ESTIMATING THE ASYMPTOTIC VARIANCE OF CLOSED LOOP SUBSPACE ESTIMATORS

Alessandro Chiuso and Giorgio Picci *,1

* *Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy.*
e-mail:{*chiuso,picci*}*@dei.unipd.it*

Abstract: Subspace identification for closed loop systems has been recently studied by several authors. Recent results are available which express the asymptotic variance of the estimated parameters (and hence of any system invariant) as a function of the "true" underlying system parameters and of certain conditional covariance matrices.
When it comes to using these formulas in practice one is faced with the problem of computing an estimator for the variance from input-output data alone.
In this paper we discuss this problem, we propose an algorithm which computes an estimate of the variance from data alone and we show, through some simple simulation examples, how this estimate behaves as compared both to the "true" asymptotic variance and to its Monte Carlo estimate.

Keywords: Closed-loop Identification, Subspace Methods, Asymptotic Properties, Variance.

1. INTRODUCTION

It is well known that subspace methods compare very favorably to prediction error methods (PEM hereafter) for identification of linear multi-input multi-output (MIMO) stochastic models. Only very recently however (see for instance (Chiuso and Picci, 2003; Chiuso and Picci, 2005*a*; Qin and Ljung, 2003; Lin *et al.*, 2004; Larimore, 2004; Jansson, 2003; Jansson, 2005)) new subspace procedures have been proposed which can effectively deal with feedback, making them very appealing for application areas where feedback is present either due to intrinsic mechanisms or physical controllers which cannot be removed due (e.g.) to safety or production quality constraints. However, when it comes to using an estimated model for the purpose of control design or decision making (e.g fault detection), it is mandatory that some "quality tags" should be attached to the estimated model telling us how reliable the estimated model is. Obtaining practically computable expressions for the (asymptotic) variance of subspace estimates has long been recognized to be a central issue in system identification, see e.g. (Ljung, 1997).

For open-loop subspace methods recent results are available regarding the computation of the variance of the system parameter estimates (see for instance (Bauer, 2005; Chiuso and Picci, 2004; Jansson, 2000) and references therein) while, for closed loop operating conditions, expressions for the asymptotic variance of the system parameters using two recently developed subspace methods can be found in the paper (Chiuso, 2004). The expressions found in the paper (Chiuso, 2004) depend on the "true" system parameters and

[1] This work has been supported in part by the *RECSYS* project of the European Community and by the national project *New methods and algorithms for identification and adaptive control of technological systems* funded by MIUR.

on certain covariance matrices involving input, output and state process. Similar considerations hold for the formulas found in (Chiuso and Picci, 2004) and for the results reported in (Bauer, 2005).

Of course when performing an identification experiment one is only given input-output data and therefore both *parameters and variance* have to be estimated. A natural question therefore arises of assessing the reliability of the asymptotic variance formulas when using solely input-output data. In this paper we shall address this question providing both an explicit algorithm to estimate the asymptotic variance from the data and also some experimental study on simulated data concerning the reliability of these variance estimators.

Our purpose is to convince the reader that the somewhat complicated-looking formulas given in the references, can in practice be implemented by rather simple algorithms. Moreover we shall show that the computation of the asymptotic variance can indeed be done from the observed data hence allowing an effective use of the theoretical results. The outline of the paper is as follows: Section 2 we state the problem and set up notation while Section 3 recalls the variance formula derived in (Chiuso, 2004). In Section 4 we describe an algorithm to compute the variance which can be implemented using few lines of Matlab and in Section 5 we report some experimental results comparing the estimated variance with its theoretical value. We refer the reader to the work (Chiuso, 2004) for comparison between asymptotic and sample variance. Also 5% and 95% percentiles for the estimated variance obtained from Monte Carlo simulations will be reported.

2. STATEMENT OF THE PROBLEM AND NOTATION

Let $\{\mathbf{y}(t)\}, \{\mathbf{u}(t)\}$ be jointly (weakly) stationary second-order ergodic stochastic processes of dimension p and m respectively, which are representable as the output and input signals of a linear stochastic system in innovation form

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + \mathbf{e}(t) \end{cases} \quad t \geq t_0. \tag{2.1}$$

where in general there may be *feedabck* from $\{\mathbf{y}(t)\}$ to $\{\mathbf{u}(t)\}$ (Granger, 1963; Caines and Chan, 1976; Gevers and Anderson, 1982), so that we shall consider "closed loop" identification hereafter. Without loss of generality we shall assume that the dimension n of the state vector $\mathbf{x}(t)$ is as small as possible, i.e. the representation (2.1) is minimal. For simplicity we assume that $D = 0$, i.e. there is no direct feedthrough. For future reference we define $\bar{A} := A - KC$ and let $\rho := \lambda_{max}(\bar{A})$ be an eigenvalue of maximum modulus of $\bar{A}$; we shall assume that $|\rho| < 1$. We shall denote the "joint" process as $\mathbf{z} := \left[\mathbf{y}^\top \ \mathbf{u}^\top\right]^\top$ of dimension $q := p+m$.

The white noise process $\mathbf{e}$, the innovation of $\mathbf{y}$ given the joint past of $\mathbf{y}, \mathbf{u}$, is defined as the one step ahead prediction error of $\mathbf{y}(t)$ given the joint (strict) past of $\mathbf{u}$ and $\mathbf{y}$ up to time t. It variance matrix is denote with the symbol Λ_e.

The symbol I_p shall denote the identity matrix of dimension p, $A^\top$ shall indicate the transpose of the matrix A.

In general the aim of an identification experiment is to:

(a) identify the system parameters (A, B, C, K) (in a suitable canonical form), or equivalently the transfer functions $F(z) = C(zI - A)^{-1}B$ and $G(z) = C(sI - A)^{-1}K + I$, starting from input-output data $\{y_s, u_s\}$, $s \in [t_0, T + N]$, generated by the system (2.1)
(b) compute the asymptotic variance of the estimated parameters (of any system invariant to be precise) from input output data $\{y_s, u_s\}$, $s \in [t_0, T + N]$ alone.

In this paper we shall be concerned with point (b) above. In particular we shall consider the asymptotic variance of the estimated frequency response using some recently proposed closed loop subspace methods. We shall report results for the version of the "innovation estimation" algorithm by Qin and Ljung and of the "predictor based" subspace identification as described in (Chiuso and Picci, 2005*a*; Chiuso, 2004; Chiuso and Picci, 2005*b*). For reasons of space we shall not report the details of the algorithms for which we refer the interested reader to the paper (Chiuso, 2004). Moreover we shall only describe the algorithm to compute the asymptotic variance for the "predictor based" subspace identification of (Chiuso and Picci, 2005*b*), the other being completely similar [2].

We shall use the standard notation of boldface (lowercase) letters to denote random variables or semi-infinite tail sequences. Lower case letters will denote sample values of the corresponding random variables. More specifically, we shall denote by $\mathbf{y}(t)$ the output random vector at time t or the semi-infinite tail $[y_t \ y_{t+1}, \ \dots \ y_{t+k} \ \dots]$ where y_t is the sample value of $\mathbf{y}(t)$. It can be shown (see (Lindquist and Picci, 1996)) that the Hilbert spaces of second order stationary random variables and the Hilbert space of semi-infinite tails containing sample values of a (second order) stationary stochastic process are isometrically isomorphic and therefore random variables and semi-infinite tails can be regarded as being the same

[2] Matlab code is available upon request from the authors.

object. For this reason we shall use the same symbol without risk of confusion.

We shall use capitals to denote a "tail matrix" of length N. For instance $Y_t := [y_t\ y_{t+1},\ \dots\ y_{t+N-1}]$, $U_t := [y_t\ y_{t+1},\ \dots\ y_{t+N-1}]$ and $Z_t := [Y_t^\top\ U_t^\top]^\top$. These are the block rows of the usual *data Hankel matrices* which appear in subspace identification.

For $-\infty \le t_0 \le t \le T \le +\infty$ we define the Hilbert space of scalar zero-mean random variables

$$\mathcal{U}_{[t_0,\,t)} := \overline{\text{span}}\,\{\mathbf{u}_k(s);\ k=1,\dots,m,\ t_0 \le s < t\}$$

and similarly for $\mathcal{Y}_{[t_0,\,t)}$ where the bar denotes closure in mean square, i.e. in the metric defined by the inner product $\langle \xi, \eta \rangle := \mathbb{E}\{\xi\eta\}$, the operator $\mathbb{E}$ denoting mathematical expectation. These are the *past spaces* at time t of the processes $\mathbf{u}$ and $\mathbf{y}$. Similarly, let $\mathcal{U}_{[t,\,T]}$, $\mathcal{Y}_{[t,\,T]}$ be the future input and output spaces up to time T.

The *joint* future, $\mathcal{Z}_{[t,\,T]}$ and joint past $\mathcal{Z}_{[t_0,\,t)}$ spaces are defined as $\mathcal{U}_{[t,\,T]} \vee \mathcal{Y}_{[t,\,T]}$ and $\mathcal{U}_{[t_0,\,t)} \vee \mathcal{Y}_{[t_0,\,t)}$ respectively, the $\vee$ denoting closed vector sum. By convention the past spaces do not include the present. When $t_0 = -\infty$ we shall use the shorthands $\mathcal{U}_t^-$, $\mathcal{Y}_t^-$ for $\mathcal{U}_{[-\infty,\,t)}$, $\mathcal{Y}_{[-\infty,\,t)}$, and $\mathcal{Z}_t^- := \mathcal{U}_t^- \vee \mathcal{Y}_t^-$. Subspaces spanned by random variables at just one time instant (e.g. $\mathcal{U}_{[t,\,t]}$, $\mathcal{Y}_{[t,\,t]}$, etc) are simply denoted $\mathcal{U}_t$, $\mathcal{Y}_t$, etc. while for the spaces generated by $\mathbf{u}(s)$ and $\mathbf{y}(s)$, $-\infty < s < \infty$ we shall use the symbols $\mathcal{U}$, $\mathcal{Y}$, respectively. For convenience of notation we denote by $f := T - t$ the future horizon.

We shall let $\Sigma_{\mathbf{ab}} := \mathbb{E}\left[\mathbf{a}\mathbf{b}^\top\right]$ denote the covariance matrix of the random vectors [3] $\mathbf{a}$ and $\mathbf{b}$.

Given a subspace $\mathcal{C} \subseteq \mathcal{U} \vee \mathcal{Y}$, we shall denote with a slight abuse of notation, by $E[\mathbf{a} \mid \mathcal{C}]$ the orthogonal projection of the random variable $\mathbf{a}$ onto $\mathcal{A}$; in the Gaussian case the linear projection coincides with conditional expectation, i.e. $\mathbb{E}[\cdot \mid \mathcal{A}] = E[\cdot \mid \mathcal{A}]$.

In the finite dimensional case $E[\mathbf{a} \mid \mathcal{C}]$ will be given by the usual formula [4]

$$E[\mathbf{a}|\mathcal{C}] = \Sigma_{\mathbf{ac}}\Sigma_{\mathbf{cc}}^{-1}\mathbf{c}. \tag{2.2}$$

Given a subspace $\mathcal{C}$ we define the projection errors $\tilde{\mathbf{a}} := \mathbf{a} - E[\mathbf{a}|\mathcal{C}]$ and $\tilde{\mathbf{b}} := \mathbf{b} - E[\mathbf{b}|\mathcal{C}]$; the symbol $\Sigma_{\mathbf{ab|c}}$ (or sometimes also $\Sigma_{\mathbf{ab}|\mathcal{C}}$) will denote projection error covariance (conditional covariance in the Gaussian case) $\Sigma_{\mathbf{ab|c}} := \Sigma_{\tilde{\mathbf{a}}\tilde{\mathbf{b}}} = \Sigma_{\mathbf{ab}} - \Sigma_{\mathbf{ac}}\Sigma_{\mathbf{cc}}^{-1}\Sigma_{\mathbf{cb}}$.

For column vectors formed by stacking past and/or future random variables (or semi-infinite Hankel matrices) we shall use the notation: $\mathbf{y}_{[t,s]} := \left[\,\mathbf{y}^\top(t)\ \mathbf{y}^\top(t+1)\ \dots\ \mathbf{y}^\top(s)\,\right]^\top$. Finite Hankel data matrices will be denoted by capitals, i.e. $Y_{[t,s]} := \left[\,Y_t^\top\ Y_{t+1}^\top\ \dots\ Y_s^\top\,\right]^\top$

For ease of notation we shall reserve the following special symbols for the finite past and future Hankel data matrices:

$$\begin{array}{ll} Y^- := Y_{[t_0,t)} & \overline{Y}^- := Y_{[t_0,t]} \\ Y^+ := Y_{[t,T-1]} & \underline{Y}^+ := Y_{[t+1,T]} \end{array}. \tag{2.3}$$

The same notation shall be used for all signals involved (e.g. U^-,U^+,Z^-, Z^+ etc.). Spaces generated by finite tails, i.e. row spaces generated by finite Hankel data matrices, will be denoted with the same symbol used for the matrix itself. Sample covariances will be denoted with the same symbol used for the corresponding random variables with a "hat" on top. For example, given finite sequences $A_t := [a_t, a_{t+1}.., a_{t+N-1}]$ and $B_t := [b_t, b_{t+1}.., b_{t+N-1}]$ we shall define the sample covariance matrix

$$\hat{\Sigma}_{\mathbf{ab}} := \frac{1}{N}\sum_{i=0}^{N-1} a_{t+i}b_{t+i}^\top.$$

Under our ergodic assumption $\lim_{N\to\infty} \hat{\Sigma}_{\mathbf{ab}} \overset{a.s}{=} \Sigma_{\mathbf{ab}}$.

The orthogonal projection onto the row space of a matrix will be denoted by the symbol $\hat{E}$; for instance, given a matrix $C_t := [c_t, c_{t+1}, .., c_{t+N-1}]$, $\hat{E}[\cdot \mid C_t]$ will be the orthogonal projection onto the row space of the matrix C_t; the symbol $\hat{E}[A_t|C_t]$ shall denote the orthogonal projection of the rows of the matrix A_t onto the row space of C_t, and is given by the formula

$$\hat{E}[A_t|C_t] = \hat{\Sigma}_{\mathbf{ac}}\hat{\Sigma}_{\mathbf{cc}}^{-1}C_t \tag{2.4}$$

As above, given a matrix C_t, we define the projection errors $\tilde{A}_t := A_t - \hat{E}[A_t|C_t]$ and $\tilde{B}_t := B_t - \hat{E}[B_t|C_t]$. The sample covariance (conditional sample covariance) of the projection errors is denoted with the symbol $\hat{\Sigma}_{\mathbf{ab|c}} := \hat{\Sigma}_{\tilde{\mathbf{a}}\tilde{\mathbf{b}}}$ and computed by the formula

$$\hat{\Sigma}_{\mathbf{ab|c}} := \hat{\Sigma}_{\mathbf{ab}} - \hat{\Sigma}_{\mathbf{ac}}\hat{\Sigma}_{\mathbf{cc}}^{-1}\hat{\Sigma}_{\mathbf{cb}}.$$

3. PRELIMINARY RESULTS

In this section we shall provide simplified formulas for the asymptotic variance of closed loop subspace identification methods. For convenience we shall only discuss the variance expression for the "predictor based" algorithm [5]. The simplified formulas below are valid when either (i) $\rho = 0$ (i.e. the system is of the ARX type) or (ii) the past horizon $t - t_0$ goes to infinity at a suitable rate with N (see for instance (Bauer and

[3] Zero mean.
[4] Provided $\Sigma_{\mathbf{cc}}$ is invertible.
[5] Referred to as "whitening filter " algorithm in (Chiuso, 2004).

Ljung, 2002)). Denote by $\hat{A}_N, \hat{B}_N, \hat{C}_N$ the estimators with N data points and by A_N, B_N, C_N the "true" (but unknown) matrices in a suitable data dependent basis; it is shown in the literature that the basis in which A_N, B_N, C_N is represented can be chosen in such a way that $(A_N, B_N, C_N) \to (A, B, C)$ as $N \to \infty$. Define also the error $\tilde{A}_N := \hat{A}_N - A_N$; it is shown in (Chiuso, 2004) that $\sqrt{N}\text{vec}\left(\tilde{A}_N\right)$, $\sqrt{N}\text{vec}\left(\tilde{B}_N\right)$ and $\sqrt{N}\text{vec}\left(\tilde{C}_N\right)$ are jointly asymptotically normal with asymptotic covariance matrix given by the formula

$$\begin{pmatrix} M_{A1}P & M_{A2}\bar{P} \\ M_{B1}P & M_{B2}\bar{P} \\ M_C P & 0 \end{pmatrix} \Sigma \begin{pmatrix} M_{A1}P & M_{A2}\bar{P} \\ M_{B1}P & M_{B2}\bar{P} \\ M_C P & 0 \end{pmatrix}^\top \quad (3.1)$$

where:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{pmatrix}$$

and

(a) P and $\bar{P}$ are suitable permutation matrices
(b) Σ_{ij} are conditional covariance matrices of input-output data to be defined later on
(c) M_{A1}, M_{A2}, etc. are matrices which depend only on the system parameters and on some conditional covariances involving input, output, state and innovation processes.

For the extact definition of P and $\bar{P}$ we refer the reader to (Chiuso, 2004), while we report in the sequel the quantities in points (b) and (c) above.

Let $\bar{\Gamma} := \left[C^\top, \bar{A}^\top C^\top, .., \left(\bar{A}^{f-1}\right)^\top C^\top\right]^\top$ and $M_{x1} := \left(K[I\ 0] - A\bar{\Gamma}^{-L}\right)$, $M_{x2} := \bar{\Gamma}^{-L}$, $M_y := \left([I\ 0] - C\bar{\Gamma}^{-L}\right)$, where $^{-L}$ denotes the left inverse $\bar{\Gamma}^{-L} := \left(\bar{\Gamma}^\top W^\top W \bar{\Gamma}\right)^{-1} \bar{\Gamma}^\top W^\top W$ for a suitable weighting matrix W.

Using the symbol $\hat{\mathbf{x}}(t)$ for the Kalman state $E\left[\mathbf{x}(t) \mid \mathcal{Z}_{[t_0,t)}\right]$ and defining the transient innovation $\hat{\mathbf{e}}(t) := \mathbf{y}(t) - E\left[\mathbf{y}(t) \mid \mathcal{Z}_{[t_0,t)}\right]$ and $\mathbf{g}(t) := [\mathbf{u}^\top(t)\ \hat{\mathbf{e}}^\top(t)]^\top$ we have

$$\begin{aligned} M_{A1} &:= \left[\left(\Sigma^{-1}_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{g}}\ \Sigma^\top_{\mathbf{z}^-\hat{\mathbf{x}}|\mathbf{g}}\right) \otimes M_{x1}\right] \\ M_{A2} &:= \left[\left(\Sigma^{-1}_{\hat{\mathbf{x}}\hat{\mathbf{x}}|\mathbf{g}}\ \Sigma^\top_{\bar{\mathbf{z}}^-\hat{\mathbf{x}}|\mathbf{g}}\right) \otimes M_{x2}\right] \\ M_{B1} &:= \left[\left(\Sigma^{-1}_{\mathbf{uu}|(\hat{\mathbf{x}},\hat{\mathbf{e}})}\ \Sigma^\top_{\mathbf{z}^-\mathbf{u}|(\hat{\mathbf{x}},\hat{\mathbf{e}})}\right) \otimes M_{x1}\right] \\ M_{B2} &:= \left[\left(\Sigma^{-1}_{\mathbf{uu}|(\hat{\mathbf{x}},\hat{\mathbf{e}})}\ \Sigma^\top_{\bar{\mathbf{z}}^-\mathbf{u}|(\hat{\mathbf{x}},\hat{\mathbf{e}})}\right) \otimes M_{x2}\right] \\ M_C &:= \left[\left(\Sigma^{-1}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}\ \Sigma^\top_{\mathbf{z}^-\hat{\mathbf{x}}}\right) \otimes M_y\right] \end{aligned} \quad (3.2)$$

In order to define the matrices Σ_{ij} in (3.1) we need to introduce [6] the following quantities

$$\mathbf{z}^{zk}_{[t_0+\tau,t+\tau)} := E\left[\mathbf{z}_{[t_0+\tau,t+\tau)} \mid \mathcal{Z}^\perp_{[t+\tau,t+\tau+k]}\right],$$

$$\bar{\mathbf{z}}^{zk}_{[t_0+\tau,t+1+\tau)} := E\left[\mathbf{z}_{[t_0+\tau,t+1+\tau)} \mid \mathcal{Z}^\perp_{[t+1+\tau,t+\tau+k]}\right],$$

and let

$$\Sigma_{\mathbf{z}^{zk}\mathbf{z}^{zh}}(\tau) := E\left[\mathbf{z}^{zk}_{[t_0+\tau,t+\tau)}\left(\mathbf{z}^{zh}_{[t_0,t)}\right)^\top\right].$$

Similarly we introduce $\Sigma_{\bar{\mathbf{z}}^{zk}\bar{\mathbf{z}}^{zh}}(\tau)$ and $\Sigma_{\mathbf{z}^{zk}\bar{\mathbf{z}}^{zh}}(\tau)$. Let also define $\tau := h - k$, $Q_k := \left(\Sigma^{-1}_{\mathbf{zz}|\mathbf{z}^k} \otimes I_p\right)$ and [7] $\bar{Q}_k := \left(\Sigma^{-1}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\underline{\mathbf{z}}^k} \otimes I_p\right)$.

With this notation the matrices Σ_{11}, Σ_{22} and Σ_{12} have $f \times f$ blocks given, for $h = 1, .., f$, $k = 1, .., f$, by

$$\Sigma_{11}(k,h) := Q_{k-1}\left(\Sigma_{\mathbf{z}^{z(k-1)}\mathbf{z}^{z(h-1)}}(\tau) \otimes \Lambda_e\right)(Q_{h-1})^\top \quad (3.3)$$

$$\Sigma_{22}(k,h) := \bar{Q}_k\left(\Sigma_{\bar{\mathbf{z}}^{z(k)}\bar{\mathbf{z}}^{z(h)}}(\tau) \otimes \Lambda_e\right)\left(\bar{Q}_h\right)^\top \quad (3.4)$$

and

$$\Sigma_{12}(k,h) := Q_{k-1}\left(\Sigma_{\mathbf{z}^{z(k-1)}\bar{\mathbf{z}}^{z(h)}}(\tau) \otimes \Lambda_e\right)\left(\bar{Q}_h\right)^\top. \quad (3.5)$$

4. ALGORITHM

The purpose of this section is to present and algorithm to estimate the quantities in formula (3.1) from data. We shall assume that the following quantities are given as a byproduct of the subspace identification algorithm:

(a) estimates of the system matrices $\hat{A}, \hat{B}, \hat{C}, \hat{K}$ and of the innovation variance $\hat{\Lambda}_e$
(b) the finite sequences $\hat{X}_t$ (state) $\hat{E}_t := Y_t - \hat{E}\left[Y_t \mid Z_{[t_0,t)}\right]$ (transient innovation)
(c) the QR decompositions [8]

$$Z := \begin{bmatrix} Z^+ \\ Z^- \end{bmatrix} = R \cdot Q \quad \overline{Z} := \begin{bmatrix} \underline{Z}^+ \\ \overline{Z}^- \end{bmatrix} = \overline{R} \cdot \overline{Q} \quad (4.1)$$

Let also Ind and $\overline{Ind}$ be set of indexes so that, using Matlab notation, $Z(Ind,:) = Z^-$ and $\overline{Z}(\overline{Ind},:) = \overline{Z}^-$. Then, for $k = 0, .., f-2$, define the finite tails of length [9] $N - f$:

$$\begin{aligned} Z^k &:= R(Ind, k*q+1 : (f-1)*q)* \\ &\quad *Q(k*q+1 : (f-1)*q, f-k : N-k-1) \\ \overline{Z}^k &:= \overline{R}(\overline{Ind}, k*q+1 : f*q)* \\ &\quad *\overline{Q}(k*q+1 : f*q, f-k-1 : N-k-2) \end{aligned}$$

and their "normalized" versions:

$$Z^k_n := \left[\left(\frac{Z^k\left(Z^k\right)^\top}{N-f}\right)^{-1} \otimes I_p\right]\left(Z^k \otimes \hat{\Lambda}_e^{1/2}\right)$$

$$\overline{Z}^k_n := \left[\left(\frac{\overline{Z}^k\left(\overline{Z}^k\right)^\top}{N-f}\right)^{-1} \otimes I_p\right]\left(\overline{Z}^k \otimes \hat{\Lambda}_e^{1/2}\right)$$

[6] The superscript ^{zk} is meant to remind that the process $\mathbf{z}$ is taken conditionally on the k lags of the joint future.

[7] Note that $\Sigma_{\mathbf{zz}|\mathbf{z}^k} = \Sigma_{\mathbf{z}^{zk}\mathbf{z}^{zk}}(0)$.

[8] At the price of some complication one can compute just one QR decomposition, however for ease of exposition we prefer to assume both decomposition in (4.1) are available.

[9] Finite tails here will be shorter than N since we need to consider shifted version of the data.

Then Σ can be estimated as:

$$\hat{\Sigma} := \frac{1}{N-f} \begin{bmatrix} Z_n^0 \\ \vdots \\ Z_n^{f-2} \\ \overline{Z}_n^0 \\ \vdots \\ \overline{Z}_n^{f-2} \end{bmatrix} * \begin{bmatrix} Z_n^0 \\ \vdots \\ Z_n^{f-2} \\ \overline{Z}_n^0 \\ \vdots \\ \overline{Z}_n^{f-1} \end{bmatrix}^\top$$

Remark 4.1 The estimation of Σ needs to be done with some care. For instance, the naive approach of estimating all the block matrices $\Sigma_{ij}(h,k)$ and then putting them together to form $\hat{\Sigma}_{ij}$ would not guarantee that the estimated $\hat{\Sigma}$ is positive definite. Note instead the we have constructed $\hat{\Sigma}$ by taking a "square" which guarantees positive semi-definiteness. $\diamond$

The task to estimate M_{A1}, M_{A2} etc. is simpler as one just need to substitute true parameters (A, B, C, K, Λ_e) with their estimates $(\hat{A}, \hat{B}, \hat{C}, \hat{K}, \hat{\Lambda}_e)$ and replace the conditional covariance matrices with their sample version which are easily computable using the Hankel data matrices Z^-, $\bar{Z}^-$, U_t and the state and innovation sequences $\hat{X}_t$, $\hat{E}_t$ which are available as a byproduct of the identification experiment as mentioned in point (b) above. Note that the finite tail corresponding to $\mathbf{g}(t)$ can be defined as $G_t := \begin{bmatrix} U_t^\top & \hat{E}_t^\top \end{bmatrix}^\top$.

5. EXPERIMENTS

We consider two systems ($i = 1, 2$) in state space form

$$\begin{cases} \mathbf{x}(t+1) = A_i \mathbf{x}(t) + B_i \mathbf{u}(t) + K_i \mathbf{e}(t) \\ \mathbf{y}(t) = C_i \mathbf{x}(t) + \mathbf{e}(t) \end{cases}$$

where $\mathbf{e}(t)$ is unit variance white noise.

Example 1 is a first order system $A_1 = 0.9$, $B_1 = 1$, $C_1 = 1.4$, $K_1 = 1$, with a proportional controller $\mathbf{u}(t) = 5\mathbf{r}(t) - \mathbf{y}(t)$; the reference signal $\mathbf{r}(t)$ is unit variance white noise.

Example 2 instead is a fifth order (marginally stable) system

$$A_2 = \begin{bmatrix} 4.40 & 1 & 0 & 0 & 0 \\ -8.09 & 0 & 1 & 0 & 0 \\ 7.83 & 0 & 0 & 1 & 0 \\ -4 & 0 & 0 & 0 & 1 \\ 0.86 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0.00098 & 0.01299 & 0.01859 & 0.0033 & -0.00002 \end{bmatrix}^\top$$

$$C_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 2.3 & -6.64 & 7.515 & -4.0146 & 0.86336 \end{bmatrix}^\top$$

controlled with $\mathbf{u}(t) = \mathbf{r}(t) - H(z)\mathbf{y}(t)$ where

$$H(z) = \frac{0.63 - 2.083z^{-1} + 2.8222z^{-2} - 1.865z^{-3} + 0.4978z^{-4}}{1 - 2.65z^{-1} + 3.11z^{-2} - 1.75z^{-3} + 0.39z^{-4}}$$

	$t - t_0$	f	N
Example 1	10	5	1000
Example 2	25	15	2000

Table 1. Simulation setup.

and again $\mathbf{r}(t)$ is unit variance white noise. The experiment is carried out as follows: 1000 Monte Carlo runs are repeated using the parameters reported in Table 1. For each run the variances of the transfer functions are estimated using the algorithm in Section 4. Then we have computed the mean of the estimated variance (solid line in figure 1) together with 5% and 95% percentiles. These are obtained frequency by frequency (not to be understood as 5% and 95% percentiles over the whole frequency axis) using the function `prctile` of the Matlab Statistics Toolbox.

As one would expect, the variation of the estimated variance grows large as the variance itself grows.

6. CONCLUSIONS

In this paper we have addressed the problem of asymptotic variance computation of subspace estimates from data. We have shown that the formulas derived in (Chiuso, 2004) can be effectively used to assess the quality of estimated models; the implementation just requires computing some sample covariance matrices and relies on quantities which come as a byproduct of the identification experiment. Matlab code is available upon request from the first author.

A limitation of the theory is the fact that the variance expressions rely on the (rather strong) assumption that the true model belongs to the model class, i.e. no undermodeling occurs or, in other words, that the "true" order is chosen. In this sense, even though order selection is one of the steps in subspace identification, the asymptotic results are valid "conditionally" given the true order. Further research is necessary to overcome this limitation; this is in fact of paramount importance in practice, where identification has to be understood as model approximation/reduction rather than estimation of a "true" system.

REFERENCES

Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica* **41**, 359–376.

Bauer, D. and L. Ljung (2002). Some facts about the choice of the weighting matrices in Larimore type of subspace algorithm. *Automatica* **38**, 763–773.

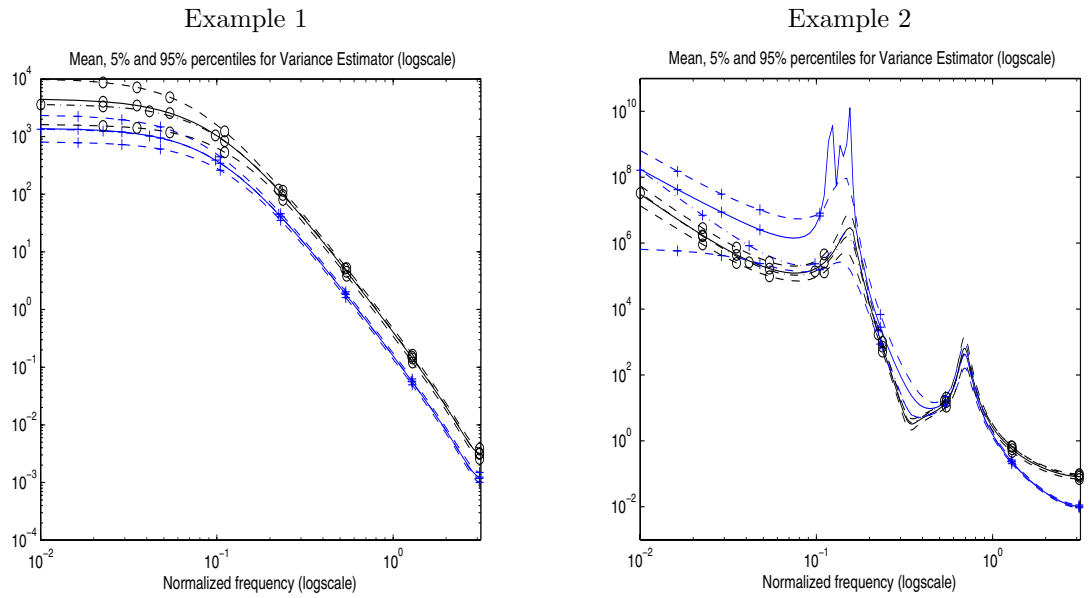

Fig. 1. Asymptotic variance estimated using data (1000 Monte Carlo runs) vs. true asymptotic variance. Solid: mean estimated variance. Dashed: 5% and 95% (frequency by frequency) percentiles computed from the Monte Carlo experiment. Dashed-dotted: true asymptotic variance. Black with crosses (+): predictor based subspace identification. Blue with circles (o): innovation estimation algorithm.

Caines, P.E. and C.W. Chan (1976). Estimation, identification and feedback. In: *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, Eds.). Academic. pp. 349–405.

Chiuso, A. (2004). Asymptotic variance of closed-loop subspace identification algorithms. *Submitted to IEEE Trans. on Aut. Control*. available at http://www.dei.unipd.it/∼chiuso.

Chiuso, A. and G. Picci (2003). Constructing the state of random processes with feedback. In: *Proc. of the IFAC Int. Symposium on System Identification (SYSID)*. Rotterdam.

Chiuso, A. and G. Picci (2004). The asymptotic variance of subspace estimates. *Journal of Econometrics* **118**(1-2), pp. 257–291.

Chiuso, A. and G. Picci (2005*a*). Consistency analysis of some closed-loop subspace identification methods. *Automatica* **41**(3), 377–391.

Chiuso, A. and G. Picci (2005*b*). Prediction error vs. subspace methods in closed-loop identification. In: *Proc. of the 16th IFAC World Congress*. Prague.

Gevers, M. and B.D.O. Anderson (1982). On jointly stationary feedback-free stochastic processes. *IEEE Trans. Automat. Contr.* **27**, 431–436.

Granger, C.W.J. (1963). Economic processes involving feedback. *Information and Control* **6**, 28–48.

Jansson, M. (2000). Asymptotic variance analysis of subspace identification methods. In: *Proceedings of SYSID2000*. S. Barbara Ca.

Jansson, M. (2003). Subspace identification and ARX modeling. In: *Proceedings of SYSID 2003*. Rotterdam.

Jansson, M. (2005). A new subspace identification method for open and closed loop data. In: *Proc. of the 16th IFAC World Congress*. Prague.

Larimore, W.E. (2004). Large sample efficiency for adaptix subspace identification with unknown feedback. In: *Proc. of IFAC DYCOPS'04*. Boston, MA, USA.

Lin, W., S.J. Qin and L. Ljung (2004). A framework for closed-loop subspace identification with
innovation estimation. Technical Report 07. Texas-Wisconsin Modeling and Control Consortium. Available at www.che.wisc.edu/jbr-group/tech-reports/twmcc-2004-07.pdf.

Lindquist, A. and G. Picci (1996). Canonical correlation analysis, approximate covariance extension and identification of stationary time series. *Automatica* **32**, 709–733.

Ljung, L. (1997). *System Identification; Theory for the User*. Prentice Hall.

Qin, S.J. and L. Ljung (2003). Closed-loop subspace identification with innovation estimation. In: *Proceedings of SYSID 2003*. Rotterdam.