the basic subspace algorithm; the second one (see Bauer et al., 1999; Bauer & Jansson, 2000) developed the asymptotic theory for subspace identification algorithms, in the purely stochastic case (exogeneous input $\mathbf{u}_t$ missing) and for the MOESP class of algorithms.

In this section we will give an overview of such results.

3.1. The statistical analysis of Viberg et al.

In view of (8), one has that the error committed in estimating $\mathbf{A}$ can be written as

$$\tilde{\mathbf{A}}=\hat{\mathbf{A}}-\mathbf{A}=(\mathbf{J}_1\hat{\mathbf{U}}_n)^{\dagger}(\mathbf{J}_2\hat{\mathbf{U}}_n-\mathbf{J}_1\hat{\mathbf{U}}_n\mathbf{A})$$
$$\simeq(\mathbf{J}_1\mathbf{U}_n)^{\dagger}(\mathbf{J}_2\hat{\mathbf{U}}_n-\mathbf{J}_1\hat{\mathbf{U}}_n\mathbf{A}). \quad (12)$$

If now we indicate by $\mathbf{A}_d$ the diagonal form of $\mathbf{A}$, a matrix $\mathbf{T}$ exists such that $\mathbf{A}=\mathbf{T}^{-1}\mathbf{A}_d\mathbf{T}$. It is further possible to show that the error in the estimated matrix can be related to the error in the estimated eigenvalues as follows:

$$\tilde{\boldsymbol{\mu}}_k=\mathbf{T}_{k.}\tilde{\mathbf{A}}\mathbf{T}_{.k}^{-1}, \quad (13)$$

where by $\mathbf{T}_{k.}$ we denote the kth column of matrix $\mathbf{T}$, i.e., the kth eigenvector of $\mathbf{A}$, and by $\mathbf{T}_{.k}$ the kth row.

One can simplify the expression of the estimation errors for the eigenvalues:

$$\tilde{\boldsymbol{\mu}}_k\simeq\mathbf{J}_1\mathbf{\Gamma}_{i,k.}(\mathbf{J}_2-\boldsymbol{\mu}_k\mathbf{J}_1)\tilde{\mathbf{U}}_n\mathbf{U}_n^{*}\mathbf{\Gamma}_{i,.k} \quad (14)$$

and defining

$$\mathbf{a}_k=[(\mathbf{J}_1\mathbf{\Gamma}_{i,k.})^{\dagger}(\mathbf{J}_2-\boldsymbol{\mu}_k\mathbf{J}_1)]^{*}, \quad (15)$$

one has

$$\tilde{\boldsymbol{\mu}}_k\simeq\mathbf{a}_k^{*}\tilde{\mathbf{U}}_n\mathbf{U}_n^{*}\mathbf{\Gamma}_{i,.k}. \quad (16)$$

From the above expression for $\tilde{\boldsymbol{\mu}}$, one can derive closed-form expressions for its second moment, given by (see Viberg et al., 1991 for additional details)

$$NE[\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^{*}]=\sigma^2\sum_{p=1}^{li}\sum_{q=1}^{li}\{\mathrm{diag}(\mathbb{A}_{q.})\mathbb{B}^{*}\mathbf{\Gamma}_i\mathbf{P}(q-p)\mathbf{\Gamma}_i^{*}\mathrm{diag}(\mathbb{A}_{p.})^{*}\}^{\mathrm{T}}+\sigma^4\sum_{|p|<li}\{\mathrm{diag}(\mathbb{A}^{*}\mathbf{T}_p\mathbb{B})\,\mathrm{diag}(\mathbb{A}^{*}\mathbf{T}_p\mathbb{B})^{*}+\mathrm{diag}(\mathbb{A}^{*}\mathbb{T}_p\mathbb{B})\mathrm{diag}(\mathbb{A}^{*}\mathbb{T}_{-p}\mathbb{B})^{*}\}, \quad (17)$$

where $\mathbb{A}$ is the matrix whose columns are $\mathbf{a}_k$, $\mathbf{P}(k)$ is defined as

$$\mathbf{P}(k)=\lim_{N\to\infty}\frac{1}{N}X\mathbf{\Pi}^{\perp}\mathbb{T}_k\mathbf{\Pi}^{\perp}X^{\mathrm{T}}. \quad (18)$$

$\mathbb{T}_k$ is a square matrix of suitable dimentions the elements of which are given by $\{\mathbb{T}_k\}_{pq}=\delta_{q-p-k}$ and the columns $\mathbf{b}_k$ of $\mathbb{B}$ are defined as

$$\mathbf{b}_k=\mathbf{\Gamma}_i^{\dagger *}P_{.k}^{-1}. \quad (19)$$

In Viberg et al. (1993) the above analysis was extended to the case of IV subspace identification algorithms and similar closed-form expressions for the variance of the poles' estimates were obtained.

The analytical expressions for the variance of the pole estimates obtained by Viberg et al. have proven very accurate in various simulation studies (see the cited papers and the simulation results given later in this section).

However, the analysis which has been summarized in the previous sections is characterized by two assumptions:

(1) It is not applicable in the case of multiple eigenvalues.
(2) It relies on the Gaussianity assumption for the noise acting on the system.

In the section dedicated to simulation results a comparison between the accuracy offered by these asymptotic results and the bootstrap estimates for standard deviation will be given.

3.2. The analysis of Bauer et al.

More recently, an analysis of the asymptotic variance of subspace algorithms has been proposed in Bauer et al.

(1999). The discussion in the cited papers refers to the so-called CCA algorithm (see Bauer et al., 1995) and is confined to the case when the input acting on the system is an unobserved white noise and the system itself is asymptotically stable and minimum phase.

The above results have been recently extended to the MOESP class of subspace identification algorithms (see Bauer & Jansson, 2000). In this framework it is possible to prove, under a few additional technical assumptions, that the estimates of the matrices in the estimated state-space realization are characterized by a multivariate standard normal limiting distribution.

This result implies a consistency proof, together with variance expressions for the estimated state-space matrices.

4. The bootstrap

The bootstrap is a statistical method which was originally introduced in order to solve the following problem (Efron & Tibshirani, 1993; Shao & Tu, 1995): given a random i.i.d. sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$ drawn from an unknown distribution F, one computes an estimate $\hat{\theta}$ of the parameter $\theta = t(F) = t[\mathbf{x}]$ on the basis of the available data, and would like to assess the accuracy of the obtained estimate, in terms of its standard deviation or its variance.

In general, the variance of $\hat{\theta}$ will be given by

$$\mathrm{var}(\hat{\theta}) = \int\left[t[\mathbf{x}] - \int t[\mathbf{y}]\mathrm{d}\prod_{i=1}^{n} F(y_i)\right]^2 \mathrm{d}\prod_{i=1}^{n} F(x_i). \tag{20}$$

For simple statistics, it is possible to obtain explicit expressions for $\mathrm{var}(\theta)$; this, however, is not always feasible, so one normally resorts to asymptotic results in order to simplify the analysis. For example, one can sometimes establish results of the form

$$\lim_{N\to\infty} N[\mathrm{var}(\hat{\theta})] = \sigma_F^2, \tag{21}$$

where σ_F^2 is a simple function of F.

However, when a classical asymptotic analysis is not possible or too complex, or if the asymptotic approximation holds with satisfactory accuracy only for very (too) large values of N, then it may be useful and advantageous to resort to alternative techniques.

The bootstrap is one of such methods.

From a theoretical point of view, the bootstrap principle is based on the idea of replacing the (unknown) distribution F of the data with an estimate $\hat{F}$ of it.

For the case of variance estimation, the bootstrap estimate of variance is then given by

$$\hat{\sigma}^2_{\mathrm{boot}} = \int\left[t[\mathbf{x}] - \int t[\mathbf{y}]\mathrm{d}\prod_{i=1}^{n} \hat{F}(\mathbf{y}_i)\right]^2 \mathrm{d}\prod_{i=1}^{n} \hat{F}(x_i). \tag{22}$$

Whenever $\hat{\sigma}^2_{\mathrm{boot}}$ is an explicit function of the sample x, one can compute the bootstrap estimate of variance in a direct way. Unfortunately, this situation is fairly uncommon, so that in general $\hat{\sigma}^2_{\mathrm{boot}}$ cannot be evaluated exactly and one has to make use of approximation methods. The approach which is normally followed for the computation of the estimate is a Monte Carlo one. Suppose for a moment that the distribution F is known. Then, it would be easy to estimate σ^2_{boot} by repeatedly drawing new random data sets from F, computing new estimates of θ from such data sets and approximating σ^2_{boot} with the sample variance of the obtained estimates. As F is normally not known, one has to replace it with a data-based estimate $\hat{F}$.

Therefore, the bootstrap approach to the problem of variance estimation proceeds as follows:

(1) Obtain an estimate $\hat{F}$ of F. The estimate can be either parametric or non-parametric, according to the assumptions in the statistical model underlying the analysis.
(2) From the distribution $\hat{F}$, draw (with replacement) B samples of size n each, which we indicate by $\mathbf{x}^{*(i)}$, $i = 1, \dots, B$.
(3) From each of the obtained samples, compute a *replication* of the estimate of the statistic of interest, $\hat{\theta}^{*(i)}$, $i = 1, \dots, B$.
(4) The estimate of standard error is then given by

$$\hat{\sigma}_{\mathrm{boot}} = \frac{1}{\sqrt{B-1}}\left(\sum_{i=1}^{B} (\hat{\theta}^{*(i)} - \hat{\theta}^{*(\cdot)})^2\right)^{1/2}, \tag{23}$$

where

$$\hat{\theta}^{*(\cdot)} = \frac{\sum_{i=1}^{B}(\hat{\theta}^{*(i)})}{B}. \tag{24}$$

When no a priori assumptions on F can be made, $\hat{F}$ is defined as the empirical discrete distribution, obtained by associating a probability of $1/n$ to each of the data points in sample x. On the other hand, whenever prior information on F is available, it is also possible to use a parametric approach for the estimation of F, by assuming that it belongs to some specified class of probability distributions. For example, assuming that F is Gaussian, it can be determined in a straightforward way from the available data by estimating the corresponding sufficient statistics (mean and variance).

The bootstrap is now a well-established tool in statistics, as it allows to work out numerical solutions to many problems which cannot be given a proper analytical treatment.

In particular, the consistency of the bootstrap procedure can be proved for a large class of statistics of practical interest (see Shao & Tu, 1995). It should also be mentioned that the bootstrap is not restricted in scope to the

problem of variance estimation as discussed herein, but has a much wider applicability.

Unfortunately, there is a price to be paid for the generality of the bootstrap method, which is represented by the following facts:

(1) The computational cost implied by the replication of the estimation procedure. The number of replications B required for the bootstrap estimate of standard error to converge is usually very large (see Shao & Tu, 1995, for a discussion of the problem of choosing B): values of B from at least 100 on are a common indication in the literature, so that the application of the bootstrap can be considered in practice only if one is capable of computing the estimate $\hat{\theta}$ in a very efficient way. As will be demonstrated in the following, the numerical efficiency of subspace identification algorithms can compensate for this requirement.
(2) The loss of the insight which might be provided even by an approximate asymptotic analysis.

5. Bootstrapping dynamic models

The definition of the boostrap given in the previous section clearly applies only to the case of i.i.d. samples.

When dealing with more complicated data structures such as time series and dynamic models, different schemes have to be derived in order to obtain correct results from a bootstrap procedure. This is indeed necessary as the bootstrap has been developed to deal with i.i.d. data, while this is certanly not the case in, e.g., a time-series analysis problem. Various approaches have been proposed in order to apply the bootstrap for variance estimation in time-series analysis (Shao & Tu, 1995; Tjarnstrom, 1999) and signal processing (Shamsunder, 1998). In particular, two main classes of methods can be recognized:

- Moving block methods.
- Bootstrapping residuals methods.

In the following, a brief outline of the two approaches will be given.

5.1. The moving block method

Given a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ which is known NOT to be i.i.d., one computes an estimate $\hat{\theta}$ of the parameter $\theta = t(\mathbf{x})$ on the basis of the available data, and would like to assess the standard deviation of the obtained estimate.

The moving block idea is to partition the set of available measurements into a number of overlapping blocks, which are subsequently treated as independent. Precisely, suppose that each block contains h data and define $\mathbf{B}_i = (x_i, \ldots, x_{i+h-1})$, $i = 1, \ldots, n-h+1$ as the ith data block.

Now from $\mathbf{B}_1, \ldots, \mathbf{B}_{n-h+1}$ perform B i.i.d extractions of k blocks each $\mathbf{B}_1^*, \ldots, \mathbf{B}_k^*$ and compute B times the estimate θ_l^* of the statistic θ based on the $l = kh$ data points contained in $\mathbf{B}_1^*, \ldots, \mathbf{B}_k^*$. Then we have for the estimate of standard error the following expression:

$$\hat{\sigma} = \frac{1}{\sqrt{B-1}} \left(\sum_{i=1}^{B} (\hat{\theta}_l^{*(i)} - \hat{\theta}_l^{*(\cdot)})^2 \right)^{1/2}, \tag{25}$$

where

$$\hat{\theta}_l^{*(\cdot)} = \frac{\sum_{i=1}^{B} (\hat{\theta}_l^{*(i)})}{B}. \tag{26}$$

This method (which has been succesfully applied to the problem of determining accurate confidence intervals in non-parametric spectral estimation (see Politis, Romano & Lai, 1992)) is clearly based on the assumption that the correlation structure of the data makes it possible to consider each block as being independent of the others. Obviously, the choice of the block size parameter h is very critical and has to be based upon some prior knowledge about the process which generated the data.

The major disadvantage of this method lies in the required prior knowledge about the correlation structure of the data. This limitation can be at least partially overcome by means of the residuals-based method, which is outlined in the following.

5.2. The bootstrapping residuals method

In the statistical literature, the attention has focused mainly on AR and ARMA time-series models. In particular, a heuristic bootstrap procedure was first proposed in Chatterjee (1986), while a detailed proof of consistency for such a procedure can be found in Bose (1988). As in the case of the moving block method, the idea is to turn the non-i.i.d. problem into an i.i.d. one. Consider, for example, the problem of AR modelling for a stationary time series:

$$\mathbf{y}(t) = a_1 \mathbf{y}(t-1) + \cdots + a_n \mathbf{y}(t-n) + e_t, \tag{27}$$

where $e(\cdot)$ is a white noise process with probability distribution F_e.

It is a well-known fact from the asymptotic theory of optimal linear prediction that the prediction error $\varepsilon = y_t - \hat{y}(t|t-1)$ will be given by a white noise process with a (a priori unknown) probability distribution F_e.

Based on this key observation, one can devise the following bootstrapping scheme.

First perform the estimation of the parameters for the time-series model and compute the prediction error based on the estimated model.

Assuming subsequently that the sequence of the prediction error is a realization of a white noise process, use

it to approximate the distribution (either in a parametric or a non-parametric way) of the true residual. Finally, resample from the estimated distribution in order to obtain replications of the residual and feed them to the estimated model, in order to obtain the required replications for the original data set. Repeating the estimation procedure for the replicated data and computing the standard deviation of the replicated estimates completes the procedure.

The bootstrapping residuals method of dealing with time-series data has been proved to be consistent for various kinds of correlation structures. In particular, a detailed proof of consistency for the case of ARMA modelling has been given in Bose (1988), where the residual-based approach for the estimation of the standard error of ARMA parameters is analysed.

It is also clear that this idea can be readily extended from the case of time series to the case of input/output dynamic models, with only the minor modifications required in order to take into account the presence of the additional deterministic input.

6. Bootstrap-based estimates of uncertainty in SMI

With reference to AR and ARMA models, one normally wants to establish the standard error for the estimates of the model parameters directly. In the framework of subspace identification, however, the absence of a parameterization for the model class and the indetermination in the choice of the state-space basis for the obtained models direct the analysis towards confidence intervals for input/output characteristics of the model like, for example, its frequency response or towards realization independent features such as the location of its poles and zeros.

Therefore, one is looking for an estimate of the standard deviation of a prescribed statistic which is a function of the identified model.

For the present purposes and with reference to the problem of evaluating the standard deviation for the frequency response of the estimated model, the method of bootstrapping residuals can be synthesized as follows:

(1) Estimate the linear model $[\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}]$ from the available input/output data (u, y) and compute the estimate for the points of interest of its frequency response $\hat{G}(\mathrm{e}^{\mathrm{j}\omega_k})$, $k = 1, \ldots, N$.
(2) Compute the prediction error for the identified model:

$$e_t = y_t - \hat{y}_t. \tag{28}$$

(3) Obtain an estimate $\hat{F}_e$ for the distribution F_e of the prediction error. The estimate can be either parametric or non parametric. In this work only the parametric case will be considered and the normality assumption for the distribution of the residual will be made.
(4) Generate B replications $(u^{*(i)}, y^{*(i)})$ of the original data set (u, y), with $u^{*(i)} = u$ and $y^{*(i)}$ obtained by feeding the identified model $[\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}]$ with the deterministic input $u^{*(i)} = u$ and the stochastic input $\mathrm{e}^{*(i)}$, $k = 1, \ldots, B$ where $\mathrm{e}^{*(i)}$ is constructed by resampling (with replacement) from the distribution $\hat{F}_e$.
(5) Estimate B replications of the identified model and of the points of interest for the frequency response $\hat{G}^{*(i)}(\mathrm{e}^{\mathrm{j}\omega_k})$, $k = 1, \ldots, N$.
(6) The estimate of the standard error for the frequency response of the model is finally given by

$$\hat{\sigma}_{\hat{G}(\mathrm{e}^{\mathrm{j}\omega_k})} = \frac{1}{\sqrt{B-1}} \left(\sum_{i=1}^{B} (\hat{G}^{*(i)}(\mathrm{e}^{\mathrm{j}\omega_k}) - \bar{\hat{G}}^{*(i)}(\mathrm{e}^{\mathrm{j}\omega_k}))^2 \right)^{1/2}, \tag{29}$$

where

$$\bar{\hat{G}}^{*(i)}(\mathrm{e}^{\mathrm{j}\omega_k}) = \frac{\sum_{i=1}^{B} \hat{G}^{*(i)}(\mathrm{e}^{\mathrm{j}\omega_k})}{B}. \tag{30}$$

In a similar way, one can obtain estimates of the standard deviation for the poles and zeros of the estimated model.

7. Computational issues

It should be clear from the above discussion of the bootstrap that its main disadvantage lies in the large amount of computations it requires. As a matter of fact, replicating the estimation process B times can be extremely time consuming even for moderate samples sizes, unless one is capable of computing the required estimates in a very efficient way. Therefore, the reduction of the computational cost of the estimation procedure is a major issue for the practical application of the bootstrap.

When it comes to SMI, it turns out that the replication process can be sped up considerably by taking advantage of the structure of the algorithms and by considering the following crucial fact: *when working with replicated data the noise processes acting on the system are known a priori*. This fact implies that the generation of a replica of the estimate can be conceived as a two-step procedure:

(1) Generation of a noise free replica.
(2) Perturbation of the noise free replica in order to take into account the effect of noise in the estimation process.

Clearly, the noise free replica only has to be computed once. On the other hand, taking the noise into account requires only the computation of *perturbations* to the

noise free replica. Within the SMI framework, such a computation can be performed in a much cheaper way.

In this section we will therefore outline how such computational savings can be achieved for the ordinary MOESP and the PI/PO versions.

In all cases, the considered framework is the following: from the dataset u, y an estimate $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$ of the system matrices has been obtained and the corresponding sequence of the prediction error e has been computed. Now, a replica $\mathbf{e}^*$ of the prediction error is drawn and the replica $\hat{\mathbf{A}}^*$, $\hat{\mathbf{B}}^*$, $\hat{\mathbf{C}}^*$, $\hat{\mathbf{D}}^*$ of the estimate has to be computed from the data set $\mathbf{u}^* = \mathbf{u}$, $\mathbf{y}^*$, where $\mathbf{y}^*$ is related to $\mathbf{e}^*$ and $\mathbf{u}$ according to the considered model structure.

Consider the case of the ordinary MOESP algorithm. In this case the data equation is given by

$$\mathbf{Y}^*_{t,i,j} = \hat{\mathbf{\Gamma}}_i \mathbf{X}^{\mathrm{nf}}_{t,j} + \hat{\mathbf{H}}_i \mathbf{U}_{t,i,j} + \mathbf{V}^*_{t,i,j} = \mathbf{Y}^{\mathrm{nf}}_{t,i,j} + \mathbf{V}^*_{t,i,j}, \tag{31}$$

where matrices $\hat{\mathbf{\Gamma}}_i$ and $\hat{\mathbf{H}}$ are formed from the state-space representation of the estimated model and $\mathbf{Y}_{nf}$, $\mathbf{X}_{nf}$ (where the subscript nf stands for noise free) are, respectively, the output and the state sequences obtained by simulating the estimated model with input $\mathbf{u}^* = \mathbf{u}$ and neglecting the effect of noise.

The first step in the replication of the estimate is the RQ factorization of the data matrix:

$$\begin{bmatrix} \mathbf{U}_{t,i,j} \\ \mathbf{Y}^*_{t,i,j} \end{bmatrix} = \begin{bmatrix} \mathbf{R}^*_{11} & 0 \\ \mathbf{R}^*_{21} & \mathbf{R}^*_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}^*_1 \\ \mathbf{Q}^*_2 \end{bmatrix}. \tag{32}$$

This factorization can be simply related to the noise-free data matrix as

$$\begin{aligned} \begin{bmatrix} \mathbf{U}_{t,i,j} \\ \mathbf{Y}^*_{t,i,j} \end{bmatrix} &= \begin{bmatrix} \mathbf{U}_{t,i,j} \\ \mathbf{Y}^{\mathrm{nf}}_{t,i,j} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{V}^* \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_{11} & 0 \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{V}^* \end{bmatrix}. \end{aligned} \tag{33}$$

If now we express the noise-dependent term in factored form, we have

$$\begin{aligned} \begin{bmatrix} \mathbf{U}_{t,i,j} \\ \mathbf{Y}^*_{t,i,j} \end{bmatrix} &= \begin{bmatrix} \mathbf{U}_{t,i,j} \\ \mathbf{Y}^{\mathrm{nf}}_{t,i,j} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{V}^* \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_{11} & 0 \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{R}^*_{v11} & 0 \\ \mathbf{R}^*_{v21} & \mathbf{R}^*_{v22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix}, \end{aligned} \tag{34}$$

where clearly

$$\mathbf{R}^*_{v11} = 0 \quad \mathbf{R}^*_{v21} = \mathbf{V}^*\mathbf{Q}_1^{\mathrm{T}} \quad \mathbf{R}^*_{v22} = \mathbf{V}^*\mathbf{Q}_2^{\mathrm{T}}$$

and therefore

$$\mathbf{R}^*_{22} = \mathbf{R}_{22} + \mathbf{V}^*\mathbf{Q}_2^{\mathrm{T}}. \tag{35}$$

So, it is possible to determine the correction required to compute the replica of the $\mathbf{R}_{22}$ term in the factorization by projecting the noise matrix $\mathbf{V}^*$ onto the row space of $\mathbf{Q}_2$.

From the computational point of view, this leads to a significant saving: the full RQ factorization would require (using the Householder method) $2j(m+l)^2i^2 - 2(m+l)^3i^3/3$ flops, while the above scheme involves only one matrix product, $2l^2i^2j$ flops only.

For example, with $m = l = 2$, $i = 10$ and $j = 100$, the full RQ would cost approximately 480 000 flops, while the updating scheme would cost 80 000 flops. The computation of the noise free factorization is not included in the flop count as it needs be performed only once, while the other operations must be repeated B times.

A similar analysis can be carried out for the PO MOESP algorithm, leading to a comparable computaitonal saving.

Remark. It is well known that in the implementation of SMI the use of the RQ factorization is particularly efficient as the $\mathbf{Q}$ factor needs not be accumulated. It appears from the above schemes, however, that this is indeed necessary in this case. However, the reader should keep in mind that the computation of $\mathbf{Q}$ is performed *only once* and on the basis of the noise-free data. In this way, the increased computational cost is widely compensated by the saving achieved in the replications, at no expense in terms of numerical accuracy as noiseless data are involved in the computation of the first RQ factorization.

8. Simulation results

In this section, some results obtained in applying the bootstrap method to subspace identification are presented. In particular, we will first compare the bootstrap and the asymptotic formulas of Viberg et al. in the assessment of the uncertainty of the eigenvalues of the identified model.

Subsequently, the application of the bootstrap the estimation of the uncertainty in the frequency response estimate will be considered.

8.1. Two SISO systems

Consider the following first-order SISO system:

$$x_{t+1} = 0.5x_t + u_t, \tag{36}$$

$$y_t = x_t + u_t + v_t, \tag{37}$$

Table 1
Estimates of standard error for $V = 0.1$

$V = 0.1$	σ_{MC}	$\hat{\sigma}_{boot}$	$\hat{\sigma}_{asym}$
$N = 50$	0.0125	0.0131	0.0123
$N = 100$	0.0074	0.0080	0.0078
$N = 150$	0.0060	0.0056	0.0055
$N = 200$	0.0046	0.0043	0.0044

Table 2
Estimates of standard error for $V = 0.5$

$V = 0.5$	σ_{MC}	$\hat{\sigma}_{boot}$	$\hat{\sigma}_{asym}$
$N = 50$	0.0720	0.0671	0.0527
$N = 100$	0.0396	0.0458	0.0441
$N = 150$	0.0322	0.0281	0.0270
$N = 200$	0.0256	0.0247	0.0252

Table 3
Estimates of standard error for $V = 1$

$V = 1$	σ_{MC}	$\hat{\sigma}_{boot}$	$\hat{\sigma}_{asym}$
$N = 50$	0.1928	0.2673	0.1131
$N = 100$	0.0835	0.0763	0.0628
$N = 150$	0.0850	0.0826	0.0755
$N = 200$	0.0479	0.0389	0.0366

where v and u are uncorrelated zero mean Gaussian white noise processes with variance V^2 and 1, respectively.

The identification of a first-order model, using the ordinary MOESP algorithm has been performed, for three different values of the standard deviation of v ($V = 0.1, 0.5, 1$) and for various lengths of the data set ($N = 50, 100, 150, 200$), with $i = 5$. The standard deviation of the estimated eigenvalue was evaluated by means of Monte Carlo simulations and was compared with the estimates provided by the asymptotic analysis of Viberg et al. and by the bootstrap method, performing 1000 replications in both cases.

The results of this analysis are presented in Tables 1–3.

As can be seen, the asymptotic expressions and the bootstrap provide comparable estimates of the Monte Carlo standard deviation. However, the bootstrap is also applicable to other features of the model, like, e.g., the zeros of the transfer function or the frequency response.

As an example, 50 points of the frequency response of the identified model (corresponding to frequencies $k\pi/50$, $k = 1, \ldots, 50$) were computed and their standard deviation was estimated by means of a conventional Monte Carlo method and by means of the (parametric) method of bootstrapping residuals, starting from a data set of 500 points.

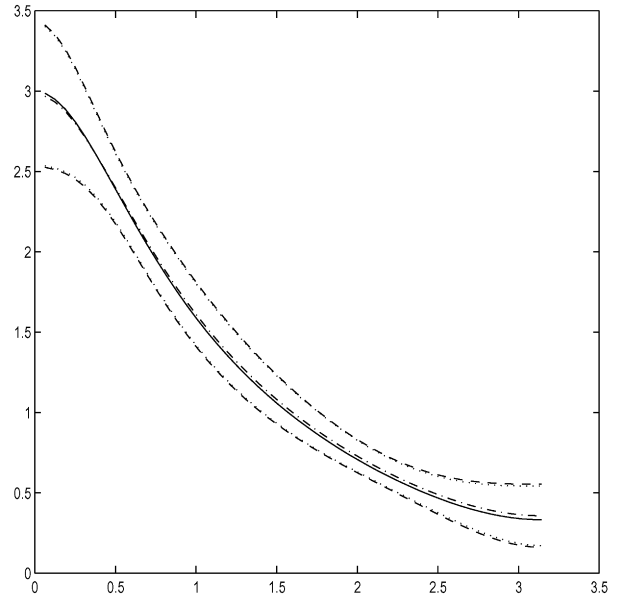

Fig. 1. True (solid line) and estimated (dash–dotted line) frequency responses, with Monte Carlo (dotted lines) and bootstrap (dashed lines) $3-\sigma$ confidence intervals.

Fig. 1 shows a comparison of the Monte Carlo and bootstrap $3-\sigma$ confidence intervals for the estimated frequency response. As can be seen the agreement is very good.

As a second example consider the second-order system:

$$x_{1,t+1} = 0.5x_{1,t} - 0.8x_{2,t} + w_t + u_t, \tag{38}$$

$$x_{2,t+1} = x_{1,t}, \tag{39}$$

$$y_t = x_{2,t} + v_t, \tag{40}$$

where $V = 1, W = 0.5$, $N = 500$ and the applied input u is again given by a zero mean white Gaussian noise with unit variance, independent from v and w.

A data set was generated from this system and was used for the identification of a second-order model, using the PO MOESP algorithm with $i = 5$.

Fig. 2 shows a comparison of the Monte Carlo and bootstrap $3-\sigma$ confidence intervals for the estimated frequency response and gives the values of the ratios $\sigma_{boot}/\sigma_{M.C.}$ for the 50 considered frequency points, computed on the basis of 500 replications: the agreement of the estimated (bootstrap) standard deviation with the "true" (Monte Carlo) one is very satisfactory.

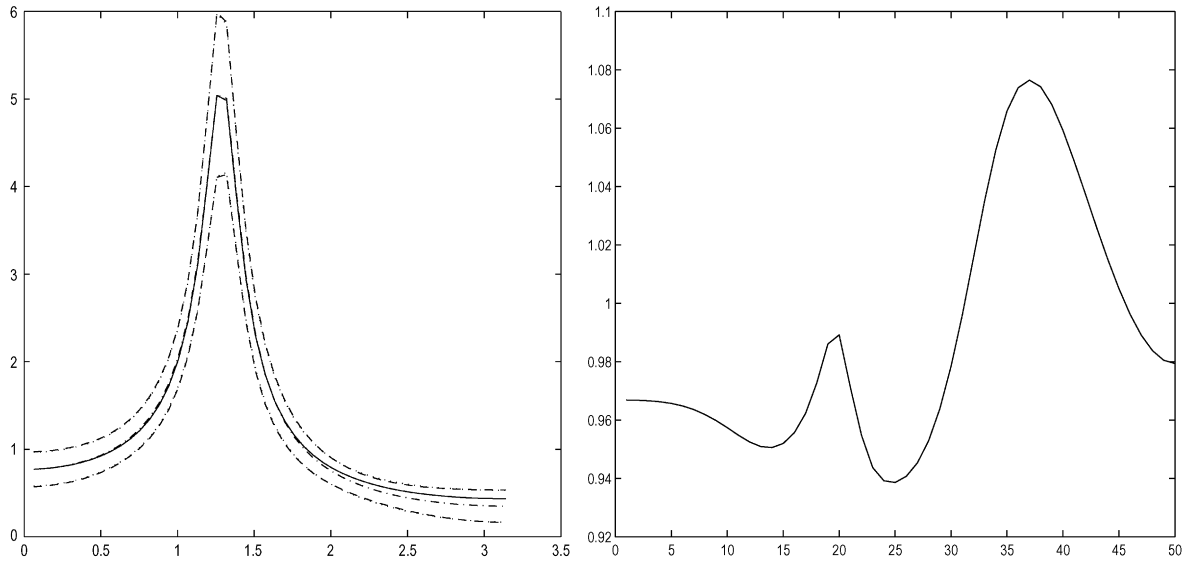

Fig. 2. Left: True (solid line) and estimated (dash–dotted line) frequency responses, with Monte Carlo (dotted lines) and bootstrap (dashed lines) $3-\sigma$ confidence intervals. Right: Ratio between the true (Monte Carlo) and estimated (bootstrap) standard error for the amplitude of the frequency response.

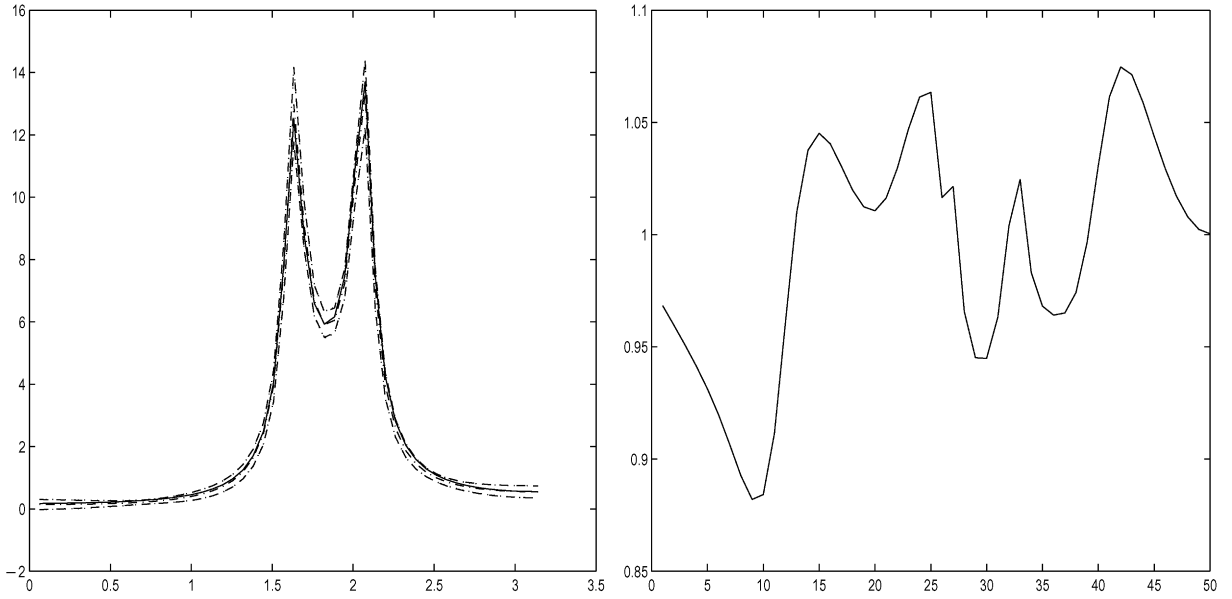

Fig. 3. Left: True (solid line) and estimated (dash–dotted line) frequency responses (output 1), with Monte Carlo (dotted lines) and bootstrap (dashed lines) $3-\sigma$ confidence intervals. Right: Ratio between the true (Monte Carlo) and estimated (bootstrap) standard error for the amplitude of the frequency response (output 1).

8.2. A MIMO system

Consider the system

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}u_t, \tag{41}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t, \tag{42}$$

where

$$\mathbf{A} = \begin{bmatrix} -1.0000 & -1.8900 & -0.9000 & -0.8100 \\ 1.0000 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \end{bmatrix}, \tag{43}$$

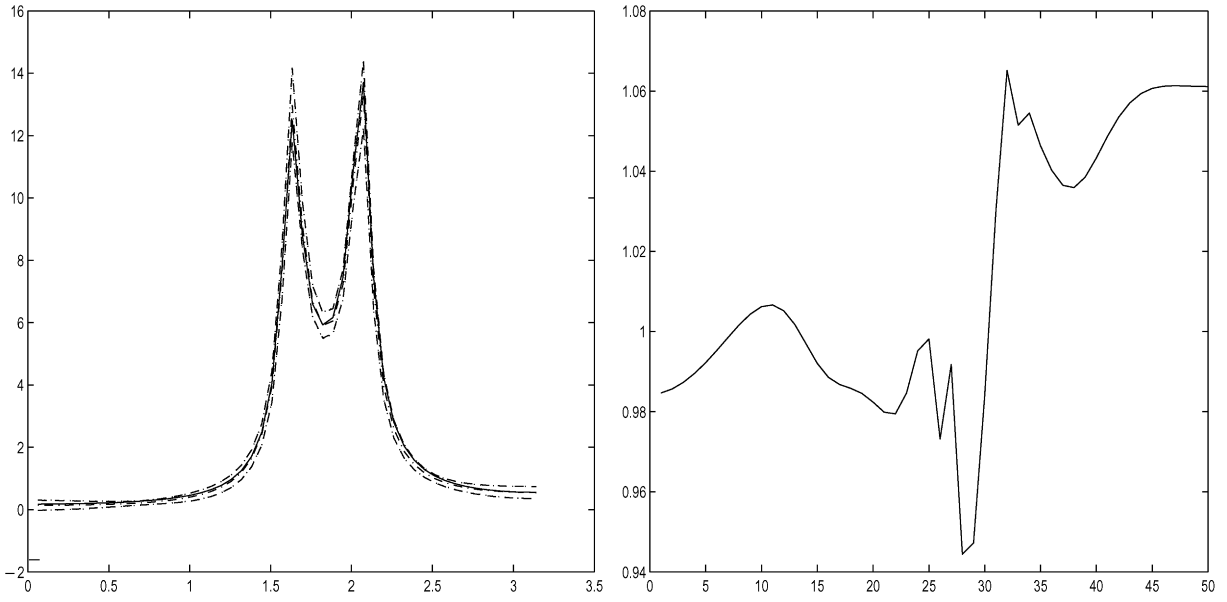

Fig. 4. Left: True (solid line) and estimated (dash–dotted line) frequency responses (output 2), with Monte Carlo (dotted lines) and bootstrap (dashed lines) $3-\sigma$ confidence intervals. Right: Ratio between the true (Monte Carlo) and estimated (bootstrap) standard error for the amplitude of the frequency response (output 2).

$$\mathbf{B} = [1 \quad 0 \quad 0 \quad 0]^{\mathrm{T}} \tag{44}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix}, \tag{45}$$

$$\mathbf{D} = [0 \quad 0]^{\mathrm{T}}, \tag{46}$$

where v_t and u are uncorrelated zero mean Gaussian white noise processes with variance I and 1, respectively.

In this case, 500 replications of the identification procedure (based on the ordinary MOESP algorithm with $i = 10$) were made.

The results obtained in the estimation of the standard error for the frequency responses of the system are given in Figs. 3 and 4.

9. Concluding remarks

The application of the bootstrap to the problem of evaluating model uncertainty in the framework of subspace identification methods is proposed and a procedure for the estimation of the standard error of the poles and zeros and of the frequency response of the identified models has been outlined. Simulation results demonstrate that accurate estimates of standard error can be obtained by this approach. Future work will aim at establishing a complete theoretical analysis, along the lines of Bose (1988).

Acknowledgements

Paper partially supported by the MURST project "Identification and control of industrial systems".

References

Bauer, D., Deistler, M., & Scherrer, W. (1995). Consistency and relative efficiency of subspace methods. *Automatica*, *31*(2), 1865–1875.

Bauer, D., Deistler, M., & Scherrer, W. (1999). Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica*, *35*(7), 1243–1254.

Bauer, D., Jansson, M. (2000). Analysis of the asymptotic properties of the MOESP type of subspace algorithms. *Automatica*, to appear.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics*, *16*(4), 1709–1722.

Chatterjee, S. (1986). Bootstrapping ARMA models: Some simulations. *IEEE Transactions on Systems, Man and Cybernetics*, *16*(2), 294–299.

Cho, Y., & Kailath, T. (1995). Fast subspace-based identification: An instrumental variable approach. *Automatica*, *31*(6), 903–905.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.

Larimore, W. (1994). The optimality of canonical variate identification by example. In *Proceedings of the 10th IFAC symposium on system identification*, Copenhagen, Denmark,

Lovera, M. (1998). *Subspace identification methods: Theory and applications*. Ph.D thesis, Politecnico di Milano.

Van Overschee, P. (1994). Subspace identification: Theory, implementation, application. Ph.D. thesis, K.U. Leuven.

Van Overschee, P., & De Moor, B. (1996). *Subspace identification: Theory, implementation, application*. Dordrecht: Kluwer Academic Publishers.

Politis, D., Romano, J., & Lai, T. (1992). Bootstrap confidence bands for spectra and cross-spectra. *IEEE Transactions on Signal Processing*, *40*(2).

Shamsunder, S. (1998). Signal processing applications of the bootstrap. *IEEE Signal Processing Magazine*, *15*, 38.

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. Berlin: Springer.

Tjarnstrom, F. (1999). *The use of bootstrap in system identification*. Technical Report LiTH-ISY-R-2113, Dept. of Electrical Engineering, Linkoping University.

Verhaegen, M. (1994). Identification of the deterministic part of MIMO state space models given in innovations form from input–output data. *Automatica*, *30*(1), 61–74.

Verhaegen, M., & Dewilde, P. (1992). Subspace model identification, part 1: Output error state space model identification class of algorithms. *International Journal of Control*, *56*(5), 1187–1210.

Viberg, M. (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, *31*(3), 1835–1852.

Viberg, M., Ottersten, B., Wahlberg, B., & Ljung, L. (1991). A statistical perspective on state-space modelling using subspace methods. Technical Report Report LiTH-ISY-I-1269, Dept. of Electrical Engineering, Linkoping University.

Viberg, M., Ottersten, B., Wahlberg, B., & Ljung, L. (1993). Performance of subspace based state space system identification methods. In *Proceedings of the 12th IFAC world congress*, vol. 7, Sydney, Australia (pp. 369–372).

Sergio Bittanti received the degree in Electronic Engineering in 1970 from the Politecnico di Milano, Italy. Since then, he has been with the Faculty of Engineering of the same university, working in the Dipartimento di Elettronica e Informazione. After starting as an assistant of Giorgio Quazza, he has been teaching various courses, such as Automatic Control, System Theory and Model Identification and Data Analysis.

Besides being active on various theoretical topics of system identification and control, such as periodic optimization, adaptive control and recursive identification, he has developed a number of application studies in cooperation with many companies.

He is the author of more than a hundred of papers in the best internationl journals, and the editor of a number of books. In particular, together with Giorgio Picci, he is editor of Identification, adaptation, learning — the science of learning models from data (Springer-Verlag, 1996), which contains a collection of selected lectures given at the NATO Advanced Study Institute from Identification to Learning organized in Como (Italy) in the summer of 1994. Moreover, together with A. J. Laub and J. C. Willems, he is editor of The Riccati Equation (Springer-Verlag, 1991), a book published after the IFAC/IEEE/SIAM workshop on The Riccati Equation in Control, Systems, and Signals organized in Como, Italy, June 1989. In addition to the already mentioned conferences, he substantially contributed to the organization of the 3rd European Control Conference (Rome, September 1995) and, together with R. E. Kalman, to the workshop on New Horizons in System Theory (Como, Italy, 1986). He will chair the IFAC Workshop on Periodic Control (Como, Italy, August 27 and 28, 2001) and the IFAC Workshop on Adaptation and Learning in Control and Signal Processing (Como, Italy, August 29–31, 2001).

He serves or served as associate editor for: Journal of Mathematical Systems, Estimation, and Control; Statistics and Computing; European Journal of Control.

For 7 years, from 1988 to 1995, he acted as national coordinator of the Italian research network Model identification, system control and signal processing, which connected about 70 Italian professor and researchers distibuted in different universities. Since 1999, he is Chairman of the IFAC Technical Committee on Adaptive Control.

Marco Lovera was born in Milan, Italy, in 1967. He obtained the Laurea degree in Electronic Engineering from the Politecnico di Milano in 1993 with a thesis on the active control of vibration in helicopters and the Ph.D. from the same University in April 1998 with a dissertation on subspace identification methods. He joined Carlo Gavazzi Space SpA (Milan, Italy) in May 1998, as control engineering specialist, working on spacecraft attitude determination and control systems and contributed to the design of the attitude control software for the Italian spacecraft MITA and AGILE. Since August 1999 he is Assistant Professor at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano. His main research interests include model identification, parameter estimation methods and advanced active control applications.