

# Asymptotic variance of closed-loop subspace identification methods.

Alessandro Chiuso

## Abstract

In this paper the asymptotic properties of a version of the “innovation estimation” algorithm by Qin and Ljung as well as of a version of the “whitening filter” based algorithm introduced by Jansson are studied. Expressions for the asymptotic error as the sum of a “bias” term plus a “variance” term are given. The analysis is performed under rather mild assumptions on the spectrum of the joint input-output process; however, in order to avoid unnecessary complications, the asymptotic variance formulas are computed explicitly only for finite memory systems, i.e. of the ARX type. This assumption could be removed at the price of some technical complications; the simulation results confirm that when the past horizon is large enough (as compared to the predictor dynamics) the asymptotic expressions provide a good approximation of the asymptotic variance also for ARMAX systems.

## Index Terms

Subspace Identification, Closed-Loop Identification, Asymptotic Variance

## I. INTRODUCTION

Assessing the quality of estimated models is undoubtedly a central issue in system identification, see [36], [41]. Statistical properties are also fundamental tools to study the relative performance of different approaches and constitute a starting point for optimizing choices of various user parameters. While for classical prediction error methods (PEM hereafter) there are well established results concerning asymptotic statistical properties of the estimators both for open (see [36], [41]) and closed loop (see [20],

This work has been supported in part by the *RECSYS* project of the European Community and by the national project *Identification and adaptive control of industrial systems* funded by MIUR.

Alessandro Chiuso is with the Department of Information Engineering, University of Padova, Via Gradenigo 6/A, 35131 Padova, Italy, ph. +39-049-8277709, fax. +39-049-8277699, e-mail: [chiuso@dei.unipd.it](mailto:chiuso@dei.unipd.it)

[21]) operating conditions, the same does not hold for the so called “subspace methods”. This class of methods has been introduced in the last decades by several authors; see for instance [38], [33], [49], [50], [48], [43], [44], [51], [39] and references therein.

The asymptotic properties of subspace identification methods operating in open loop have been recently studied by several researchers (see [52], [1], [31], [6], [7], [29], [4], [13], [15], [14] and references therein) and the field may be considered, even though not yet mature, quite developed. Some of these results allowed to perform comparison between different subspace methods as in [6], [29], [8], [26], [14] and also in some cases to show the link with prediction error methods (see [8], [2]).

Unfortunately the analysis for the open loop case cannot be directly applied to data involving feedback. Several complications arise, first of which the fact that most subspace algorithms do not deliver consistent estimators when applied to closed loop data. In fact the presence of stochastic feedback (see [25] for a formal definition) has always been recognized to be an obstacle for subspace methods to work properly and various attempts have been made in the last decades to extend existing subspace algorithms to deal with data involving feedback (see for instance [42], [47], [37], [18], [46]). We stress that in many practical cases feedback is present also if one cannot directly recognize physical controllers which “close the loop”; therefore it is of great interest to have algorithms which work without any specific assumption on the feedback structure.

For these reasons much work remains to be done and closed loop subspace algorithms, together with their asymptotic properties, have been indicated as an “important and open research question” also by D. Bauer in the Semiplenary Lecture given at the recent International Symposium on System Identification (see [3]).

Very recently two subspace methods (see [40], [30]) have been proposed which may be regarded, in our opinion, as a significant step towards a satisfactory subspace identification algorithm working with closed-loop data. These algorithms may be seen as implementation of some stochastic realization procedures introduced in the recent works [12], [11]. Using these tools we have analyzed in [17] the consistency of these algorithms. Independently a similar analysis has been developed in [34] for the algorithm first introduced in [40].

In this paper we study the asymptotic behavior of a version of the estimators obtained by the algorithm introduced by Qin and Ljung in [40] (referred to as “innovation estimation” later on) and by the version of the algorithm by Jansson as analyzed in [17] (referred to as “whitening filter” later on)<sup>1</sup>. In particular we

<sup>1</sup>For a motivation of this terminology we refer to the original works [40], [30] or to the paper [17].

derive asymptotic variance formulas under a simplifying assumption which guarantees that the estimators are asymptotically unbiased<sup>2</sup> and provide tools to deal with the general case. We consider a version of Qin-Ljung which uses the so-called “state approach” [5], partly motivated by the results obtained in [10] concerning optimality of this approach in certain particular cases.

The original algorithm by Jansson [30] requires that a window of Markov parameters be preliminarily estimated. This two steps procedure seems to require a separate analysis and will be addressed in future work; in this paper we stick to procedures which *do not* require any *a priori* knowledge or preliminary estimation step.

The structure of the paper is as follows: In Section II we state the problem, recall some preliminary results and set up notations.

In Section III we provide a brief summary of the algorithms we study.

After this introductory part we analyze the algorithm by Qin and Ljung as a “prototypical” example of subspace algorithm to illustrate the main ideas which allows to derive the asymptotic variance formulas. The very same ideas would apply to virtually any subspace algorithm.

Section IV contains a description of the main ideas which allow to derive the asymptotic distribution of the errors.

In Section V we show how errors originate when working with data sequences of finite length ( $N$  finite).

In Section VI we obtain recursions involving the estimated state with finite sequences.

In Section VII we introduce approximations of the error terms which are used in Section VIII to compute the asymptotic distribution.

Section IX contains some experimental results comparing the asymptotic variance computed using the formulas of this paper with the sample variance obtained from Monte Carlo simulation. As a demonstrating example we shall consider the identification of a first order ARX and a first order ARMAX systems with a proportional controller.

Section X contains some conclusions and open questions for future research.

Most technicalities together with the analysis of the “whitening filter” approach are postponed to Appendices.

Unfortunately the paper turns out to be rather technical and requires some effort to be read. Nevertheless we believe the analysis reported is worth the effort both from the practical side, as it allows to assess

<sup>2</sup>Asymptotically in the number of data  $N$  while keeping fixed the past horizon  $t - t_0$ . See later on for details.

quality of estimated models, and from the theoretical one, as it constitutes a step towards understanding and hopefully optimizing these kind of identification procedures.

The paper has been structured in a way so that, hopefully, the complicated but seemingly unavoidable notation is gradually introduced bringing the reader step by step to the proof of the main results.

The reader not interested in the details of the derivation may just read Sections II and III, learn about the general ideas in Section IV, understand Propositions 5.3, 6.1 keeping track of the notation, understand the linearizations in Section VII summarized in Propositions 7.1 and study the notation of Section VIII which contains the main result stated in Theorem 8.1. A parallel derivation for the “whitening filter” may be found in Appendix B

## II. BASIC NOTATION AND PRELIMINARIES

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

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

where in general there may be feedback from  $\{\mathbf{y}(t)\}$  to  $\{\mathbf{u}(t)\}$  [25], [9], [23], [22]. 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 (1) is minimal. For identifiability reasons [23], [22] we assume that  $D = 0$ , i.e. there is no direct feedthrough from<sup>3</sup>  $\mathbf{u}$  to  $\mathbf{y}$ . Under this assumption the feedback (if any) can be quite general. For future reference we define  $\bar{A} := A - KC$ .

The white noise process  $\mathbf{e}$ , the innovation of  $\mathbf{y}$  given the joint past of  $\mathbf{y}, \mathbf{u}$ , is defined as the (linear) one step ahead prediction error of  $\mathbf{y}(t)$  given the joint (strict) past of  $\mathbf{u}$  and  $\mathbf{y}$  up to time  $t$ . We do not make any assumption on the correlation structure of  $\mathbf{u}$  and  $\mathbf{e}$ ; it is well known that  $\mathbf{e}(t)$  is uncorrelated with  $\mathbf{u}(s)$ ,  $-\infty < s < \infty$  if and only if there is no feedback from  $\mathbf{y}$  to  $\mathbf{u}$ .

Our aim is to identify the system parameters  $(A, B, C)$ , or equivalently the transfer function  $F(z) = C(zI - A)^{-1}B$ , starting from input-output data  $\{y_s, u_s\}$ ,  $s \in [t_0, T + N]$ , generated by the system (1) *without any a priori knowledge* of the (possible) feedback from  $\mathbf{y}$  to  $\mathbf{u}$ . This is in contrast with the so-called joint input-output identification, where a full model of the joint input-output process is sought. Joint

<sup>3</sup>We shall use interchangeably the notation  $\{\mathbf{u}(t)\}$  and the simpler symbol  $\mathbf{u}$  to denote the random process, i.e. the collection of random variables  $\mathbf{u}(t)$ ,  $t \in \mathbb{Z}$ .

input-output identification is well-established but needs an assumption of linearity (and time invariance) of the feedback channel and requires to work with more complex (joint) models. Also, obtaining the parameters  $(A, B, C)$  of the direct chain transfer function from the joint identified model requires in general to solve a nontrivial model reduction problem, see e.g. [32].

In most of this paper we shall deal with random fluctuations due to finite sample length (e.g. approximating expectations with finite time averages, etc.).

We shall use the standard notation of boldface (lowercase) letters to denote random variables (or semi-infinite tails). Lowercase letters denote sample values of a certain random variable. For example we shall denote with  $\mathbf{y}(t)$  the random vector denoting the output or equivalently the semi-infinite tail (a  $p \times \infty$  matrix)  $[y_t \ y_{t+1}, \dots \ y_{t+k} \ \dots]$  where  $y_t$  is the sample value of  $\mathbf{y}(t)$ . It can be shown (see [35], [16]) that the Hilbert spaces of random variables of second order stationary and ergodic process and the Hilbert space of semi-infinite tails containing sample values of the same process are isometrically isomorphic and therefore random variables and semi-infinite tails can be regarded as being the same object. For this reason we shall use the same symbol without risk of confusion.

We shall instead use boldface letters with a subscript  $N$  to denote the finite tails of length  $N$ . For instance  $\mathbf{y}_N(t) := [y_t \ y_{t+1}, \dots \ y_{t+N-1}]$ ,  $\mathbf{u}_N(t) := [u_t \ u_{t+1}, \dots \ u_{t+N-1}]$  and  $\mathbf{z}_N(t) := [\mathbf{y}_N^\top(t) \ \mathbf{u}_N^\top(t)]^\top$ . These are the block rows of the usual *data Hankel matrices* which appear in subspace identification.

Our  $N$  parameter is  $j$  in the notation of [44].

Still, in order to deal with realistic algorithms which can only regress on a finite amount of data, we shall keep *finite past and future horizons* (which play a similar role to the “ $i$ ” parameter of [44]) in subspace identification. This setting we shall describe as using data from a *finite observation interval* later on. In fact, in this paper finite (and generally fixed) past and future horizons will hold even when the sample size  $N$  is let going to  $\infty$  ( $j \rightarrow \infty$ ) for the purpose of asymptotic analysis. Because of this intrinsic limitation, the effect of initial conditions has to be taken into account and will play an important role as we shall see.

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

$$\begin{aligned} \mathcal{U}_{[t_0, t]} &:= \overline{\text{span}} \{ \mathbf{u}_k(s); k = 1, \dots, m, t_0 \leq s < t \} \\ \mathcal{Y}_{[t_0, t]} &:= \overline{\text{span}} \{ \mathbf{y}_k(s); k = 1, \dots, p, t_0 \leq s < t \} \end{aligned}$$

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.

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

We adopt the notation  $\Sigma_{\mathbf{ab}} := \mathbb{E}[\mathbf{ab}^\top]$  to denote the covariance matrix between the random vectors<sup>4</sup>  $\mathbf{a}$  and  $\mathbf{b}$ . In the finite dimensional case the orthogonal projection of the random variable  $\mathbf{a}$  onto the space spanned by the vector  $\mathbf{c}$ , i.e.  $\mathcal{C} := \text{span}\{\mathbf{c}\}$ , will be given by the usual formula<sup>5</sup>

$$E[\mathbf{a}|\mathcal{C}] = \Sigma_{\mathbf{ac}}\Sigma_{\mathbf{cc}}^{-1}\mathbf{c}. \quad (2)$$

When  $\mathbf{a}$  is a random vector,  $E[\mathbf{a} | \mathcal{C}]$  denotes the orthogonal projection of the components of  $\mathbf{a}$  onto  $\mathcal{C}$ .

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  $\Sigma_{\mathbf{ab}|\mathcal{C}}$  (or sometimes also  $\Sigma_{\mathbf{ab}|\mathcal{C}}$ ) will denote projection error covariance (conditional covariance in the Gaussian case)  $\Sigma_{\mathbf{ab}|\mathcal{C}} := \Sigma_{\tilde{\mathbf{a}}\tilde{\mathbf{b}}} = \Sigma_{\mathbf{ab}} - \Sigma_{\mathbf{ac}}\Sigma_{\mathbf{cc}}^{-1}\Sigma_{\mathbf{cb}}$ . Given two subspaces  $\mathcal{B} \subseteq \mathcal{U} \vee \mathcal{Y}$ ,  $\mathcal{C} \subseteq \mathcal{U} \vee \mathcal{Y}$ , with trivial intersection  $\mathcal{B} \cap \mathcal{C} = \{0\}$ ,  $E_{\parallel\mathcal{B}}[\cdot | \mathcal{C}]$  shall denote the oblique projection onto  $\mathcal{C}$  along  $\mathcal{B}$  (see [24], [16]). If  $\mathbf{b}$  and  $\mathbf{c}$  are basis for  $\mathcal{B}$  and  $\mathcal{C}$  respectively the oblique projection  $E_{\parallel\mathcal{B}}[\mathbf{a}|\mathcal{C}]$  can be computed using the formula:

$$E_{\parallel\mathcal{B}}[\mathbf{a}|\mathcal{C}] = \Sigma_{\mathbf{ac}|\mathbf{b}}\Sigma_{\mathbf{cc}|\mathbf{b}}^{-1}\mathbf{c}. \quad (3)$$

For column vectors formed by stacking past and/or future random variables (or semi-infinite Hankel matrices) we shall use the following notation:  $\mathbf{y}_{[t,s]} := \begin{bmatrix} \mathbf{y}^\top(t) & \mathbf{y}^\top(t+1) & \dots & \mathbf{y}^\top(s) \end{bmatrix}^\top$ . Similarly the (finite) Hankel data matrices will be denoted as  $(\mathbf{y}_{[t,s]})_N := \begin{bmatrix} \mathbf{y}_N^\top(t) & \mathbf{y}_N^\top(t+1) & \dots & \mathbf{y}_N^\top(s) \end{bmatrix}^\top$

For ease of notation we shall reserve the special symbols for the finite (future) Hankel data matrices:

$$\mathbf{y}_N^+ := (\mathbf{y}_{[t,T-1]})_N \quad \underline{\mathbf{y}}_N^+ := (\mathbf{y}_{[t+1,T]})_N. \quad (4)$$

The same notation shall be used for all signals involved (e.g.  $\mathbf{u}_N^+$ ,  $\mathbf{z}_N^+$  etc.). Spaces generated by finite tails, i.e. row spaces generated by finite Hankel data matrices, will be denoted with capital letters (in

<sup>4</sup>Zero mean.

<sup>5</sup>Provided  $\Sigma_{\mathbf{cc}}$  is invertible.

lieu of script capitals) with the appropriate subscripts. For instance  $U_{[t_0, t]}$  will be the space generated by the rows (length  $N$  vectors) of the matrix  $(\mathbf{u}_{[t_0, t]})_N$ . 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  $\mathbf{a}_N(t) := [a_t, a_{t+1}, \dots, a_{t+N-1}]$  and  $\mathbf{b}_N(t) := [b_t, b_{t+1}, \dots, b_{t+N-1}]$  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  $\lim_{N \rightarrow \infty} \hat{\Sigma}_{\mathbf{ab}} \stackrel{a.s.}{=} \Sigma_{\mathbf{ab}}$ .

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

$$\hat{E}[\mathbf{a}_N(t)|C_t] = \hat{\Sigma}_{\mathbf{ac}} \hat{\Sigma}_{\mathbf{cc}}^{-1} \mathbf{c}_N(t) \quad (5)$$

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

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

The symbol  $\hat{E}_{\parallel B_t}[\cdot|C_t]$  will denote the oblique projection along the space  $B_t$  onto the space  $C_t$  [24]. As above, the oblique projection can be computed using the formula:

$$\hat{E}_{\parallel B_t}[\mathbf{a}_N(t)|C_t] = \hat{\Sigma}_{\mathbf{ac}|\mathbf{b}} \hat{\Sigma}_{\mathbf{cc}|\mathbf{b}}^{-1} \mathbf{c}_N(t). \quad (6)$$

The difference between population and sample covariance matrices shall be denoted with a “tilde”, for instance

$$\tilde{\Sigma}_{\mathbf{ab}} := \hat{\Sigma}_{\mathbf{ab}} - \Sigma_{\mathbf{ab}} \quad \tilde{\Sigma}_{\mathbf{ab}|\mathbf{c}} := \hat{\Sigma}_{\mathbf{ab}|\mathbf{c}} - \Sigma_{\mathbf{ab}|\mathbf{c}}$$

All through this paper we shall assume that the joint process is “sufficiently rich”, in the sense that  $\mathcal{Z}_{[t_0, T]}$  admits the direct sum decomposition

$$\mathcal{Z}_{[t_0, T]} = \mathcal{Z}_{[t_0, t]} + \mathcal{Z}_{[t, T]}, \quad t_0 \leq t \leq T \quad (7)$$

the  $+$  sign denoting direct sum of subspaces. The symbol  $\oplus$  will be reserved for *orthogonal* direct sum. Various conditions ensuring sufficient richness are known. For example, it is well-known that for a full-rank purely non deterministic (p.n.d.) process  $\mathbf{z}$  to be sufficiently rich it is necessary and sufficient that the determinant of the spectral density matrix  $\Phi_z$  has no zeros on the unit circle [28].

This we formalize with the following assumption:

*Assumption 2.1:* The joint process  $\mathbf{z}(t)$  is a full rank (purely non deterministic) stochastic process whose spectral density matrix  $\Phi_{\mathbf{z}}$  is strictly positive on the unit circle

$$\Phi_{\mathbf{z}}(e^{j\omega}) > cI \quad c > 0, \omega \in [0, 2\pi]$$

Note that, for the joint spectrum to satisfy Assumption 2.1, reference signals such as sums of sinusoids are in general not allowed. This assumption can be relaxed since only full rank of the Hankel data matrix  $Z_{[t_0, T]}$  is required. This condition of course can be checked directly on the data.

We believe entering into the details of such aspects (as done for instance in the book [27]), is outside of the scope of this paper and therefore we prefer to work under Assumption 2.1.

With the notation introduced above the innovation  $\mathbf{e}(t)$ , i.e. the one step ahead prediction error of  $\mathbf{y}(t)$  based on the joint past  $\mathcal{Z}_t^-$  is written in the form

$$\mathbf{e}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) | \mathcal{Z}_t^-]. \quad (8)$$

We shall use the symbol  $\mathcal{E}_t$  to denote the space generated by the components of  $\mathbf{e}(t)$ .

We define the innovation process of  $\mathbf{u}$  based on the joint past

$$\mathbf{w}(t) := \mathbf{u}(t) - E[\mathbf{u}(t) | \mathcal{Z}_t^- \vee \mathcal{Y}_t] \quad t \in \mathbb{Z}$$

Let us define the transient innovations

$$\hat{\mathbf{e}}(t) := \mathbf{y}(t) - E[\mathbf{y}(t) | \mathcal{Z}_{[t_0, t]}]. \quad (9)$$

and

$$\hat{\mathbf{w}}(t) := \mathbf{u}(t) - E[\mathbf{u}(t) | \mathcal{Z}_{[t_0, t]} \vee \mathcal{Y}_t]. \quad (10)$$

We use the “hat” notation on random vectors  $\hat{\mathbf{e}}(t)$  (and similarly for  $\hat{\mathbf{x}}(t)$  etc.) to remind that this random vector are associated with the transient Kalman filter and are a function of the initial time instant  $t_0$ ; they should not be confused with the stationary innovation  $\mathbf{e}(t)$ . For  $t - t_0 \rightarrow \infty$   $\hat{\mathbf{e}}(t)$  tends to  $\mathbf{e}(t)$  (in mean square). Define also the transient Kalman state  $\hat{\mathbf{x}}(t) := E[\mathbf{x}(t) | \mathcal{Z}_{[t_0, t]}]$ . The spaces generated by the transient innovations are defined as

$$\hat{\mathcal{E}}_t := \text{span}\{\hat{\mathbf{e}}(t)\}, \quad \hat{\mathcal{W}}_t := \text{span}\{\hat{\mathbf{w}}(t)\}; \quad (11)$$

we also introduce

$$\hat{\mathbf{g}}(t) := \begin{bmatrix} \hat{\mathbf{e}}(t) \\ \mathbf{u}(t) \end{bmatrix} \quad \hat{\mathcal{G}}_{[t, s]} = \hat{\mathcal{E}}_{[t, s]} \vee \mathcal{U}_{[t, s]}. \quad (12)$$

Furthermore, recall that, given sequence of random variables  $\mathbf{w}(N)$ , we say that  $\sqrt{N}\mathbf{w}(N)$  is asymptotically normal if  $\sqrt{N}\mathbf{w}(N)$  converges in distribution to a normal random variable. The variance of this normal random variable is called *asymptotic variance* of  $\sqrt{N}\mathbf{w}(N)$ .

The symbol  $\stackrel{\cdot}{=}$  shall denote equality up to terms of order  $o(\frac{1}{\sqrt{N}})$  in probability, while  $O(1/\sqrt{N})$  will be used for terms which are of the order of  $1/\sqrt{N}$  in probability. We shall also denote with  $\text{vec}(A)$  the column vector obtained by stacking one on top of the other the columns of the matrix  $A$ . The superscript  $-L$  shall denote a matrix left inverse; particular choices of left inverses will be made later on.

### A. Some preliminary results

For future reference we report the following proposition from [17] without proof:

*Proposition 2.1:* Define

$$\tilde{\mathbf{x}}(t) := E[\mathbf{x}(t) | \mathcal{Z}_{[t_0, t+1]}] - \hat{\mathbf{x}}(t) \quad (13)$$

Then the following representation holds:

$$\hat{\mathbf{x}}(t+1) = A\hat{\mathbf{x}}(t) + B\mathbf{u}(t) + K\hat{\mathbf{e}}(t) + \bar{A}\tilde{\mathbf{x}}(t) \quad (14)$$

$$\mathbf{y}(t) = C\hat{\mathbf{x}}(t) + \hat{\mathbf{e}}(t) \quad (15)$$

*Lemma 2.2:* The projection error  $\bar{A}\tilde{\mathbf{x}}(t)$  can be expressed as

$$\bar{A}\tilde{\mathbf{x}}(t) = \bar{A}^{t-t_0+1} \left( E[\mathbf{x}(t_0) | \hat{\mathcal{E}}_t] + E[\mathbf{x}(t_0) | \hat{\mathcal{W}}_t] \right). \quad (16)$$

**Remark II.1** Equation (16) guarantees that the term  $\bar{A}\tilde{\mathbf{x}}(t)$  vanishes when  $\bar{A}$  is nilpotent, i.e.  $\bar{A}^{t-t_0} = 0$ . This happens when the “true” system is of the ARX type. Note that virtually non assumptions on the feedback mechanism have been made so far, besides that it should guarantee that the joint process is second order stationary.  $\diamond$

Under the simplifying assumption that feedback is generated through a finite dimensional linear system, i.e.

$$\begin{cases} \mathbf{s}(t+1) = F\mathbf{s}(t) + G\mathbf{y}(t) + L\mathbf{w}(t) \\ \mathbf{u}(t) = H\mathbf{s}(t) + J\mathbf{y}(t) + \mathbf{w}(t) \end{cases} \quad t \geq t_0 \quad (17)$$

the expressions in (16) can be made more explicit.

*Lemma 2.3:* Assume that feedback is generated according to (17) and define  $\bar{F} := F - LH$ . Let also  $\Lambda_e := \text{Var}\{\mathbf{e}(t)\}$ ,  $\Lambda_w := \text{Var}\{\mathbf{w}(t)\}$  and<sup>6</sup>

$$\Lambda_e(t-t_0) := \text{Var}\{\hat{\mathbf{e}}(t)\} = \Lambda_e + C\bar{A}^{t-t_0}\Sigma_{\mathbf{x}(t_0)\mathbf{x}(t_0)|\mathcal{Z}_{[t_0, t]}} \left( \bar{A}^\top \right)^{t-t_0} C^\top$$

<sup>6</sup>In this lemma we keep the time indexes also as subscripts of covariance matrices to avoid confusion.

$$\Lambda_w(t - t_0) := \text{Var} \{ \hat{\mathbf{w}}(t) \} = \Lambda_w + H \bar{F}^{t-t_0} \Sigma_{\mathbf{s}(t_0) \mathbf{s}(t_0) | \mathcal{Z}_{[t_0, t]} \vee \mathcal{Y}_t} (\bar{F}^\top)^{t-t_0} H^\top;$$

then the two projections in (16) can be written as:

$$E \left[ \mathbf{x}(t_0) \mid \hat{\mathcal{E}}_t \right] = \Sigma_{\mathbf{x}(t_0) \mathbf{x}(t_0) | \mathcal{Z}_{[t_0, t]}} (\bar{A}^\top)^{t-t_0} C^\top \Lambda_e^{-1}(t - t_0) \hat{\mathbf{e}}(t) \quad (18)$$

and

$$E \left[ \mathbf{x}(t_0) \mid \hat{\mathcal{W}}_t \right] = \Sigma_{\mathbf{x}(t_0) \mathbf{s}(t_0) | \mathcal{Z}_{[t_0, t]} \vee \mathcal{Y}_t} (\bar{F}^\top)^{t-t_0} H^\top \Lambda_w^{-1}(t - t_0) \hat{\mathbf{w}}(t) \quad (19)$$

**Remark II.2** Note that, under the finite dimensionality assumption of the feedback channel, the projection  $E \left[ \mathbf{x}(t_0) \mid \hat{\mathcal{W}}_t \right]$  vanishes when  $\bar{F}$  is nilpotent, i.e. the subsystem (17) generating the feedback signal is of the ARX type. This fact will be useful in the analysis of the “innovation estimation” algorithm. From now on we *shall not* make the assumption that feedback is generated according to (17), unless explicit reference to the matrix  $\bar{F}$  is made. The reader may just think that, if the feedback is not generated according to (17), any condition required on  $\bar{F}$  is *not satisfied*.  $\diamond$

In this paper we shall also make use of the “inverse” model of (14). The following lemma has been proved [17] (see eq. (4.20) and (4.21)).

*Lemma 2.4:* Let  $K(t)$  be a suitably defined time-varying gain.<sup>7</sup> Then the following holds:

$$\hat{\mathbf{x}}(t+1) = (A - K(t)C) \hat{\mathbf{x}}(t) + B \mathbf{u}(t) + K(t) \hat{\mathbf{e}}(t) - \tilde{B}(t) \hat{\mathbf{w}}(t) \quad (20)$$

$$\hat{\mathbf{e}}(t) = -C \hat{\mathbf{x}}(t) + \mathbf{y}(t) \quad (21)$$

where  $\tilde{B}(t) \rightarrow 0$  when at least one of  $\bar{A}^{t-t_0} \rightarrow 0$ ,  $\bar{F}^{t-t_0} \rightarrow 0$  holds and  $K(t) \rightarrow K$  when  $\bar{A}^{t-t_0} \rightarrow 0$ . We define, for convenience of notation  $\bar{A}(t) := A - K(t)C$ ,  $\Phi(t, s) := \prod_{i=0}^{s-1-t} \bar{A}(s-i-1)$  where  $\Phi(t, t) := I$  and

$$\bar{\Gamma}(t, s) := \left[ C^\top \quad \Phi^\top(t, t+1) C^\top \quad \dots \quad \Phi^\top(t, s) C^\top \right]^\top.$$

### III. SUMMARY OF THE ALGORITHMS

In this section we summarize the algorithms that will be analyzed in the rest of the paper. We analyze a version of the algorithms which uses the so called “state” approach [4], which refers to the fact that the state matrices are estimated solving a least squares problem starting from the estimated state sequences (see eq. (27) and (33)). To avoid confusion we shall reserve the symbol  $\hat{\xi}_N$  for the state estimate provided by the algorithm by Qin and Ljung and  $\hat{\zeta}_N$  for those obtained by the algorithm based on the whitening

<sup>7</sup>Since in this paper we are not interested in modeling the stochastic part, it suffices to note that  $K(t) \rightarrow K$  as  $\bar{A}^{t-t_0}$  goes to zero (see [17] for details).

filter. As we shall see these state estimators correspond to different choice of basis. Therefore we shall denote with  $A_N^{ql}, B_N^{ql}, C_N^{ql}, K_N^{ql}$  the state matrices in the basis

$$\hat{T}_N := \hat{\Gamma}_N^{-L} \Gamma \quad (22)$$

which follows from (26), i.e.  $A_N^{ql} := \hat{T}_N A \hat{T}_N^{-1}$  etc. Similarly  $A_N^{wf}, B_N^{wf}, C_N^{wf}, K_N^{wf}$  shall be the state matrices expressed in the basis

$$\hat{\hat{T}}_N := \hat{\hat{\Gamma}}_N^{-L} \bar{\Gamma} \quad (23)$$

which follows from (32), i.e.  $A_N^{wf} := \hat{\hat{T}}_N A \hat{\hat{T}}_N^{-1}$  etc. The superscripts  $ql$  and  $wf$  stand for “Qin-Ljung” and “whitening filter” respectively while the subscript  $N$  reminds that the basis in (26) and (32) depend on the data length  $N$ . This observation will be crucial in the rest of the paper. In fact we shall compare the estimated system matrices with the “true” ones in a certain “data dependent” basis corresponding to (22) and (23).

#### A. The “innovation estimation” by Qin and Ljung

The basic steps of the algorithm in [40] are as follows:

- 1) Construct the future innovations  $\hat{\mathbf{e}}_N(s) = \mathbf{y}_N(s) - \hat{E}[\mathbf{y}_N(s) | Z_{[t_0, s]}]$ ,  $s = t, \dots, T$ . Define  $\hat{\mathbf{g}}_N(t) := [\hat{\mathbf{e}}_N^\top(t) \mathbf{u}_N^\top(t)]^\top$  and denote with  $\hat{G}_{[t_1, t_2]}$  the space generated by the rows of the matrix  $(\hat{\mathbf{g}}_{[t_1, t_2]})_N$ .
- 2) Estimate the (extended) observability matrix from factorization (weighted SVD) of

$$W^{-1} \hat{E}_{\|\hat{G}_{[t, T-1]}}^{\dagger} [\mathbf{y}_N^+ | Z_{[t_0, t]}] = USV \simeq U_n S_n V_n^\top \quad (24)$$

where  $W$  is a weighting matrix which is usually introduced in subspace identification [45]. The matrices  $U_n, V_n$ , denote respectively the first  $n$  columns of the matrices  $U$  and  $V$ , where  $n$  is the number of “significant” singular values retained in  $S_n$  which is the  $n$ -dimensional upper left block of  $S$ . The same notation will be used throughout.

An estimate of the observability matrix is obtained discarding the “less significant” singular values (i.e. pretending  $\tilde{S}_n \simeq 0$ ) from

$$\hat{\Gamma}_N = W U_n \hat{T} \quad (25)$$

where  $\hat{T}$  can be any non-singular matrix providing a choice of basis. It will be useful for the analysis reported in this paper to make specific choices of  $\hat{T}$  which shall guarantee that the estimated system matrices converge regardless of the possible ambiguities<sup>8</sup> in the SVD (24).

<sup>8</sup>Due to orientation of singular vectors and multiple singular values (if any).

Of course when implementing the algorithms and also when computing the variance formulas, the specific  $\hat{T}$  will not matter. Typical choices of  $\hat{T}$  are  $\hat{T} = I$  or  $\hat{T} = S_n^{1/2}$

- 3) Introducing the (W-weighted) left inverse  $\hat{\Gamma}_N^{-L} := \left( \hat{\Gamma}_N^\top W^{-\top} W^{-1} \hat{\Gamma}_N \right)^{-1} \hat{\Gamma}_N^\top W^{-\top} W^{-1} = \hat{T}^{-1} U_n^\top W^{-1}$  construct basis for the state space as follows :

$$\hat{\xi}_N(t) := \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\mathbf{y}_N^+ | Z_{[t_0, t]}] \quad \hat{\xi}_N(t+1) := \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t+1, T]}} [\mathbf{y}_N^+ | Z_{[t_0, t+1]}] \quad (26)$$

- 4) Solve in the least squares sense

$$\begin{cases} \hat{\xi}_N(t+1) \simeq A_N^{ql} \hat{\xi}_N(t) + B_N^{ql} \mathbf{u}_N(t) + K_N^{ql} \hat{\mathbf{e}}_N(t) \\ \mathbf{y}_N(t) \simeq C_N^{ql} \hat{\xi}_N(t) + \hat{\mathbf{e}}_N(t) \end{cases} \quad (27)$$

to obtain  $\hat{A}_N^{ql}$ ,  $\hat{B}_N^{ql}$ ,  $\hat{C}_N^{ql}$ , the estimators of  $A_N^{ql}$ ,  $B_N^{ql}$ ,  $C_N^{ql}$  using  $N$  data points.

As shown in [17]  $\hat{\xi}_N(t)$  is an (asymptotically) biased estimator of  $\hat{T}_N \hat{\mathbf{x}}_N(t) := \hat{E} \left[ \hat{T}_N \mathbf{x}_N(t) | Z_{[t_0, t]} \right]$ .

As a consequence also the equations (27) should be regarded as only ‘‘approximately’’ satisfied.

As mentioned above, for the purpose of analysis, it is useful to make a specific choice of  $\hat{T}$  so that the data dependent basis  $\hat{T}_N = \hat{\Gamma}_N^{-L} \Gamma = \hat{T}^{-1} U_n^\top W^{-1} \Gamma$  converges almost surely to a non-singular matrix  $T_\infty$ , which without loss of generality can be chosen equal to the identity matrix.

For instance a straightforward (but not realizable) choice is of course

$$\hat{T} = U_n^\top W^{-1} \Gamma \quad (28)$$

which guarantees that  $\hat{T}_N = I$ .

There are many other possible choices of  $\hat{T}$  which serve to this purpose, among which the one proposed in [8]; it also possible to choose  $\hat{T}$  without knowledge of the true system parameters ( $\Gamma$  in (28)). We stress however that this choice is inessential for the implementation of both algorithms and variance formulas and is just needed here for the purpose of analysis. Given this for granted, in order to simplify exposition we prefer not to stick to a specific choice of  $\hat{T}$  and rather make the following assumption:

*Assumption 3.1:* For the purpose of analysis we assume that  $\hat{T}$  in (25) has been chosen so that  $\hat{T}_N$  in (22) converges to a non singular matrix  $T_\infty$  as  $N$  goes to infinity.

### B. ‘‘Whitening filter’’-based algorithm

We analyze a version of the algorithm introduced in [17]. A conceptually similar algorithm, but with an implementation differing in the preliminary estimation of certain Markov parameters which are used in place of the oblique projection, has been introduced in [30]. The basic steps are as follows:

- 1) Construct the oblique predictors  $\hat{\mathbf{y}}_N(t+k|t-1) = \hat{E}_{\|Z_{[t,t+k]}}[\mathbf{y}_N(t+k) | Z_{[t_0,t]}]$ ,  $k = 0, \dots, T-t-1$  and  $\hat{\mathbf{y}}_N(t+k|t) = \hat{E}_{\|Z_{[t+1,t+k]}}[\mathbf{y}_N(t+k) | Z_{[t_0,t+1]}]$ ,  $k = 1, \dots, T-t$ .
- 2) Estimate the predictor (extended) observability matrix  $\bar{\Gamma}(t, T-1)$  from factorization (weighted SVD) of

$$\bar{W}^{-1} \begin{bmatrix} \hat{\mathbf{y}}_N(t|t-1) \\ \hat{\mathbf{y}}_N(t+1|t-1) \\ \vdots \\ \hat{\mathbf{y}}_N(T-1|t-1) \end{bmatrix} = \bar{U}\bar{S}\bar{V} \simeq \bar{U}_n\bar{S}_n\bar{V}_n^\top \quad (29)$$

letting

$$\hat{\Gamma}_N = \bar{W}\bar{U}_n\hat{T} \quad (30)$$

Also for the “whitening-filter” algorithm for instance the choice

$$\hat{T} := \bar{U}_n^\top \bar{W}^{-1} \bar{\Gamma} \quad (31)$$

where  $\bar{\Gamma}$  is the “true” (but unknown) predictor observability matrix, guarantees that  $\hat{T}_N = I$ . Also in this case we shall not stick to a specific  $\hat{T}$  but rather work under an assumption similar to Assumption 3.1.

- 3) Defining the left inverse  $\hat{\Gamma}_N^{-L} := \left( \hat{\Gamma}_N^\top \bar{W}^{-\top} \bar{W}^{-1} \hat{\Gamma}_N \right)^{-1} \hat{\Gamma}_N^\top \bar{W}^{-\top} \bar{W}^{-1}$ , construct basis for the state space as follows:

$$\hat{\zeta}_N(t) := \hat{\Gamma}_N^{-L} \begin{bmatrix} \hat{\mathbf{y}}_N(t|t-1) \\ \hat{\mathbf{y}}_N(t+1|t-1) \\ \vdots \\ \hat{\mathbf{y}}_N(T-1|t-1) \end{bmatrix} \quad \hat{\zeta}_N(t+1) := \hat{\Gamma}_N^{-L} \begin{bmatrix} \hat{\mathbf{y}}_N(t+1|t) \\ \hat{\mathbf{y}}_N(t+2|t) \\ \vdots \\ \hat{\mathbf{y}}_N(T|t) \end{bmatrix} \quad (32)$$

- 4) Solve in the least squares sense

$$\begin{cases} \hat{\zeta}_N(t+1) \simeq A_N^{wf} \hat{\zeta}_N(t) + B_N^{wf} \mathbf{u}(t) + K_N^{wf} \hat{\mathbf{e}}_N(t) \\ \mathbf{y}_N(t) \simeq C_N^{wf} \hat{\zeta}_N(t) + \hat{\mathbf{e}}_N(t) \end{cases} \quad (33)$$

to obtain  $\hat{A}_N^{wf}$ ,  $\hat{B}_N^{wf}$ ,  $\hat{C}_N^{wf}$ , the estimators of  $A_N^{wf}$ ,  $B_N^{wf}$ ,  $C_N^{wf}$  using  $N$  data points.

As discussed in [17]  $\hat{\zeta}_N(t)$  is a biased estimator of  $\hat{T}_N \hat{\mathbf{x}}_N(t)$ . As a consequence also the equations (33) should be regarded as only “approximately” satisfied. As we shall see there is one more difficulty in this case related to the fact that  $\bar{\Gamma}(t, T-1)$  is not equal to  $\bar{\Gamma}(t+1, T)$  for finite  $t - t_0$  unless  $\bar{A}$  is nilpotent.

#### IV. OUTLINE OF THE DERIVATION

This section is meant to provide a very “high level” description of the main steps which allow to derive the asymptotic distribution of the errors as well as a critical discussion of the “changing basis” approach.

As should be clear from the description in the previous section, both algorithms<sup>9</sup> are based on the two steps:

- 1) construct matrices which form basis for the state space at time  $t$  and  $t + 1$  as done in (26) and (32)
- 2) solve the state equation in the least squares sense to obtain the system matrices as done in (27) and (33).

Our strategy to obtain the asymptotic formulas follows the two steps above. For simplicity of presentation from now on we refer to the “innovation estimation” algorithm; the discussion and parallel results for the “whitening filter” approach are postponed to Appendix B. Recall that for the purpose of analysis we shall always work under Assumption 3.1 which guarantees that  $\hat{T}_N$  converges almost surely to a non-singular  $T_\infty$ .

- 1) First we derive exact expressions for the estimated state sequences of the form:

$$\begin{aligned}\hat{\xi}_N(t) &:= \hat{T}_N \hat{\mathbf{x}}_N(t) + \Delta \mathbf{x}_N(t) \\ \hat{\xi}_N(t+1) &:= \hat{T}_N \hat{\mathbf{x}}_N(t+1) + \Delta \mathbf{x}_N(t+1)\end{aligned}\tag{34}$$

which express the estimated state sequences  $\hat{\xi}_N(t)$  and  $\hat{\xi}_N(t+1)$  as a non-singular change of basis of the “true” Kalman state (finite tail) plus perturbation terms  $\Delta \mathbf{x}_N(t)$  and  $\Delta \mathbf{x}_N(t+1)$ ; these terms shall include both a “bias” as well as a “variance” part and will be analyzed at the beginning of Section VI.

- 2) Then, in Proposition 6.1, we derive “state-like” recursions for  $\hat{\xi}_N(t)$  and  $\hat{\xi}_N(t+1)$  which involve the true state matrices in the data dependent basis  $\hat{T}_N$  of the form

$$\left\{ \begin{array}{l} \hat{\xi}_N(t+1) = A_N^{ql} \hat{\xi}_N(t) + B_N^{ql} \mathbf{u}_N(t) + K_N^{ql}(t) \hat{\mathbf{e}}_N(t) + \\ \quad + \Delta \mathbf{x}_N \\ \mathbf{y}_N(t) = C_N^{ql} \hat{\xi}_N(t) + \hat{\mathbf{e}}_N(t) + \\ \quad + \Delta \mathbf{y}_N \end{array} \right.\tag{35}$$

where  $\Delta \mathbf{x}_N$  and  $\Delta \mathbf{y}_N$  are error terms.

<sup>9</sup>Virtually all subspace algorithms based on the so called “state approach” [5] follows the same lines.

3) From equations (35) it will be easy to obtain expressions for the errors:

$$\sqrt{N}\tilde{A}_N^{ql} := \sqrt{N}(\hat{A}_N^{ql} - A_N^{ql}) \quad \sqrt{N}\tilde{B}_N^{ql} := \sqrt{N}(\hat{B}_N^{ql} - B_N^{ql}) \quad \sqrt{N}\tilde{C}_N^{ql} := \sqrt{N}(\hat{C}_N^{ql} - C_N^{ql}) \quad (36)$$

which express the estimation error in the data depend basis  $\hat{T}_N$ . This will be done in Proposition 6.2

In order to derive the asymptotic distribution of (36) we shall need to derive asymptotic equivalent versions of these error terms via linearization, which is a standard tool in asymptotic analysis [19]. See sections VII and VIII.

4) The asymptotic distribution of any system invariant  $g(\theta)$  which depends differentiably on the system parameters  $\theta = (\text{vec}^\top(A), \text{vec}^\top(B), \text{vec}^\top(C))^\top$  can be found by using well-known linearization techniques [19].

Some remarks are now in order, especially concerning point 3). The errors are expressed in a data dependent basis, which means that the estimators (say  $\hat{A}_N^{ql}$ ) are confronted with the ‘‘moving targets’’ (say  $A_N^{ql}$ ).

Under Assumption 3.1  $\hat{T}_N$  converges and therefore also the moving targets  $A_N^{ql}$ ,  $B_N^{ql}$ ,  $C_N^{ql}$  converge; this will be needed in the proof.

When it comes to computing asymptotic distribution of (say) transfer function estimators

$$\hat{F}_N^{ql}(z) = \hat{C}_N^{ql} (zI - \hat{A}_N^{ql})^{-1} \hat{B}_N^{ql}$$

of  $F(z) = C_N^{ql} (zI - A_N^{ql})^{-1} B_N^{ql}$  the simple linearization (see for instance [29], [13]):

$$\begin{aligned} \hat{F}_N^{ql}(z) - F(z) &\doteq (\hat{C}_N^{ql} - C_N^{ql})(zI - A_N^{ql})^{-1} B_N^{ql} + C_N^{ql}(zI - A_N^{ql})^{-1}(\hat{B}_N^{ql} - B_N^{ql}) + \\ &\quad + C_N^{ql}(zI - A_N^{ql})^{-1}(\hat{A}_N^{ql} - A_N^{ql})(zI - A_N^{ql})^{-1} B_N^{ql} \doteq \\ &\doteq (\hat{C}_N^{ql} - C_N^{ql})(zI - \hat{A}_N^{ql})^{-1} \hat{B}_N^{ql} + \hat{C}_N^{ql}(zI - \hat{A}_N^{ql})^{-1}(\hat{B}_N^{ql} - B_N^{ql}) + \\ &\quad + \hat{C}_N^{ql}(zI - \hat{A}_N^{ql})^{-1}(\hat{A}_N^{ql} - A_N^{ql})(zI - \hat{A}_N^{ql})^{-1} \hat{B}_N^{ql} \end{aligned} \quad (37)$$

shows that indeed, as system invariants are concerned, errors around a ‘‘moving target’’ do the job. Note in fact that the linearization above makes sense as long as  $(\hat{A}_N^{ql} - A_N^{ql}) = O(1/\sqrt{N})$ ,  $(\hat{B}_N^{ql} - B_N^{ql}) = O(1/\sqrt{N})$  and  $(\hat{C}_N^{ql} - C_N^{ql}) = O(1/\sqrt{N})$  in probability, which indeed holds true.

Furthermore, as already mentioned, the specific basis  $\hat{T}$  does not matter when it comes to compute invariants. When computing in practice the asymptotic variance from data, one will just use the estimated quantities, which will be obtained from realizable choice of  $\hat{T}$  (usually  $\hat{T} = I$  or  $\hat{T} = \hat{S}_n^{1/2}$ ). This can be easily seen rewriting equation (37) using different choice of basis  $\hat{T}$ .

## V. FINITE DATA MODELS

In order to compute the error terms  $\Delta \mathbf{x}_N$  and  $\Delta \mathbf{y}_N$  in (35) we need to consider the finite data counterparts of state update equations found in Section II and taken from [17].

We recall from the previous section that the transient innovation  $\hat{\mathbf{e}}_N(t)$  is given by<sup>10</sup>

$$\hat{\mathbf{e}}_N(t) := \mathbf{y}_N(t) - \hat{E}[\mathbf{y}_N(t) \mid Z_{[t_0,t]}] \quad (38)$$

and the transient Kalman state  $\hat{\mathbf{x}}_N(t) := \hat{E}[\mathbf{x}_N(t) \mid Z_{[t_0,t]}]$ . Similarly we define

$$\hat{\mathbf{w}}_N(t) := \mathbf{u}_N(t) - \hat{E}[\mathbf{u}_N(t) \mid Z_{[t_0,t]} \vee Y_t] \quad (39)$$

The spaces generated by the transient innovations shall be denoted

$$\hat{E}_t := \text{span}\{\hat{\mathbf{e}}_N(t)\}, \quad \hat{W}_t := \text{span}\{\hat{\mathbf{w}}_N(t)\}. \quad (40)$$

Define also

$$\hat{\mathbf{e}}_N(t) := \hat{E}[\mathbf{e}_N(t) \mid Z_{[t_0,t]}]; \quad (41)$$

note that  $\hat{\mathbf{e}}_N(t)$  goes to zero as  $N$  goes to infinity.

**Remark V.3** This projection residual is the source of errors due to “finite tails”, i.e. approximation of expected values with finite sums.  $\diamond$

The following proposition is the analogous of Proposition 2.1 and is the fundamental tool to study asymptotic properties.

*Proposition 5.1:* Define

$$\tilde{\mathbf{x}}_N(t) := \hat{E}[\mathbf{x}_N(t) \mid Z_{[t_0,t+1]}] - \hat{\mathbf{x}}_N(t) \quad (42)$$

Then the following representation holds:

$$\begin{aligned} \hat{\mathbf{x}}_N(t+1) &= A\hat{\mathbf{x}}_N(t) + B\mathbf{u}_N(t) + K\hat{\mathbf{e}}_N(t) + \\ &\quad + K\hat{\mathbf{e}}_N(t) + \bar{A}\tilde{\mathbf{x}}_N(t) \end{aligned} \quad (43)$$

$$\mathbf{y}_N(t) = C\hat{\mathbf{x}}_N(t) + \hat{\mathbf{e}}_N(t) + \hat{\mathbf{e}}_N(t) \quad (44)$$

*Proof:* The proof is basically the same as that of Proposition 2.1 and can easily be adapted from the proof found in [17].  $\blacksquare$

<sup>10</sup>Note that  $\hat{\mathbf{e}}_N(t)$  is not the truncation of  $\hat{\mathbf{e}}(t)$ ; as a mnemonic rule the reader may think that truncation of sequences is done before projection is taken. Therefore first  $\mathbf{y}(t)$  is truncated to  $\mathbf{y}_N(t)$  and then projected onto the finite past  $\hat{E}[\mathbf{y}_N(t) \mid Z_{[t_0,t]}]$ . Recall that the “hat” on  $\hat{\mathbf{e}}_N(t)$  reminds that this is a *transient* innovation (finite  $t_0$ ).

*Lemma 5.2:* The projection error  $\tilde{\mathbf{x}}_N(t)$  satisfies

$$\begin{aligned}\bar{A}\tilde{\mathbf{x}}_N(t) &= \bar{A}^{t-t_0+1} \left( \hat{E} \left[ \mathbf{x}_N(t_0) \mid \hat{E}_t \right] + \hat{E} \left[ \mathbf{x}_N(t_0) \mid \hat{W}_t \right] \right) \\ &= \hat{\Theta}_t \hat{\mathbf{e}}_N(t) + \hat{\Psi}_t \hat{\mathbf{w}}_N(t)\end{aligned}\quad (45)$$

Furthermore there exist constants  $\alpha$  and  $\beta$  so that<sup>11</sup>  $\|\hat{\Theta}_t\| \leq \alpha \|\bar{A}^{t-t_0}\|$  and  $\|\hat{\Psi}_t\| \leq \beta \|\bar{A}^{t-t_0}\|$ . If the feedback is generated according to (17) then also  $\|\hat{\Psi}_t\| = O\left(\|\bar{A}^{t-t_0}\| \cdot \left(\|\bar{F}^{t-t_0}\| + \frac{1}{\sqrt{N}}\right)\right)$  in probability holds. Of course as  $N \rightarrow \infty$  we obtain exactly the expressions of Lemmas 2.2 and 2.3.

*Proof:* Again the proof follows the same lines of Lemmas 2.2 and 2.3 and can be adapted from [17]. ■

**Remark V.4** Lemma 5.2 implies that, as in the ideal case of infinite data ( $N = \infty$ ) of Lemma 2.2, the bias term disappears if  $\bar{A}$  is nilpotent. To be precise, as in the infinite data setup (see [17]),  $\|\bar{A}^{t-t_0}\| \|\bar{F}^{t-t_0}\| = 0$  is a sufficient condition for  $\hat{A}_N^{ql}$ ,  $\hat{B}_N^{ql}$ ,  $\hat{C}_N^{ql}$  obtained by the Qin-Ljung algorithm [40] to be consistent estimators of  $A_N^{ql}$ ,  $B_N^{ql}$ ,  $C_N^{ql}$ . In other words the ‘‘innovation estimation’’ algorithm is consistent if at least one of the two conditions is verified:

- 1) the ‘‘true’’ system is of the ARX type
- 2) feedback is generated according to (17) and  $\bar{F}^{t-t_0} = 0$  i.e. it is of the ARX type.

Note that, if  $\bar{A}^{t-t_0} \neq 0$ , the bias due to  $\hat{\Psi}_t$  is  $O(1/\sqrt{N})$  in probability; therefore this term cannot be neglected when it comes to compute variance<sup>12</sup>. This would introduce some complications in the formulas and therefore, for the purpose of exposition, we shall always make the assumption that  $\bar{A}^{t-t_0} = 0$  also for the algorithm by Qin and Ljung. ◇

Using the results of Proposition 5.1 and Lemma 5.2 it is convenient to rewrite the state recursions directly in terms of  $\hat{\mathbf{w}}_N(t)$ ,  $\hat{\mathbf{e}}_N(t)$ , and  $\hat{\mathbf{e}}_N(t)$ .

*Proposition 5.3:* Let  $K(t) := K + \hat{\Theta}_t$ ; then the truncated Kalman state  $\hat{\mathbf{x}}_N(t)$  satisfies the recursion

$$\begin{aligned}\hat{\mathbf{x}}_N(t+1) &= A\hat{\mathbf{x}}_N(t) + B\mathbf{u}_N(t) + K(t)\hat{\mathbf{e}}_N(t) + \\ &\quad + K\hat{\mathbf{e}}_N(t) + \hat{\Psi}_t\hat{\mathbf{w}}_N(t)\end{aligned}\quad (46)$$

$$\mathbf{y}_N(t) = C\hat{\mathbf{x}}_N(t) + \hat{\mathbf{e}}_N(t) + \hat{\mathbf{e}}_N(t)\quad (47)$$

<sup>11</sup> $\|\cdot\|$  denotes any matrix norm; we use here the Frobenius norm.

<sup>12</sup>The author would like to thank an anonymous reviewer for pointing out this fact which had been overlooked in a preliminary version of the manuscript.

which can be written in “predictor form” as follows:

$$\begin{aligned}\hat{\mathbf{x}}_N(t+1) &= \bar{A}(t)\hat{\mathbf{x}}_N(t) + B\mathbf{u}_N(t) + K\mathbf{y}_N(t) + \\ &\quad -\hat{\Theta}_t\hat{\mathbf{e}}_N(t) + \hat{\Psi}_t\hat{\mathbf{w}}_N(t)\end{aligned}\quad (48)$$

$$\mathbf{y}_N(t) = C\hat{\mathbf{x}}_N(t) + \hat{\mathbf{e}}_N(t) + \hat{\mathbf{e}}_N(t) \quad (49)$$

where we recall that  $\bar{A}(t) := A - K(t)C$ .

*Proof:* The proof consist in a simple algebraic manipulation of (43), (44) and (45).  $\blacksquare$

Using the machinery developed so far, and in particular equations (46) and (47) we have basically proved the following technical lemma.

*Lemma 5.4:* Let  $\nu := T - t$  and define the matrices

$$\hat{H}_{\mathbf{w}}(t, T-1) := \begin{bmatrix} 0 & 0 & \dots & 0 \\ C\hat{\Psi}_t & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\nu-3}\hat{\Psi}_t & \dots & C\hat{\Psi}_{T-2} & 0 \end{bmatrix} \quad H_s := \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\nu-3}K & \dots & CK & I \end{bmatrix}$$

and

$$\hat{H}_{\mathbf{e}}(t, T-1) := \begin{bmatrix} I & 0 & \dots & 0 \\ CK(t) & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\nu-3}K(t) & \dots & CK(T-2) & I \end{bmatrix} \quad H_{\mathbf{u}} := \begin{bmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\nu-3}B & \dots & CB & 0 \end{bmatrix}.$$

Let us also define

$$\begin{aligned}\mathbf{b}_N(t) &:= \hat{H}_{\mathbf{w}}(t, T-1)\hat{\mathbf{w}}_N^+ \\ \mathbf{b}_N(t+1) &:= \hat{H}_{\mathbf{w}}(t+1, T)\hat{\mathbf{w}}_N^+\end{aligned}\quad (50)$$

and

$$\begin{aligned}\mathbf{v}_N(t) &:= H_s\hat{\mathbf{e}}_N^+ \\ \mathbf{v}_N(t+1) &:= H_s\hat{\mathbf{e}}_N^+\end{aligned}\quad (51)$$

Then the future output sequences  $\mathbf{y}_N^+ := [\mathbf{y}_N(t)^\top, \mathbf{y}_N(t+1)^\top, \dots, \mathbf{y}_N(T-1)^\top]^\top$  can be decomposed as follows:

$$\mathbf{y}_N^+ = \Gamma\hat{\mathbf{x}}_N(t) - \mathbf{b}_N(t) + \mathbf{v}_N(t) + H_{\mathbf{e}}(t, T-1)\hat{\mathbf{e}}_N^+ + H_{\mathbf{u}}\mathbf{u}_N^+ \quad (52)$$

and similarly

$$\underline{\mathbf{y}}_N^+ = \Gamma\hat{\mathbf{x}}_N(t+1) - \mathbf{b}_N(t+1) + \mathbf{v}_N(t+1) + H_{\mathbf{e}}(t+1, T)\hat{\mathbf{e}}_N^+ + H_{\mathbf{u}}\underline{\mathbf{u}}_N^+ \quad (53)$$

**Remark V.5** We would like to recall at this point that  $\mathbf{b}_N(t)$ , and  $\mathbf{v}_N(t)$  are finite tails formed with sample values of a random process and therefore are subject to random fluctuations. However, we shall call  $\mathbf{b}_N(t)$  “bias term” as it converges to zero as  $\|\bar{A}^{t-t_0}\| \left( \|\bar{F}^{t-t_0}\| + 1/\sqrt{N} \right)$  does, while we shall call “variance term” the other as it originates solely from approximation of expectations with finite sums.

It is common use in the subspace identification literature to assume that  $t - t_0$  goes to infinity “fast enough” (see for instance the assumptions in [8]) as  $N \rightarrow \infty$  so that the bias term can be neglected in the asymptotic variance computation. This approach would involve some technical complications which would make the paper even harder to access. We prefer therefore to follow a different line of thought: we will decompose the bias term  $\mathbf{b}_N(t)$  in its asymptotic value  $\mathbf{b}_\infty(t)$  plus a random fluctuation  $\Delta\mathbf{b}_N(t)$  which will contribute to the asymptotic variance. In the simplifying assumption that at least one of  $\bar{A}$ ,  $\bar{F}$  is nilpotent we have that  $\mathbf{b}_\infty(t) = 0$  and the more classical kind of results are obtained.  $\diamond$

## VI. STATE ESTIMATION WITH FINITE SEQUENCES

The first step in subspace identification is the construction of the state; we have seen in Section III that it basically consists of an oblique projection followed by a truncation step using the SVD.

Our program now, as outlined in Section IV, is to compute the terms  $\Delta\mathbf{x}_N(t)$ ,  $\Delta\mathbf{x}_N(t+1)$  in (34)

The state sequences  $\hat{\xi}_N(t)$  and  $\hat{\xi}_N(t+1)$  are constructed according to (26). From equations (52) and (53) we can see how the error terms  $\mathbf{b}_N$  and  $\mathbf{v}_N$  affect the state estimators. Defining  $\hat{T}_N := \hat{\Gamma}_N^{-L}\Gamma$  a simple manipulation shows in fact that

$$\hat{\xi}_N(t) = \hat{T}_N \hat{\mathbf{x}}_N(t) + \Delta\mathbf{x}_N(t) \quad (54)$$

and

$$\hat{\xi}_N(t+1) = \hat{T}_N \hat{\mathbf{x}}_N(t+1) + \Delta\mathbf{x}_N(t+1) \quad (55)$$

where

$$\Delta\mathbf{x}_N(t) := -\hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t,T-1]}} [\mathbf{b}_N(t) \mid Z_{[t_0,t]}] + \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t,T-1]}} [\mathbf{v}_N(t) \mid Z_{[t_0,t]}]$$

and

$$\Delta\mathbf{x}_N(t+1) := -\hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t+1,T]}} [\mathbf{b}_N(t+1) \mid Z_{[t_0,t+1]}] + \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t+1,T]}} [\mathbf{v}_N(t+1) \mid Z_{[t_0,t+1]}].$$

**Remark VI.6 [Non singularity<sup>13</sup> of  $\hat{T}_N$ ]**

<sup>13</sup>The author is indebted to an anonymous reviewer for suggesting this remark.

The data dependent change of basis  $\hat{T}_N$  needs to be non singular of course. This is guaranteed if, given  $\hat{\Gamma}_N$  estimated from (24),  $\hat{\Gamma}_N^\top W^{-\top} W^{-1} \Gamma$  is non singular; this certainly holds for  $N$  large enough when the column space of  $\Gamma$  is estimated consistently. Unfortunately this is not the case here since the bias term  $\mathbf{b}_N(t)$  in (52) does not vanish for  $N \rightarrow \infty$ . As we have seen the magnitude of  $\mathbf{b}_\infty(t)$  decreases as  $\bar{A}^{t-t_0}$  and  $\bar{F}^{t-t_0}$  do and therefore one way to guarantee that the effect of  $\mathbf{b}_N(t)$  is negligible is to chose  $t - t_0$  large enough. In any case this question deserves further analysis which we postpone to future work.  $\diamond$

These decompositions are the starting point for obtaining recursions for  $\hat{\xi}_N(t)$  involving the system parameters. The result, which we summarize in the next Proposition, follows from (54), (55) and Proposition 5.3.

*Proposition 6.1:* Let  $A_N^{ql} := \hat{T}_N A \hat{T}_N^{-1}$ ,  $B_N^{ql} := \hat{T}_N B$ ,  $C_N^{ql} := C \hat{T}_N^{-1}$ ,  $K_N^{ql} := \hat{T}_N K$ ,  $\bar{A}_N^{ql} := A_N^{ql} - K_N^{ql} C_N^{ql}$  be the state matrices in the (data dependent) basis  $\hat{T}_N$  and

$$\begin{aligned} (\Delta \mathbf{b}_x)_N &:= \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t+1, T]}} [\mathbf{b}_N(t+1) | Z_{[t_0, t+1]}] + \\ &\quad - A_N^{ql} \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\mathbf{b}_N(t) | Z_{[t_0, t]}] - \hat{T}_N \hat{\Psi}_t \hat{\mathbf{w}}_N(t) \\ (\Delta \mathbf{v}_x)_N &:= \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t+1, T]}} [\mathbf{v}_N(t+1) | Z_{[t_0, t+1]}] + \\ &\quad - A_N^{ql} \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\mathbf{v}_N(t) | Z_{[t_0, t]}] + K_N^{ql} \hat{\epsilon}_N(t) \\ (\Delta \mathbf{b}_y)_N &:= -C_N^{ql} \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\mathbf{b}_N(t) | Z_{[t_0, t]}] \\ (\Delta \mathbf{v}_y)_N &:= \hat{\epsilon}_N(t) - C_N^{ql} \hat{\Gamma}_N^{-L} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\mathbf{v}_N(t) | Z_{[t_0, t]}] \end{aligned}$$

Defining  $K_N^{ql}(t) := K_N^{ql} + \hat{T}_N \hat{\Theta}_t$ , the estimated state sequences  $\hat{\xi}_N(t)$  and  $\hat{\xi}_N(t+1)$  satisfy the following recursion:

$$\begin{cases} \hat{\xi}_N(t+1) = A_N^{ql} \hat{\xi}_N(t) + B_N^{ql} \mathbf{u}_N(t) + K_N^{ql}(t) \hat{\epsilon}_N(t) + \\ \quad - (\Delta \mathbf{b}_x)_N + (\Delta \mathbf{v}_x)_N \\ \mathbf{y}_N(t) = C_N^{ql} \hat{\xi}_N(t) + \hat{\epsilon}_N(t) + \\ \quad - (\Delta \mathbf{b}_y)_N + (\Delta \mathbf{v}_y)_N \end{cases} \quad (56)$$

which is of the form (35) where  $\Delta \mathbf{x}_N := -(\Delta \mathbf{b}_x)_N + (\Delta \mathbf{v}_x)_N$  and  $\Delta \mathbf{y}_N := -(\Delta \mathbf{b}_y)_N + (\Delta \mathbf{v}_y)_N$ .

Quite naturally Proposition 6.1 allows to compute exact expressions for the errors  $\tilde{A}_N^{ql}$ ,  $\tilde{B}_N^{ql}$ ,  $\tilde{C}_N^{ql}$  in the data dependent basis (step 3) in Section IV).

*Proposition 6.2:* Let  $\hat{\Xi}_t := \text{span}\{\hat{\xi}_N(t)\}$ ,  $\Delta \mathbf{x}_N := -(\Delta \mathbf{b}_x)_N + (\Delta \mathbf{v}_x)_N$  and  $\Delta \mathbf{y}_N := -(\Delta \mathbf{b}_y)_N +$

$(\Delta \mathbf{v}_y)_N$ . The error matrices  $\tilde{A}_N^{ql}$ ,  $\tilde{B}_N^{ql}$ ,  $\tilde{C}_N^{ql}$  obtained solving (27) (or equivalently (56)) satisfy

$$\begin{aligned}\tilde{A}_N^{ql} &= \hat{\Sigma}_{\Delta \mathbf{x} \hat{\xi} | \mathbf{g}} \hat{\Sigma}_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \\ \tilde{B}_N^{ql} &= \hat{\Sigma}_{\Delta \mathbf{x} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} \hat{\Sigma}_{\mathbf{u} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})}^{-1} \\ \tilde{C}_N^{ql} &= \hat{\Sigma}_{\Delta \mathbf{y} \hat{\xi}} \hat{\Sigma}_{\hat{\xi} \hat{\xi}}^{-1}\end{aligned}\tag{57}$$

*Proof:* We shall prove only the first of (57) since the others are completely analogous. By definition

$$\hat{A}_N^{ql} \hat{\xi}_N(t) := \hat{E}_{\|\hat{G}_t} \left[ \hat{\xi}_N(t+1) \mid \hat{\Xi}_t \right].$$

It follows from (56) that

$$\begin{aligned}\hat{E}_{\|\hat{G}_t} \left[ \hat{\xi}_N(t+1) \mid \hat{\Xi}_t \right] &= A_N^{ql} \hat{\xi}_N(t) + \hat{E}_{\|\hat{G}_t} \left[ \Delta \mathbf{x}_N \mid \hat{\Xi}_t \right] \\ &= \left( A_N^{ql} + \hat{\Sigma}_{\Delta \mathbf{x} \hat{\xi} | \mathbf{g}} \hat{\Sigma}_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \right) \hat{\xi}_N(t)\end{aligned}$$

from which  $\hat{A}_N^{ql} = A_N^{ql} + \hat{\Sigma}_{\Delta \mathbf{x} \hat{\xi} | \mathbf{g}} \hat{\Sigma}_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1}$  and therefore

$$\tilde{A}_N^{ql} = \hat{A}_N^{ql} - A_N^{ql} = \hat{\Sigma}_{\Delta \mathbf{x} \hat{\xi} | \mathbf{g}} \hat{\Sigma}_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1}$$

■

## VII. ASYMPTOTIC EXPRESSION FOR THE ERROR

The expression reported in Proposition 6.2 do not involve any approximation. Unfortunately it seems not possible to derive directly the asymptotic properties of the errors from (57). This Section is dedicated to deriving asymptotically equivalent formulas which can be used to compute the asymptotic distribution. We divide this in two steps:

- 1) We decompose the errors  $\tilde{A}_N^{ql}$ ,  $\tilde{B}_N^{ql}$ ,  $\tilde{C}_N^{ql}$ , into the sum of their asymptotic (in probability as  $N \rightarrow \infty$ ) values, which we denote  $\tilde{A}_\infty^{ql}$ ,  $\tilde{B}_\infty^{ql}$ ,  $\tilde{C}_\infty^{ql}$ , plus the random perturbations  $\Delta \tilde{A}_N^{ql}$ ,  $\Delta \tilde{B}_N^{ql}$ ,  $\Delta \tilde{C}_N^{ql}$ . The limiting values represent the bias of the algorithm, as studied in [17]. The differences ( $\Delta \tilde{A}_N^{ql}$ , etc.) between the actual errors ( $\tilde{A}_N^{ql}$ , etc.) and the limiting values ( $\tilde{A}_\infty^{ql}$ , etc.) are in fact random variables which, under reasonable assumptions [1], [27], [19], [13], can be shown to be asymptotically normal with zero mean and a certain (to be determined) covariance matrix.
- 2) We use standard techniques to obtain manageable expressions which allow to show asymptotic normality and compute the asymptotic variance as done in [1], [7], [29], [8], [13]. In fact standard results in asymptotic analysis (see [19], Theorem 6, p. 39) show that one can neglect terms which are  $o(\frac{1}{\sqrt{N}})$  in probability. Similar considerations hold for functions of the estimated parameters whose asymptotic behavior can be obtained using well known linearization techniques (see [19], Theorem 7, p. 45).

*Proposition 7.1:* The error matrices  $\tilde{A}_N^{ql}$ ,  $\tilde{B}_N^{ql}$ ,  $\tilde{C}_N^{ql}$  admit the decompositions  $\tilde{A}_N^{ql} = \tilde{A}_\infty^{ql} + \Delta\tilde{A}_N^{ql}$ ,  $\tilde{B}_N^{ql} = \tilde{B}_\infty^{ql} + \Delta\tilde{B}_N^{ql}$ ,  $\tilde{C}_N^{ql} = \tilde{C}_\infty^{ql} + \Delta\tilde{C}_N^{ql}$  where<sup>14</sup>

$$\begin{aligned}\tilde{A}_\infty^{ql} &= -\Sigma_{\Delta\mathbf{b}_x\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \\ \tilde{B}_\infty^{ql} &= -\Sigma_{\Delta\mathbf{b}_x\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1} \\ \tilde{C}_\infty^{ql} &= -\Sigma_{\Delta\mathbf{b}_y\hat{\xi}}\Sigma_{\hat{\mathbf{x}}\hat{\mathbf{x}}}^{-1}\end{aligned}$$

and

$$\begin{aligned}\Delta\tilde{A}_N^{ql} &\doteq \tilde{\Sigma}_{\Delta\mathbf{b}_x\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} + \tilde{A}_\infty^{ql}\tilde{\Sigma}_{\hat{\xi}\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} + \hat{\Sigma}_{\Delta\mathbf{v}_x\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \\ \Delta\tilde{B}_N^{ql} &\doteq \tilde{\Sigma}_{\Delta\mathbf{b}_x\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1} + \tilde{B}_\infty^{ql}\tilde{\Sigma}_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1} + \hat{\Sigma}_{\Delta\mathbf{v}_x\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1} \\ \Delta\tilde{C}_N^{ql} &\doteq \tilde{\Sigma}_{\Delta\mathbf{b}_y\hat{\xi}}\Sigma_{\hat{\xi}\hat{\xi}}^{-1} + \tilde{C}_\infty^{ql}\tilde{\Sigma}_{\hat{\xi}\hat{\xi}}\Sigma_{\hat{\xi}\hat{\xi}}^{-1} + \hat{\Sigma}_{\Delta\mathbf{v}_y\hat{\xi}}\Sigma_{\hat{\xi}\hat{\xi}}^{-1}\end{aligned}$$

The first two terms in the linearization vanish when the estimation is not biased, i.e. when  $(\Delta\mathbf{b}_x)_N = 0$  and  $(\Delta\mathbf{b}_y)_N = 0$ . Something similar happens also in the analysis of certain open loop subspace algorithms which are biased for  $t - t_0$  finite. Following the approach of [1], [2] we could guarantee that

$$\begin{aligned}&\sqrt{N}(\tilde{\Sigma}_{\Delta\mathbf{b}_x\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} + \tilde{A}_\infty^{ql}\tilde{\Sigma}_{\hat{\xi}\hat{\xi}|\mathbf{g}}\Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1}) \\ &\sqrt{N}(\tilde{\Sigma}_{\Delta\mathbf{b}_x\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1} + \tilde{B}_\infty^{ql}\tilde{\Sigma}_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\xi},\hat{\mathbf{e}})}^{-1}) \\ &\sqrt{N}(\tilde{\Sigma}_{\Delta\mathbf{b}_y\hat{\xi}}\Sigma_{\hat{\xi}\hat{\xi}}^{-1} + \tilde{C}_\infty^{ql}\tilde{\Sigma}_{\hat{\xi}\hat{\xi}}\Sigma_{\hat{\xi}\hat{\xi}}^{-1})\end{aligned}$$

go to zero in probability. Under this condition only the terms related to  $\Delta\mathbf{v}_x$  and  $\Delta\mathbf{v}_y$  would have to be accounted for variance. Both this approach and the computation of the variance in the general case (for  $t - t_0$  finite) would introduce further technical complications.

For reasons of space and ease of exposition we prefer at this point to make the assumption that at least one of the equalities  $\bar{A}^{t-t_0} = 0$ ,  $\bar{F}^{t-t_0} = 0$  holds. This, as we have shown in [17], implies that  $\tilde{A}_\infty^{ql} = 0$ ,  $\tilde{B}_\infty^{ql} = 0$ ,  $\tilde{C}_\infty^{ql} = 0$  (and also  $(\Delta\mathbf{b}_x)_N = 0$ ,  $(\Delta\mathbf{b}_y)_N = 0$ ; see Proposition 6.1, Lemma 5.2 and equation (50).

## VIII. ASYMPTOTIC VARIANCE

In this section we shall state the main result concerning asymptotic normality and give expression for asymptotic variance of the estimators. We remind the reader that this last section deals with the simplified case in which the ‘‘true’’ system is of the ARX type (i.e.  $\|\bar{A}^{t-t_0}\| = 0$ ). As discussed in Remark V.4 and

<sup>14</sup>The symbol  $\hat{\xi}$  in the asymptotic expression indicates that, also asymptotically in  $N$ , the estimated state  $\hat{\xi}_N$  is based on a finite past window of length  $t - t_0$ . The same will hold for  $\hat{\mathbf{e}}$  and  $\hat{\zeta}$ .

in [17] the assumption  $\|\bar{A}^{t-t_0}\| \|\bar{F}^{t-t_0}\| = 0$  would suffice for the algorithm by Qin and Ljung to yield consistent estimators. However, we prefer here to work under the more stringent assumption  $\|\bar{A}^{t-t_0}\| = 0$  to simplify the (already complicated) variance formulas.

Theorem 8.1 states the main results also for the “whitening filter” approach in which case the assumption that the true system is ARX (i.e.  $\|\bar{A}^{t-t_0}\| = 0$ ) is also needed. The reader is referred to Appendix B for the notation.

Similar results, but at the price of more complicated assumptions and derivations would hold if  $t - t_0 \rightarrow \infty$  at a suitable rate [8] with  $N$ . We stress however that the tools developed so far allows to deal with the general case ( $(t - t_0)$  finite and no assumptions on  $\bar{A}$  or on the feedback channel) at the price of course of more complicated formulas.

We shall need to make the following standard assumption on the innovation process  $\mathbf{e}(t)$ . Let us denote with  $\mathcal{F}_t^-(\mathbf{z})$  the  $\sigma$ -algebra generated by the past (up to time  $t$  excluded) of the process  $\{\mathbf{z}(s)\}_{s \in \mathbb{Z}}$  and denote with  $\mathbb{E}[\cdot | \mathcal{F}_t^-(\mathbf{z})]$  conditional expectation with respect to  $\mathcal{F}_t^-(\mathbf{z})$ .

- A1. The process  $\{\mathbf{e}(t)\}$  is a  $\mathcal{F}_t^-(\mathbf{z})$ -martingale difference sequence, i.e. a martingale difference with respect to the  $\sigma$ -algebra  $\mathcal{F}_t^-(\mathbf{z})$  generated by the past measurements, with constant conditional variance, i.e.

$$\begin{aligned} \mathbb{E}[\mathbf{e}(t) | \mathcal{F}_t^-(\mathbf{z})] &= 0 \\ \mathbb{E}[\mathbf{e}(t)\mathbf{e}^\top(t) | \mathcal{F}_t^-(\mathbf{z})] &= \Lambda_e. \end{aligned} \tag{58}$$

- A2. The process  $\{\mathbf{e}(t)\}$  has bounded fourth order moments, i.e.  $\mathbb{E}[\mathbf{e}_i^4(t)] < \infty$  where  $\mathbf{e}_i(t)$  is the  $i$ -th component of  $\mathbf{e}(t)$ .

We now state the main result of the paper:

*Theorem 8.1:* Assume that Assumptions 3.1, A1 and A2 hold. Assume also that in the “true” system generating the data  $\|\bar{A}^{t-t_0}\| = 0$  holds. Then  $\sqrt{N}\text{vec}\left(\tilde{A}_N^{ql}\right)$ ,  $\sqrt{N}\text{vec}\left(\tilde{B}_N^{ql}\right)$  and  $\sqrt{N}\text{vec}\left(\tilde{C}_N^{ql}\right)$  are jointly asymptotically normal with zero mean and asymptotic covariance matrix

$$\text{AsVar} \left\{ \sqrt{N} \begin{bmatrix} \text{vec}\left(\tilde{A}_N^{ql}\right) \\ \text{vec}\left(\tilde{B}_N^{ql}\right) \\ \text{vec}\left(\tilde{C}_N^{ql}\right) \end{bmatrix} \right\} = M^{ql} \begin{pmatrix} \Sigma_{11}^{ql} & \Sigma_{12}^{ql} \\ \left(\Sigma_{12}^{ql}\right)^\top & \Sigma_{22}^{ql} \end{pmatrix} \left(M^{ql}\right)^\top \tag{59}$$

The matrix  $M^{ql}$  is of the form

$$M^{ql} := \begin{pmatrix} M_{A1}^{ql} P & M_{A2}^{ql} \bar{P} \\ M_{B1}^{ql} P & M_{B2}^{ql} \bar{P} \\ M_C^{ql} P & 0 \end{pmatrix}$$

where the various symbols will be defined in the remaining part of this Section. In particular  $M_{A1}^{ql}$ ,  $M_{A2}^{ql}$ , etc. will be defined in (61),  $P$  and  $\bar{P}$  in (64) and  $\Sigma_{ij}^{ql}$  in (75), (76) and (77).

Also under the assumption that  $\bar{A}^{t-t_0} = 0$  a completely analogous statement holds also for  $\sqrt{N} \text{vec} \left( \tilde{A}_N^{wf} \right) = \sqrt{N} \text{vec} \left( \hat{A}_N^{wf} - A_N^{wf} \right)$ ,  $\sqrt{N} \text{vec} \left( \tilde{B}_N^{wf} \right) = \sqrt{N} \text{vec} \left( \hat{B}_N^{wf} - B_N^{wf} \right)$  and  $\sqrt{N} \text{vec} \left( \tilde{C}_N^{wf} \right) = \sqrt{N} \text{vec} \left( \hat{C}_N^{wf} - C_N^{wf} \right)$  computed using the “whitening filter” approach. The formulas are the same as above provided one substitutes the superscript  $^{ql}$  with the superscript  $^{wf}$ . The definition of the various symbols are found in Appendix B. In particular  $M_{A1}^{wf}$ ,  $M_{A2}^{wf}$ , etc. will be defined in (B.89), and  $\Sigma_{ij}^{wf}$  in (B.98), (B.99) and (B.100).

In the remaining of this section we shall state some intermediate results which allow to introduce the notation required; the proof of theorem 8.1 for the “innovation estimation” approach shall conclude the Section while the part concerning the “whitening filter” approach shall be found in Appendix B.

Recall that under the assumption that  $\bar{A}^{t-t_0} = 0$

$$\begin{aligned} \tilde{A}_N^{ql} &= \Delta \tilde{A}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}} \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \\ \tilde{B}_N^{ql} &= \Delta \tilde{B}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_x \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})}^{-1} \\ \tilde{C}_N^{ql} &= \Delta \tilde{C}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_y \hat{\xi}} \Sigma_{\hat{\xi} \hat{\xi}}^{-1} \end{aligned}$$

Therefore we shall need to study the asymptotic properties of  $\hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}} \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1}$ ,  $\hat{\Sigma}_{\Delta \mathbf{v}_x \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})}^{-1}$  and  $\hat{\Sigma}_{\Delta \mathbf{v}_y \hat{\xi}} \Sigma_{\hat{\xi} \hat{\xi}}^{-1}$ .

*Lemma 8.2:* Let us introduce the shorthands  $\mathbf{g}^+ := \mathbf{g}_{[t, T-1]}$ ,  $\underline{\mathbf{g}}^+ := \mathbf{g}_{[t+1, T]}$ ,  $\mathbf{z} := \mathbf{z}_{[t_0, t]}$ ,  $\bar{\mathbf{z}} := \mathbf{z}_{[t_0, t+1]}$ ,  $\hat{\epsilon}^+ := \hat{\epsilon}_{[t, T-1]}$  and  $\hat{\underline{\epsilon}}^+ := \hat{\epsilon}_{[t+1, T]}$ . Then

$$\begin{aligned} \hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}} &= \hat{M}_{x1}^{ql} \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} \hat{\Sigma}_{\mathbf{z} \hat{\xi} | \mathbf{g}} + \hat{M}_{x2}^{ql} \hat{\Sigma}_{\hat{\underline{\epsilon}}^+ \bar{\mathbf{z}} | \underline{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \underline{\mathbf{g}}^+}^{-1} \hat{\Sigma}_{\bar{\mathbf{z}} \hat{\xi} | \mathbf{g}} \\ \hat{\Sigma}_{\Delta \mathbf{v}_x \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} &= \hat{M}_{x1}^{ql} \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} \hat{\Sigma}_{\mathbf{z} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} + \hat{M}_{x2}^{ql} \hat{\Sigma}_{\hat{\underline{\epsilon}}^+ \bar{\mathbf{z}} | \underline{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \underline{\mathbf{g}}^+}^{-1} \hat{\Sigma}_{\bar{\mathbf{z}} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} \\ \hat{\Sigma}_{\Delta \mathbf{v}_y \hat{\xi}} &= \hat{M}_y^{ql} \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} \hat{\Sigma}_{\mathbf{z} \hat{\xi}} \end{aligned} \tag{60}$$

where  $\hat{M}_{x2}^{ql} := \hat{\Gamma}_N^{-L} H_s$ ,  $\hat{M}_{x1}^{ql} := \left( K_N^{ql} [I \ 0] - A_N^{ql} \hat{\Gamma}_N^{-L} H_s \right)$ ,  $\hat{M}_y^{ql} := \left( [I \ 0] - C_N^{ql} \hat{\Gamma}_N^{-L} H_s \right)$ .

*Proof:* See Appendix A ■

Using standard results in asymptotic theory [19] one can prove the following.

*Proposition 8.3:* Let as usual denote the asymptotic values (in probability as  $N \rightarrow \infty$ ) of matrices

removing the “hat” from the symbol denoting the sample version (e.g.  $M_{x2}^{ql} := \lim_{N \rightarrow \infty} \hat{M}_{x2}^{ql}$ ) and let<sup>15</sup>

$$\begin{aligned}
M_{A1}^{ql} &:= \left[ \left( \Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \Sigma_{\mathbf{z}\hat{\xi}|\mathbf{g}}^\top \right) \otimes M_{x1}^{ql} \right] \\
M_{A2}^{ql} &:= \left[ \left( \Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \Sigma_{\bar{\mathbf{z}}\hat{\xi}|\mathbf{g}}^\top \right) \otimes M_{x2}^{ql} \right] \\
M_{B1}^{ql} &:= \left[ \left( \Sigma_{\mathbf{u}\mathbf{u}|\hat{\xi},\hat{\mathbf{e}}}^{-1} \Sigma_{\mathbf{z}\mathbf{u}|\hat{\xi},\hat{\mathbf{e}}}^\top \right) \otimes M_{x1}^{ql} \right] \\
M_{B2}^{ql} &:= \left[ \left( \Sigma_{\mathbf{u}\mathbf{u}|\hat{\xi},\hat{\mathbf{e}}}^{-1} \Sigma_{\bar{\mathbf{z}}\mathbf{u}|\hat{\xi},\hat{\mathbf{e}}}^\top \right) \otimes M_{x2}^{ql} \right] \\
M_C^{ql} &:= \left[ \left( \Sigma_{\hat{\xi}\hat{\xi}}^{-1} \Sigma_{\mathbf{z}\hat{\xi}}^\top \right) \otimes M_y^{ql} \right]
\end{aligned} \tag{61}$$

Then the following first order approximation (up to  $o(\frac{1}{\sqrt{N}})$  terms in probability) of vectorized sample covariances can be used to compute the asymptotic distribution:

$$\begin{aligned}
\text{vec} \left( \hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi}|\mathbf{g}} \Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \right) &\doteq \begin{bmatrix} M_{A1}^{ql} & M_{A2}^{ql} \end{bmatrix} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{g}^+}^{-1} \right) \end{bmatrix} \\
\text{vec} \left( \hat{\Sigma}_{\Delta \mathbf{v}_x \mathbf{u}|\hat{\xi},\hat{\mathbf{e}}} \Sigma_{\mathbf{u}\mathbf{u}|\hat{\xi},\hat{\mathbf{e}}}^{-1} \right) &\doteq \begin{bmatrix} M_{B1}^{ql} & M_{B2}^{ql} \end{bmatrix} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{g}^+}^{-1} \right) \end{bmatrix} \\
\text{vec} \left( \hat{\Sigma}_{\Delta \mathbf{v}_y \hat{\xi}} \Sigma_{\hat{\xi}\hat{\xi}}^{-1} \right) &\doteq M_C^{ql} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right)
\end{aligned} \tag{62}$$

*Proof:* See Appendix A ■

### Remark VIII.7 [Convergence]

We remind the reader that convergence to a limit of  $\hat{T}_N$  (and therefore of  $A_N^{ql}$ ,  $\hat{M}_{x1}^{ql}$ , etc.) holds only under Assumption 3.1. However the reader should keep clear in mind that the actual implementation of the formulas for a specific value of  $N$  does not need any specific choice of  $\hat{T}$  which, without loss of generality, the user can choose at will. The same remark applies to the “whitening filter” approach. In other words, there is no need to transform the estimates into some canonical form to compute the variance of system invariants. ◇

The last effort left is to compute the asymptotic variance of

$$\begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{g}^+}^{-1} \right) \end{bmatrix}$$

Let now  $P_k$  be a permutation matrix so that, given a matrix  $A_k^\top := [a_1^\top, a_2^\top, \dots, a_{T-t}^\top]$  where  $a_i$  has  $p$

<sup>15</sup>The symbol  $\otimes$  denotes Kronecker product, see [24].

rows and  $k$  columns then

$$\text{vec} \{A_k\} = P_k \begin{bmatrix} \text{vec} \{a_1\} \\ \text{vec} \{a_2\} \\ \vdots \\ \text{vec} \{a_{T-t}\} \end{bmatrix}. \quad (63)$$

For ease on notation, let  $q = p + m$  and define

$$P := P_{q(t-t_0)} \quad , \quad \bar{P} = P_{q(t-t_0+1)}. \quad (64)$$

It is now easy to verify that  $\text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right\}$  can be written in the following form:

$$\text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}+\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right\} = P \begin{bmatrix} \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_t|\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right\} \\ \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_{t+1}|\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right\} \\ \vdots \\ \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_{T-1}|\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} \right\} \end{bmatrix} \quad (65)$$

and similarly

$$\text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}+\bar{\mathbf{z}}|\bar{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\bar{\mathbf{g}}^+}^{-1} \right\} = \bar{P} \begin{bmatrix} \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_{t+1}|\bar{\mathbf{z}}|\bar{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\bar{\mathbf{g}}^+}^{-1} \right\} \\ \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_{t+2}|\bar{\mathbf{z}}|\bar{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\bar{\mathbf{g}}^+}^{-1} \right\} \\ \vdots \\ \text{vec} \left\{ \hat{\Sigma}_{\hat{\epsilon}_T|\bar{\mathbf{z}}|\bar{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\bar{\mathbf{g}}^+}^{-1} \right\} \end{bmatrix} \quad (66)$$

This is useful because, as we shall show, it is rather easy to compute the asymptotic covariance of the (block) rows of the matrices on the right hand side of equations (65) and (66).

An instrumental result is the following:

*Lemma 8.4:* The oblique projection  $\hat{E}_{\|\hat{G}_{[s,T]}} [\hat{\epsilon}_N(s+k) | Z_{[t_0,s]}]$  satisfies the following:

$$\begin{aligned} \hat{E}_{\|\hat{G}_{[s,T]}} [\hat{\epsilon}_N(s+k) | Z_{[t_0,s]}] &= \hat{E}_{\|\hat{G}_{[s,s+k-1]}} [\hat{\epsilon}_N(s+k) | Z_{[t_0,s]}] = \\ &= \hat{E}_{\|\hat{G}_{[s,s+k-1]}} [\mathbf{e}_N(s+k) | Z_{[t_0,s]}]. \end{aligned} \quad (67)$$

Let us define the shorthands  $\mathbf{g}^k := \mathbf{g}_{[t,t+k]}$  and  $\bar{\mathbf{g}}^k := \mathbf{g}_{[t+1,t+k]}$ .

From the fact that  $Z_{[t_0,T]}$  is assumed to be of full row rank, (67) is equivalent to the ‘‘dual’’ equation for the coefficients ( $k = 0, \dots, T-t-1$ )

$$\hat{\Sigma}_{\hat{\epsilon}_{t+k}|\mathbf{z}|\mathbf{g}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^+}^{-1} = \hat{\Sigma}_{\hat{\epsilon}_{t+k}|\mathbf{z}|\mathbf{g}^k} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^k}^{-1} = \hat{\Sigma}_{\mathbf{e}_{t+k}|\mathbf{z}|\mathbf{g}^k} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{g}^k}^{-1} \quad (68)$$

and ( $k = 1, \dots, T-t$ )

$$\hat{\Sigma}_{\hat{e}_{t+k} \bar{\mathbf{z}} | \mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^+}^{-1} = \hat{\Sigma}_{\hat{e}_{t+k} \bar{\mathbf{z}} | \mathbf{g}^k} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^k}^{-1} = \hat{\Sigma}_{\mathbf{e}_{t+k} \bar{\mathbf{z}} | \mathbf{g}^k} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^k}^{-1} \quad (69)$$

*Proof:* Recall from (41) that  $\hat{e}_N(t+k) = \hat{E}[\mathbf{e}_N(t+k) | Z_{[t_0, t+k]}]$ ; from the decomposition  $Z_{[t_0, t+k]} = Z_{[t_0, t]} + \hat{G}_{[t, t+k-1]}$  then (67) follows.  $\blacksquare$

It follows from (65), and (66) together with (68) and (69) that the asymptotic variance we are seeking can be expressed in terms of variances and covariances of terms of the form

$$\sqrt{N} \hat{R}_k^{ql} := \sqrt{N} \hat{\Sigma}_{\mathbf{e}_{t+k} \mathbf{z} | \mathbf{g}^k} \Sigma_{\mathbf{z} \mathbf{z} | \mathbf{g}^k}^{-1} \doteq \sqrt{N} \hat{\Sigma}_{\mathbf{e}_{t+k} \mathbf{z} | \mathbf{g}^k} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^k}^{-1} \quad k = 0, \dots, T-t-1 \quad (70)$$

and

$$\sqrt{N} \hat{R}_k^{gl} := \sqrt{N} \hat{\Sigma}_{\mathbf{e}_{t+k} \bar{\mathbf{z}} | \mathbf{g}^k} \Sigma_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^k}^{-1} \doteq \sqrt{N} \hat{\Sigma}_{\mathbf{e}_{t+k} \bar{\mathbf{z}} | \mathbf{g}^k} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^k}^{-1} \quad k = 1, \dots, T-t \quad (71)$$

which we compute in the next proposition.

*Proposition 8.5:* Let assumptions A1 and A2 be satisfied. Let us define<sup>16</sup>

$$\begin{aligned} \mathbf{z}_{[t_0+\tau, t+\tau]}^{gk} &:= E \left[ \mathbf{z}_{[t_0+\tau, t+\tau]} \mid \left( \mathcal{G}_{[t+\tau, t+\tau+k]}^{t_0+\tau} \right)^\perp \right], \\ \bar{\mathbf{z}}_{[t_0+\tau, t+1+\tau]}^{gk} &:= E \left[ \bar{\mathbf{z}}_{[t_0+\tau, t+1+\tau]} \mid \left( \mathcal{G}_{[t+1+\tau, t+\tau+k]}^{t_0+\tau} \right)^\perp \right], \\ \Sigma_{\mathbf{z}^{gk} \mathbf{z}^{gh}}(\tau) &:= E \left[ \mathbf{z}_{[t_0+\tau, t+\tau]}^{gk} \left( \mathbf{z}_{[t_0, t]}^{gh} \right)^\top \right]. \end{aligned}$$

Similar definitions hold for  $\Sigma_{\bar{\mathbf{z}}^{gk} \bar{\mathbf{z}}^{gh}}(\tau)$  and  $\Sigma_{\mathbf{z}^{gk} \bar{\mathbf{z}}^{gh}}(\tau)$ . Let also<sup>17</sup>  $Q_k^{ql} := \left( \Sigma_{\mathbf{z} \mathbf{z} | \mathbf{g}^k}^{-1} \otimes I \right)$ ,  $\bar{Q}_k^{ql} := \left( \Sigma_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^k}^{-1} \otimes I \right)$  and  $\tau := h - k$ , then:

$$\Sigma_{11}^{ql}(k, h) := \text{AsCov} \left\{ \sqrt{N} \text{vec} \left( \hat{R}_k^{ql} \right), \sqrt{N} \text{vec} \left( \hat{R}_h^{ql} \right) \right\} = Q_k^{ql} \left( \Sigma_{\mathbf{z}^{gk} \mathbf{z}^{gh}}(\tau) \otimes \Lambda_e \right) \left( Q_h^{ql} \right)^\top \quad (72)$$

Similarly

$$\Sigma_{22}^{ql}(k, h) := \text{AsCov} \left\{ \sqrt{N} \text{vec} \left( \hat{R}_k^{gl} \right), \sqrt{N} \text{vec} \left( \hat{R}_h^{gl} \right) \right\} = \bar{Q}_k^{ql} \left( \Sigma_{\bar{\mathbf{z}}^{gk} \bar{\mathbf{z}}^{gh}}(\tau) \otimes \Lambda_e \right) \left( \bar{Q}_h^{ql} \right)^\top \quad (73)$$

As far as the cross terms are concerned, we have

$$\Sigma_{12}^{ql}(k, h) := \text{AsCov} \left\{ \sqrt{N} \text{vec} \left( \hat{R}_k^{ql} \right), \sqrt{N} \text{vec} \left( \hat{R}_h^{gl} \right) \right\} = Q_k^{ql} \left( \Sigma_{\mathbf{z}^{gk} \bar{\mathbf{z}}^{gh}}(\tau) \otimes \Lambda_e \right) \left( \bar{Q}_h^{ql} \right)^\top \quad (74)$$

*Proof:* See Appendix A.  $\blacksquare$

<sup>16</sup>The superscript  $^{gk}$  is meant to remind that the process  $\mathbf{z}$  is taken conditionally on the  $k$  lags of the future inputs  $\mathbf{g}$ .

<sup>17</sup>Note that  $\Sigma_{\mathbf{z} \mathbf{z} | \mathbf{g}^k} = \Sigma_{\mathbf{z}^{gk} \mathbf{z}^{gk}}(0)$ .

Defining

$$\Sigma_{11}^{ql} := \begin{bmatrix} \Sigma_{11}^{ql}(0,0) & \dots & \Sigma_{11}^{ql}(0, T-t-1) \\ \vdots & \ddots & \vdots \\ \Sigma_{11}^{ql}(T-t-1,0) & \dots & \Sigma_{11}^{ql}(T-t-1, T-t-1) \end{bmatrix}, \quad (75)$$

$$\Sigma_{22}^{ql} := \begin{bmatrix} \Sigma_{22}^{ql}(1,1) & \dots & \Sigma_{22}^{ql}(1, T-t) \\ \vdots & \ddots & \vdots \\ \Sigma_{22}^{ql}(T-t,1) & \dots & \Sigma_{22}^{ql}(T-t, T-t) \end{bmatrix}, \quad (76)$$

and

$$\Sigma_{12}^{ql} := \begin{bmatrix} \Sigma_{12}^{ql}(0,1) & \dots & \Sigma_{12}^{ql}(0, T-t) \\ \vdots & \ddots & \vdots \\ \Sigma_{12}^{ql}(T-t-1,1) & \dots & \Sigma_{12}^{ql}(T-t-1, T-t) \end{bmatrix}, \quad (77)$$

from Proposition 8.5 and equations (65), (66) it follows that

$$\text{AsVar} \left\{ \sqrt{N} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}^+ \bar{\mathbf{z}} | \bar{\mathbf{g}}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \bar{\mathbf{g}}^+}^{-1} \right) \end{bmatrix} \right\} = \begin{pmatrix} P & 0 \\ 0 & \bar{P} \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{ql} & \Sigma_{12}^{ql} \\ \left( \Sigma_{12}^{ql} \right)^\top & \Sigma_{22}^{ql} \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & \bar{P} \end{pmatrix}^\top \quad (78)$$

Collecting the results obtained so far we have basically proven the main result of the paper.

*Proof of Theorem 8.1.* Recalling that under the assumption that  $\bar{A}^{t-t_0} = 0$

$$\begin{aligned} \tilde{A}_N^{ql} &= \Delta \tilde{A}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}} \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \\ \tilde{B}_N^{ql} &= \Delta \tilde{B}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_x \mathbf{u} | (\hat{\xi}, \hat{\epsilon})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\xi}, \hat{\epsilon})}^{-1} \\ \tilde{C}_N^{ql} &= \Delta \tilde{C}_N^{ql} \doteq \hat{\Sigma}_{\Delta \mathbf{v}_y \hat{\xi}} \Sigma_{\hat{\xi} \hat{\xi}}^{-1} \end{aligned}$$

asymptotic normality follows from standard results on asymptotic properties of sample covariances [41][p. 550]. Using Propositions 8.3, 8.5 and equation (78) the asymptotic variance formula follows.  $\blacksquare$

## IX. SIMULATIONS

In this section we consider the following simulation setup: the first example is a first order ARX system

$$A_1 = 0.5 \quad C_1 = 1 \quad B_1 = 1 \quad C_1 = 1 \quad K_1 = 0.5 \quad D_1 = 0$$

while the second is a first order ARMAX model

$$A_2 = 0.5 \quad C_2 = 1 \quad B_2 = 1 \quad C_2 = 1 \quad K_2 = 1 \quad D_2 = 0$$

The input is of the form

$$\mathbf{u}(t) = 5\mathbf{r}(t) - \mathbf{y}(t)$$

where the reference signal  $\mathbf{r}(t)$  is either unit variance white noise or unit variance white noise passed through the filter with state space realization:

$$A_{\mathbf{r}} = \begin{bmatrix} 0 & 1 \\ -0.9 & 0.5 \end{bmatrix} \quad B_{\mathbf{r}} = \begin{bmatrix} 1.3 \\ .3 \end{bmatrix}$$

$$C_{\mathbf{r}} = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad D_{\mathbf{r}} = 1$$

We report results comparing the asymptotic variance of the deterministic transfer functions estimators  $\hat{F}_i(z) = \hat{C}_i(zI - \hat{A}_i)^{-1}\hat{B}_i$ ,  $i = 1, 2$ , computed using the expressions of this paper with the sample variance obtained from 500 Monte Carlo runs. Each experiment has been performed using  $N = 1000$  data points. The future and past horizon have been chosen (somewhat arbitrarily) as follows:  $t - t_0 = 5$  and  $T - t = 5$  for the ARX example and  $t - t_0 = 10$ ,  $T - t = 5$  for the ARMAX one.

Note that in this second case  $\bar{A}_2 = A_2 - K_2C_2 = 0.5 - 1 = -0.5 \neq 0$  and therefore strictly speaking the results of the paper do not apply. However the past horizon  $t - t_0$  is chosen equal to 10 so that  $\bar{A}_2^{10} = (-0.5)^{10} \simeq 9.8 \cdot 10^{-4}$  is negligible to all practical purposes.

The experimental results show that indeed the variance computed using the formulas of this paper is in very good agreement with the sample variance obtained from the simulations.

The original algorithm by Jansson [30] is indistinguishable from the “geometric version” called “whitening filter” in this paper. The asymptotic variance of “whitening filter” (computed using the formulas of this paper and estimated from the simulations) is indistinguishable from the Cramér-Rao lower bound for the first three examples. In the ARMAX case with colored reference the Cramér-Rao lower bound is instead smaller than the asymptotic variance for both algorithms analyzed in this paper (see figure 2, right plot). The interested reader may find in [10] an analysis of the “whitening filter” approach in the feedback free case.

## X. CONCLUSIONS

We have derived the asymptotic covariance of the system parameters estimated by (a version of) the algorithms proposed in [40], [30], [17] under a simplifying assumption. The tools developed in this paper allows however to deal with the general case ( $(t - t_0)$  finite and no assumptions on  $\bar{A}$  or on the feedback channel), which we postpone to future work. In fact if the nilpotency assumptions are not satisfied,

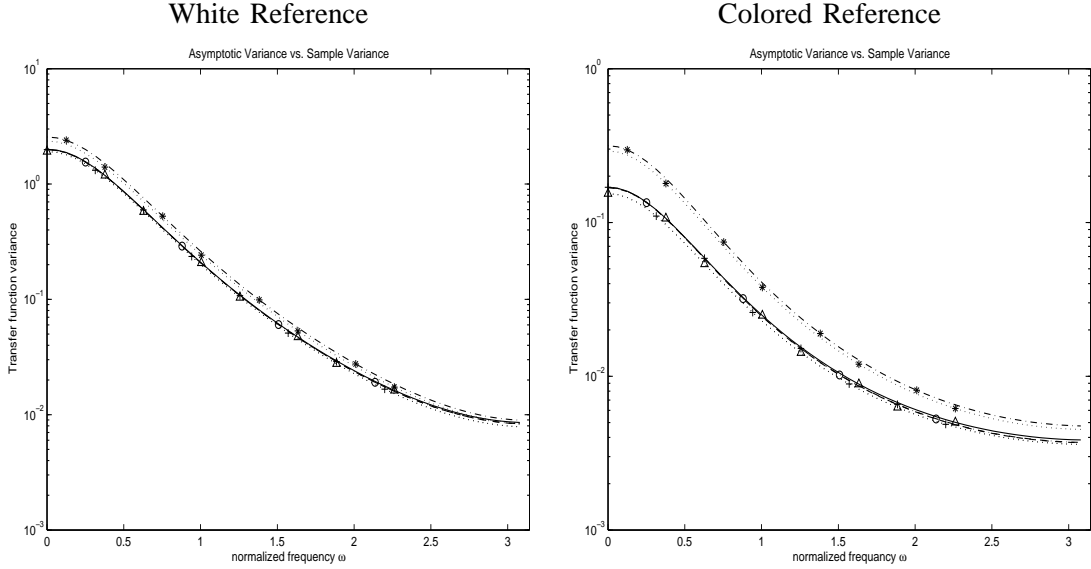


Fig. 1. EXAMPLE 1 (ARX of order 1): Asymptotic Variance and Sample Variance (Monte Carlo estimate) vs. normalized frequency ( $\omega \in [0, \pi]$ ) Solid with triangles ( $\triangle$ ) PEM, dashed with stars ( $*$ ): “innovation-estimation”, dashed with crosses ( $+$ ): “whitening-filter” algorithm, dotted with circles ( $o$ ): Jansson’s algorithm, dotted with crosses ( $+$ ): asymptotic variance for “whitening-filter”, dotted with stars ( $*$ ): asymptotic variance for “innovation-estimation”, dotted with triangles ( $\triangle$ ): Cramér Rao lower bound.

the estimators will be (asymptotically) biased ( $\tilde{A}_\infty^{ql} \neq 0$ ,  $\tilde{A}_\infty^{wf} \neq 0$ , etc.) and the deviation from the asymptotic values described by more complicated expressions (see Propositions 7.1 and B.4). However, as some simulation results confirm, the approximation is reasonable also the system is not of the ARX type.

#### APPENDIX A

*Proof of Lemma 8.2.* Recall that, since  $\hat{\epsilon}_N(t) \in Z_{[t_0, t]}$ ,  $\hat{\epsilon}_N(t) = \hat{E}_{\|\hat{G}_{[t, T]}}[\hat{\epsilon}_N(t) | Z_{[t_0, t]}]$ .

Therefore, using the definition of  $\mathbf{v}_N$  in (51) we can write  $(\Delta \mathbf{v}_x)_N$  from Proposition 6.1 in compact form as follows:

$$\begin{aligned}
 (\Delta \mathbf{v}_x)_N &= \hat{\Gamma}_N^{-L} H_s \hat{E}_{\|\hat{G}_{[t+1, T]}}[\hat{\epsilon}_N^+ | Z_{[t_0, t+1]}] + \\
 &\quad + \left( K_N^{ql} [I \ 0] - \hat{A}_N^{ql} \hat{\Gamma}_N^{-L} H_s \right) \hat{E}_{\|\hat{G}_{[t, T-1]}}[\hat{\epsilon}_N^+ | Z_{[t_0, t]}] = \\
 &= \hat{M}_{x2}^{ql} \hat{E}_{\|\hat{G}_{[t+1, T]}}[\hat{\epsilon}_N^+ | Z_{[t_0, t+1]}] + \hat{M}_{x1}^{ql} \hat{E}_{\|\hat{G}_{[t, T-1]}}[\hat{\epsilon}_N^+ | Z_{[t_0, t]}]
 \end{aligned} \tag{A.79}$$

where the last equality defines  $\hat{M}_{x2}^{ql}$ ,  $\hat{M}_{x1}^{ql}$ .

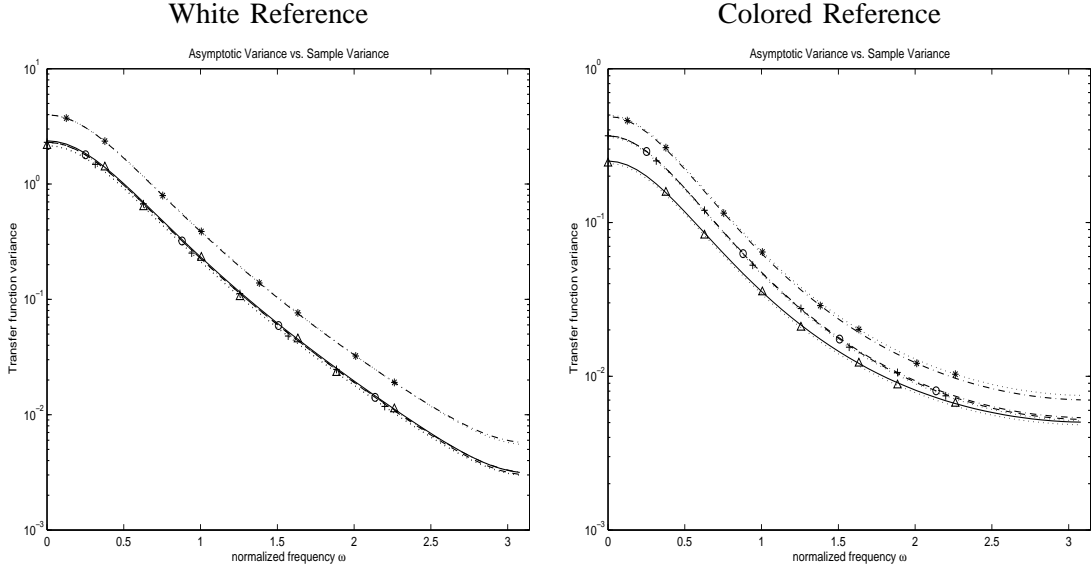


Fig. 2. EXAMPLE 2 (ARMAX of order 1): Asymptotic Variance and Sample Variance (Monte Carlo estimate) vs. normalized frequency ( $\omega \in [0, \pi]$ ) Solid with triangles ( $\Delta$ ) PEM, dashed with stars ( $*$ ): “innovation-estimation”, dashed with crosses ( $+$ ) “whitening-filter” algorithm, dotted with circles ( $o$ ): Jansson’s algorithm, dotted with crosses ( $+$ ): asymptotic variance for “whitening-filter”, dotted with stars ( $*$ ): asymptotic variance for “innovation-estimation”, dotted with triangles ( $\Delta$ ): Cramér Rao lower bound.

Similarly we have

$$\begin{aligned} (\Delta \mathbf{v}_y)_N &= \left( [I \ 0] - \hat{C}_N^{ql} \hat{\Gamma}_N^{-L} H_s \right) \cdot \hat{E}_{\|\hat{G}_{[t, T-1]}} [\hat{\epsilon}_N^+ | Z_{[t_0, t]}] = \\ &= \hat{M}_y^{ql} \hat{E}_{\|\hat{G}_{[t, T-1]}} [\hat{\epsilon}_N^+ | Z_{[t_0, t]}] \end{aligned} \quad (\text{A.80})$$

where the last equality defines  $\hat{M}_y^{ql}$ .

Rewriting (A.79) as follows

$$(\Delta \mathbf{v}_x)_N = \hat{M}_{x1}^{ql} \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} (\mathbf{z}_{[t_0, t]})_N + \hat{M}_{x2}^{ql} \hat{\Sigma}_{\hat{\epsilon}^+ \bar{\mathbf{z}} | \mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^+}^{-1} (\mathbf{z}_{[t_0, t+1]})_N$$

we obtain the sample conditional covariance  $\hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}}$  as reported in (60). Similar calculations hold for the other sample covariances.  $\blacksquare$

*Proof of Proposition 8.3.* We just need to recall that, for matrices of suitable dimensions  $\text{vec}(ABC) = (C^\top \otimes A) \text{vec}(B)$ ; therefore from (60) we get,

$$\text{vec} \left( \hat{\Sigma}_{\Delta \mathbf{v}_x \hat{\xi} | \mathbf{g}} \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \right) = \left[ \left[ \left( \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \hat{\Sigma}_{\mathbf{z} \hat{\xi} | \mathbf{g}}^\top \right) \otimes \hat{M}_{x1}^{ql} \right] \left[ \left( \Sigma_{\hat{\xi} \hat{\xi} | \mathbf{g}}^{-1} \hat{\Sigma}_{\mathbf{z} \hat{\xi} | \mathbf{g}}^\top \right) \otimes \hat{M}_{x1}^{ql} \right] \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}^+ \mathbf{z} | \mathbf{g}^+} \hat{\Sigma}_{\mathbf{z} \mathbf{z} | \mathbf{g}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}^+ \bar{\mathbf{z}} | \mathbf{g}^+} \hat{\Sigma}_{\bar{\mathbf{z}} \bar{\mathbf{z}} | \mathbf{g}^+}^{-1} \right) \end{bmatrix} \right].$$

From standard results (see [19]) the asymptotic properties of the right hand side do not change if we substitute the sample value of  $\left[ \left( \Sigma_{\hat{\xi}\hat{\xi}|\mathbf{g}}^{-1} \hat{\Sigma}_{\mathbf{z}\hat{\xi}|\mathbf{g}}^\top \right) \otimes \hat{M}_{x1}^{gl} \right]$  with its asymptotic value  $M_{A1}^{gl} := \left[ \left( \Sigma_{\xi\xi|\mathbf{g}}^{-1} \Sigma_{\mathbf{z}\xi|\mathbf{g}}^\top \right) \otimes M_{x1}^{gl} \right]$ . Similar considerations hold also for the other formulas in (60). ■

*Proof of Proposition 8.5.* We shall prove only (72), the other expressions follow completely analogous.

Recall that

$$\text{vec} \left( \hat{\Sigma}_{\mathbf{e}_{t+k}\mathbf{z}|\mathbf{g}^k} \Sigma_{\mathbf{z}\mathbf{z}|\mathbf{g}^k}^{-1} \right) = \left( \Sigma_{\mathbf{z}\mathbf{z}|\mathbf{g}^k}^{-1} \otimes I \right) \text{vec} \left( \hat{\Sigma}_{\mathbf{e}_{t+k}\mathbf{z}|\mathbf{g}^k} \right)$$

and

$$\text{vec} \left( \hat{\Sigma}_{\mathbf{e}_{t+k}\mathbf{z}|\mathbf{g}^k} \right) = \text{vec} \left( \hat{E} \left[ \mathbf{e}_N(t+k) \left( \mathbf{z}_{[t_0,t]}^{gk} \right)_N^\top \right] \right) = \hat{E} \left[ \left( \mathbf{z}_{[t_0,t]}^{gk} \right)_N \otimes \mathbf{e}_N(t+k) \right].$$

Recalling that, given column vectors  $a, b, c, d$ , it holds that  $\text{vec}(a \cdot b^\top) \text{vec}^\top(c \cdot d^\top) = (b \otimes a) \cdot (d^\top \otimes c^\top) = (b \cdot d^\top) \otimes (a \cdot c^\top)$  and using the assumption A1 (see formula (58)), some standard manipulations (see [41]) yield ( $\tau = h - k$ )

$$\lim_{N \rightarrow \infty} NE \left\{ \hat{E} \left[ \left( \mathbf{z}_{[t_0,t]}^{gk} \right)_N \otimes \mathbf{e}_N(t+k) \right] \hat{E} \left[ \left( \mathbf{z}_{[t_0,t]}^{gh} \right)_N \otimes \mathbf{e}_N(t+h) \right]^\top \right\} = \Sigma_{\mathbf{z}^{gk}\mathbf{z}^{gh}}(\tau) \otimes \Lambda_{\mathbf{e}}. \quad \blacksquare$$

## APPENDIX B

In this appendix we derive the asymptotic variance expressions for the “whitening filter” approach. The proof follows very closely the one for the “innovation estimation” algorithm and therefore only the main parts shall be reported without proof, which can be easily adapted from the parallel results established for the “innovation estimation” algorithm.

A decomposition similar to that of Lemma 5.4 holds if, instead of (46) and (47), we use (48) and (49).

*Lemma B.1:* Let us define,

$$\hat{H}_{\mathbf{w}}(t, T-1) := \begin{bmatrix} 0 & 0 & \dots & 0 \\ C\hat{\Psi}_t & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\Phi(t, T-2)\hat{\Psi}_t & \dots & C\hat{\Psi}_{T-2} & 0 \end{bmatrix}$$

$$\bar{H}_s(t, T-1) := \begin{bmatrix} I & 0 & \dots & 0 \\ -C\hat{\Theta}_t & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -C\Phi(t, T-2)\hat{\Theta}_t & \dots & -C\hat{\Theta}_{T-2} & I \end{bmatrix}$$

and

$$\bar{H}_y(t, T-1) := \begin{bmatrix} 0 & 0 & \dots & 0 \\ CK & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\Phi(t, T-2)K & \dots & CK & 0 \end{bmatrix} \quad \bar{H}_u(t, T-1) := \begin{bmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\Phi(t, T-2)B & \dots & CB & 0 \end{bmatrix}$$

Let us also define

$$\begin{aligned} \bar{\mathbf{b}}_N(t) &:= \hat{H}_w(t, T-1)\hat{\mathbf{w}}_N^+ \\ \bar{\mathbf{b}}_N(t+1) &:= \hat{H}_w(t+1, T)\hat{\mathbf{w}}_N^+ \end{aligned} \quad (\text{B.81})$$

and

$$\begin{aligned} \bar{\mathbf{v}}_N(t) &:= \bar{H}_s(t, T-1)\hat{\epsilon}_N^+ \\ \bar{\mathbf{v}}_N(t+1) &:= \bar{H}_s(t+1, T)\hat{\epsilon}_N^+ \end{aligned} \quad (\text{B.82})$$

Then the future output sequences  $\mathbf{y}_N^+ := [\mathbf{y}_N(t)^\top, \mathbf{y}_N(t+1)^\top, \dots, \mathbf{y}_N(T-1)^\top]^\top$  can be decomposed as follows:

$$\begin{aligned} \mathbf{y}_N^+ &= \bar{\Gamma}(t, T-1)\hat{\mathbf{x}}_N(t) - \bar{\mathbf{b}}_N(t) + \bar{\mathbf{v}}_N(t) + \\ &\quad + \bar{H}_u(t, T-1)\mathbf{u}_N^+ + \bar{H}_y(t, T-1)\mathbf{y}_N^+ + \hat{\mathbf{e}}_N^+ \end{aligned} \quad (\text{B.83})$$

and similarly

$$\begin{aligned} \underline{\mathbf{y}}_N^+ &= \bar{\Gamma}(t+1, T)\hat{\mathbf{x}}_N(t+1) - \bar{\mathbf{b}}_N(t+1) + \bar{\mathbf{v}}_N(t+1) + \\ &\quad + \bar{H}_u(t+1, T)\underline{\mathbf{u}}_N^+ + \bar{H}_y(t+1, T)\underline{\mathbf{y}}_N^+ + \hat{\mathbf{e}}_N^+ \end{aligned} \quad (\text{B.84})$$

Note that the dependence from the terms in  $Z_{[t, T]}$  and  $Z_{[t+1, T]}$  has a lower triangular structure.

The state sequences  $\hat{\zeta}_N(t)$  and  $\hat{\zeta}_N(t+1)$  are constructed according to (32). From equations (B.83) and (B.84) we can see how the error terms  $\bar{\mathbf{b}}_N$  and  $\bar{\mathbf{v}}_N$  affect the state estimators. Defining  $\hat{T}_N(t) := \hat{\Gamma}_N^{-L}\bar{\Gamma}(t, T-1)$  and  $\hat{T}_N(t+1) := \hat{\Gamma}_N^{-L}\bar{\Gamma}(t+1, T)$  a simple manipulation shows in fact that

$$\hat{\zeta}_N(t) = \hat{T}_N(t)\hat{\mathbf{x}}_N(t) + \Delta\bar{\mathbf{x}}_N(t) \quad (\text{B.85})$$

and

$$\hat{\zeta}_N(t+1) = \hat{T}_N(t+1)\hat{\mathbf{x}}_N(t+1) + \Delta\bar{\mathbf{x}}_N(t+1) \quad (\text{B.86})$$

where<sup>18</sup>

$$\Delta \bar{\mathbf{x}}_N(t) = -\hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{b}}_N(t) | Z_{[t_0,t]}] + \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{v}}_N(t) | Z_{[t_0,t]}]$$

and

$$\Delta \bar{\mathbf{x}}_N(t+1) = -\hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t+1,T]}} [\bar{\mathbf{b}}_N(t+1) | Z_{[t_0,t+1]}] + \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t+1,T]}} [\bar{\mathbf{v}}_N(t+1) | Z_{[t_0,t+1]}]$$

These decompositions are the starting point for obtaining recursions for  $\hat{\zeta}_N(t)$  involving the system parameters. The result, which we summarize in the next Proposition, follows from (B.85), (B.86) and Proposition 5.3.

*Proposition B.2:* Let  $A_N^{wf} := \hat{T}_N(t+1)A\hat{T}_N^{-1}(t+1)$ ,  $B_N^{wf} := \hat{T}_N(t+1)B$ ,  $C_N^{wf} := C\hat{T}_N^{-1}(t+1)$ ,  $K_N^{wf} := \hat{T}_N(t+1)K$ ,  $\bar{A}_N^{wf} := A_N^{wf} - K_N^{wf}C_N^{wf}$  be the state matrices in the (data dependent) basis  $\hat{T}_N(t+1)$  and

$$\begin{aligned} (\Delta \bar{\mathbf{b}}_x)_N &:= \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t+1,T]}} [\bar{\mathbf{b}}_N(t+1) | Z_{[t_0,t+1]}] + \\ &\quad - A_N^{wf} \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{b}}_N(t) | Z_{[t_0,t]}] - \hat{T}_N(t+1) \hat{\Psi}_t \hat{\mathbf{w}}_N(t) \\ (\Delta \bar{\mathbf{v}}_x)_N &:= \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t+1,T]}} [\bar{\mathbf{v}}_N(t+1) | Z_{[t_0,t+1]}] + \\ &\quad - A_N^{wf} \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{v}}_N(t) | Z_{[t_0,t]}] + K_N^{wf} \hat{e}_N(t) \\ (\Delta \bar{\mathbf{b}}_y)_N &:= -C_N^{wf} \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{b}}_N(t) | Z_{[t_0,t]}] \\ (\Delta \bar{\mathbf{v}}_y)_N &:= \hat{e}_N(t) - C_N^{wf} \hat{\Gamma}_N^{-L} \hat{E}_{\|Z_{[t,T-1]}} [\bar{\mathbf{v}}_N(t) | Z_{[t_0,t]}] \end{aligned}$$

Defining  $K_N^{wf}(t) := K_N^{wf} + \hat{T}_N(t+1)\hat{\Theta}_t$  and  $\Delta \hat{T}_N := \hat{T}_N(t+1)\hat{T}_N^{-1}(t)$ , the estimated state sequences  $\hat{\zeta}_N(t)$  and  $\hat{\zeta}_N(t+1)$  satisfy the following recursion:

$$\left\{ \begin{array}{l} \hat{\zeta}_N(t+1) = A_N^{wf} \hat{\zeta}_N(t) + B_N^{wf} \mathbf{u}_N(t) + K_N^{wf}(t) \hat{\mathbf{e}}_N(t) + \\ \quad - (\Delta \bar{\mathbf{b}}_x)_N + (\Delta \bar{\mathbf{v}}_x)_N + A_N^{wf} (I - \Delta \hat{T}_N) \hat{\zeta}_N(t) \\ \mathbf{y}_N(t) = C_N^{wf} \hat{\zeta}_N(t) + \hat{\mathbf{e}}_N(t) \\ \quad - (\Delta \bar{\mathbf{b}}_y)_N + (\Delta \bar{\mathbf{v}}_y)_N + C_N^{wf} (I - \Delta \hat{T}_N) \hat{\zeta}_N(t) \end{array} \right. \quad (\text{B.87})$$

The errors  $\tilde{A}_N^{wf}$ ,  $\tilde{B}_N^{wf}$ ,  $\tilde{C}_N^{wf}$  follows from the relation above similarly to the ‘‘innovation estimation’’ algorithm presented in the main part of the paper. The main difference regards the presence in the error of a term originating from the different choice of basis  $\hat{T}_N(t)$  and  $\hat{T}_N(t+1)$ .

<sup>18</sup>These expressions rely on the fact that  $\hat{E}_{\|Z_{[t,T]}} [\hat{e}_N(t+k) | Z_{[t_0,t]}] = \hat{E}_{\|Z_{[t,t+k]}} [\hat{e}_N(t+k) | Z_{[t_0,t]}]$  and  $\hat{E}_{\|Z_{[t,T]}} [\hat{\mathbf{w}}_N(t+k-1) | Z_{[t_0,t]}] = \hat{E}_{\|Z_{[t,t+k]}} [\hat{\mathbf{w}}_N(t+k-1) | Z_{[t_0,t]}]$ . The first is proved in Lemma B.6, while the second is left to the reader.

*Proposition B.3:* Let  $(\Delta\bar{\mathbf{x}})_N := -(\Delta\bar{\mathbf{b}}_x)_N + (\Delta\bar{\mathbf{v}}_x)_N + A_N^{wf}(I - \Delta\hat{T}_N)\hat{\zeta}_N(t)$  and  $(\Delta\bar{\mathbf{y}})_N := -(\Delta\bar{\mathbf{b}}_y)_N + (\Delta\bar{\mathbf{v}}_y)_N + C_N^{wf}(I - \Delta\hat{T}_N)\hat{\zeta}_N(t)$ . The error matrices  $\tilde{A}_N^{wf}$ ,  $\tilde{B}_N^{wf}$ ,  $\tilde{C}_N^{wf}$  obtained solving (33) (or equivalently (B.87)) satisfy

$$\begin{aligned}\tilde{A}_N^{wf} &= \hat{\Sigma}_{\Delta\bar{\mathbf{x}}\hat{\zeta}|\mathbf{g}}\hat{\Sigma}_{\hat{\zeta}\hat{\zeta}|\mathbf{g}}^{-1} \\ \tilde{B}_N^{wf} &= \hat{\Sigma}_{\Delta\bar{\mathbf{x}}\mathbf{u}|(\hat{\zeta},\hat{\mathbf{e}})}\hat{\Sigma}_{\mathbf{u}\mathbf{u}|(\hat{\zeta},\hat{\mathbf{e}})}^{-1} \\ \tilde{C}_N^{wf} &= \hat{\Sigma}_{\Delta\bar{\mathbf{y}}\hat{\zeta}}\hat{\Sigma}_{\hat{\zeta}\hat{\zeta}}^{-1}\end{aligned}\tag{B.88}$$

*Proof:* The proof follows the same lines as that of Proposition 6.2 and will be omitted.  $\blacksquare$

Unfortunately the analysis of the “whitening filter” based method is complicated by the fact that the estimated states at time  $t$  and time  $t + 1$  correspond in general to different choice of basis (see equations (B.85) and (B.86)) which is reflected in (B.87) by the terms containing  $(I - \Delta\hat{T}_N)$ . This difference unfortunately does not even disappear as  $N \rightarrow \infty$  and therefore contributes to the “bias” term. Under the simplifying assumption that  $\bar{A}$  is nilpotent then  $\hat{T}_N(t) = \hat{T}_N(t + 1)$ . The same equality holds asymptotically if  $t - t_0$  increases at a suitable rate with  $N$ .

**Remark B.8** To the author’s experience both the bias and the fluctuations introduced by the factor  $\Delta\hat{T}_N$  seem to have only marginal effect. A somewhat qualitative explanation of this statement could be as follows. The matrices<sup>19</sup>  $C\Phi(t, t + k)$  and  $C\Phi(t + 1, t + 1 + k)$  which constitute the  $k$ -th block row of respectively  $\bar{\Gamma}(t, T - 1)$  and  $\bar{\Gamma}(t + 1, T)$  contain in corresponding positions time varying matrices  $K(s)$  and  $K(s + 1)$  which satisfy Riccati-type recursions. If the Riccati recursion is in steady state, i.e.  $K(s) = K(s + 1) \forall s > t - t_0$  then  $\bar{\Gamma}(t, T - 1) = \bar{\Gamma}(t + 1, T)$ . In the non steady-state conditions it may still be reasonable to assume that one iteration of the Riccati equation does not change things too much. Of course this is just a qualitative statement, but somehow explains the experimental results in [17] where satisfying results are obtained also when the eigenvalues of  $A - KC$  are close to the unit circle provided those of  $F - LH$  are nearly zero (see simulation Example 4 in [17]).  $\diamond$

*Proposition B.4:* Let  $\Delta T := \Delta\hat{T}_\infty$  and  $\Delta\tilde{T}_N := \Delta T - \Delta\hat{T}_N$ . Then the error matrices  $\tilde{A}_N^{wf}$ ,  $\tilde{B}_N^{wf}$ ,  $\tilde{C}_N^{wf}$  admit the decompositions  $\tilde{A}_N^{wf} = \tilde{A}_\infty^{wf} + \Delta\tilde{A}_N^{wf}$ ,  $\tilde{B}_N^{wf} = \tilde{B}_\infty^{wf} + \Delta\tilde{B}_N^{wf}$ ,  $\tilde{C}_N^{wf} = \tilde{C}_\infty^{wf} + \Delta\tilde{C}_N^{wf}$  where

$$\begin{aligned}\tilde{A}_\infty^{wf} &= -\Sigma_{\Delta\bar{\mathbf{b}}_x\hat{\zeta}|\mathbf{g}}\Sigma_{\hat{\zeta}\hat{\zeta}|\mathbf{g}}^{-1} + A_\infty^{wf}(I - \Delta T) \\ \tilde{B}_\infty^{wf} &= -\Sigma_{\Delta\bar{\mathbf{b}}_x\mathbf{u}|(\hat{\zeta},\hat{\mathbf{e}})}\Sigma_{\mathbf{u}\mathbf{u}|(\hat{\zeta},\hat{\mathbf{e}})}^{-1} \\ \tilde{C}_\infty^{wf} &= -\Sigma_{\Delta\bar{\mathbf{b}}_y\hat{\zeta}}\Sigma_{\hat{\zeta}\hat{\zeta}}^{-1} + C_\infty^{wf}(I - \Delta T)\end{aligned}$$

<sup>19</sup>Recall that  $\Phi(t, s) := \prod_{i=0}^{s-1-t} (A - K(s - 1 - i)C)$ .

and

$$\begin{aligned}
\Delta \tilde{A}_N^{wf} &\doteq \tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_x \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} + \tilde{A}_\infty^{wf} \tilde{\Sigma}_{\hat{\zeta} \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} + \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} + \\
&\quad + A_\infty^{wf} \Delta \tilde{T}_N + (A_N^{wf} - A_\infty^{wf})(I - \Delta T) \\
\Delta \tilde{B}_N^{wf} &\doteq \tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_x \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1} + \tilde{B}_\infty^{wf} \tilde{\Sigma}_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1} + \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1} \\
\Delta \tilde{C}_N^{wf} &\doteq \tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_y \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} + \tilde{C}_\infty^{wf} \tilde{\Sigma}_{\hat{\zeta} \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} + \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_y \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} + \\
&\quad + C_\infty^{wf} \Delta \tilde{T}_N + (C_N^{wf} - C_\infty^{wf})(I - \Delta T)
\end{aligned}$$

where  $\doteq$  denotes equality up to terms which are  $o(\frac{1}{\sqrt{N}})$  in probability<sup>20</sup>.

Note that in this case the following parts of the error

$$\begin{aligned}
&\sqrt{N}(\tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_x \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} + \tilde{A}_\infty^{wf} \tilde{\Sigma}_{\hat{\zeta} \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} + A_\infty^{wf} \Delta \tilde{T}_N + (A_N^{wf} - A_\infty^{wf})(I - \Delta T)) \\
&\quad \sqrt{N}(\tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_x \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1} + \tilde{B}_\infty^{wf} \tilde{\Sigma}_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1}) \\
&\quad \sqrt{N}(\tilde{\Sigma}_{\Delta \bar{\mathbf{b}}_y \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} + \tilde{C}_\infty^{wf} \tilde{\Sigma}_{\hat{\zeta} \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} + C_\infty^{wf} \Delta \tilde{T}_N + (C_N^{wf} - C_\infty^{wf})(I - \Delta T))
\end{aligned}$$

go to zero in probability when  $\bar{A}^{t-t_0} \rightarrow 0$  at a suitable rate (see [8]) with  $N$ . Under this condition only the terms related to  $\Delta \bar{\mathbf{v}}_x$  and  $\Delta \bar{\mathbf{v}}_y$  have to be accounted for variance. Both this approach and the computation of the variance in the general case (for  $t - t_0$  finite) would introduce further technical complications.

For reasons of space and ease of exposition we prefer at this point to make the assumption that  $\bar{A}^{t-t_0} = 0$  holds. This, as we have shown in [17], implies that  $\tilde{A}_\infty^{wf} = 0, \tilde{B}_\infty^{wf} = 0, \tilde{C}_\infty^{wf} = 0$  (and also  $(\Delta \bar{\mathbf{b}}_x)_N = 0, (\Delta \bar{\mathbf{b}}_y)_N = 0, \Delta \tilde{T}_N = \Delta T = I$ ); see Proposition B.2, Lemma 5.2 and equation (B.81).

Recall now that under the assumption that  $\bar{A}^{t-t_0} = 0$

$$\begin{aligned}
\tilde{A}_N^{wf} = \Delta \tilde{A}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} \\
\tilde{B}_N^{wf} = \Delta \tilde{B}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{u} \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}^{-1} \\
\tilde{C}_N^{wf} = \Delta \tilde{C}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_y \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1}
\end{aligned}$$

we just need to study the asymptotic properties of  $\hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \hat{\zeta} | \mathbf{g}}, \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \mathbf{u} | (\hat{\zeta}, \hat{\mathbf{e}})}$  and  $\hat{\Sigma}_{\Delta \bar{\mathbf{v}}_y \hat{\zeta}}$ .

*Proposition B.5:* Let introduce the shorthands  $\mathbf{z}^+ := \mathbf{z}_{[t, T-1]}$ ,  $\mathbf{z}^- := \mathbf{z}_{[t+1, T]}$ ,  $\mathbf{z} := \mathbf{z}_{[t_0, t]}$ ,  $\bar{\mathbf{z}} := \mathbf{z}_{[t_0, t+1]}$ . Let  $\hat{M}_{x1}^{wf} := \left( K_N^{wf} [I \ 0] - A_N^{wf} \hat{\Gamma}_N^{-L} \bar{H}_s(t, T-1) \right)$ ,  $\hat{M}_{x2}^{wf} := \hat{\Gamma}_N^{-L} \bar{H}_s(t+1, T)$  and  $\hat{M}_y^{wf} := \left( [I \ 0] - C_N^{wf} \hat{\Gamma}_N^{-L} \bar{H}_s(t, T-1) \right)$ .

Let as usual denote the asymptotic values (in probability as  $N \rightarrow \infty$ ) of matrices removing the ‘‘hat’’ from the symbol denoting the sample version<sup>21</sup> (e.g.  $M_{x2}^{wf} := \lim_{N \rightarrow \infty} \hat{M}_{x2}^{wf}$ ) and let

<sup>20</sup>We remind the reader that  $A_\infty^{wf}, B_\infty^{wf}, C_\infty^{wf}$  denote the asymptotic values of  $A_N^{wf}, B_N^{wf}, C_N^{wf}$  which correspond to the asymptotic choice of basis in (32) for  $N \rightarrow \infty$ .

<sup>21</sup>We refer the reader to Remark VIII.7 concerning the issue of convergence.

$$\begin{aligned}
M_{A1}^{wf} &:= \left[ \left( \Sigma_{\hat{\zeta}\hat{\zeta}|\mathbf{g}}^{-1} \Sigma_{\bar{\mathbf{z}}\hat{\zeta}|\mathbf{g}}^\top \right) \otimes M_{x1}^{wf} \right] \\
M_{A2}^{wf} &:= \left[ \left( \Sigma_{\hat{\zeta}\hat{\zeta}|\mathbf{g}}^{-1} \Sigma_{\mathbf{z}\hat{\zeta}|\mathbf{g}}^\top \right) \otimes M_{x2}^{wf} \right] \\
M_{B1}^{wf} &:= \left[ \left( \Sigma_{\mathbf{uu}(\hat{\zeta}, \hat{\mathbf{e}})}^{-1} \Sigma_{\mathbf{zu}(\hat{\zeta}, \hat{\mathbf{e}})}^\top \right) \otimes M_{x1}^{wf} \right] \\
M_{B2}^{wf} &:= \left[ \left( \Sigma_{\mathbf{uu}(\hat{\zeta}, \hat{\mathbf{e}})}^{-1} \Sigma_{\bar{\mathbf{z}}\mathbf{u}(\hat{\zeta}, \hat{\mathbf{e}})}^\top \right) \otimes M_{x2}^{wf} \right] \\
M_C^{wf} &:= \left[ \left( \Sigma_{\hat{\zeta}\hat{\zeta}}^{-1} \Sigma_{\mathbf{z}\hat{\zeta}}^\top \right) \otimes M_y^{wf} \right]
\end{aligned} \tag{B.89}$$

Then the following first order approximation (up to  $o(\frac{1}{\sqrt{N}})$  terms in probability) of vectorized sample covariances can be used to compute the asymptotic distribution:

$$\begin{aligned}
\text{vec} \left( \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \hat{\zeta}|\mathbf{g}} \Sigma_{\hat{\zeta}\hat{\zeta}|\mathbf{z}}^{-1} \right) &\doteq \begin{bmatrix} M_{A1}^{wf} & M_{A2}^{wf} \end{bmatrix} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{z}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^+}^{-1} \right) \end{bmatrix} \\
\text{vec} \left( \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \mathbf{u}(\hat{\zeta}, \hat{\mathbf{e}})} \Sigma_{\mathbf{uu}(\hat{\zeta}, \hat{\mathbf{e}})}^{-1} \right) &\doteq \begin{bmatrix} M_{B1}^{wf} & M_{B2}^{wf} \end{bmatrix} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{z}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^+}^{-1} \right) \end{bmatrix} \\
\text{vec} \left( \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_y \hat{\zeta}} \Sigma_{\hat{\zeta}\hat{\zeta}}^{-1} \right) &\doteq M_C^{wf} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} \right)
\end{aligned} \tag{B.90}$$

*Proof:* Since the proof of this proposition follows the same lines of the proof of Lemma 8.2 and Proposition 8.3 it will be omitted. ■

The last effort left is to compute the asymptotic covariance of

$$\begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\mathbf{e}}+\bar{\mathbf{z}}|\mathbf{z}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^+}^{-1} \right) \end{bmatrix} \tag{B.91}$$

Using the same permutation matrices  $P$  and  $\bar{P}$  defined in (64) it is possible to write (B.91) in terms of  $\text{vec} \left\{ \hat{\Sigma}_{\hat{\mathbf{e}}_{t+i}\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} \right\}$ ,  $i = 0, \dots, T-t-1$  and  $\text{vec} \left\{ \hat{\Sigma}_{\hat{\mathbf{e}}_{t+i}\bar{\mathbf{z}}|\mathbf{z}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^+}^{-1} \right\}$  for  $i = 1, \dots, T-t$ .

An instrumental result is the following:

*Lemma B.6:* The oblique projection  $\hat{E}_{\|Z_{[s,T]}} [\hat{\mathbf{e}}_N(s+k) | Z_{[t_0,s]}]$  satisfies the following:

$$\begin{aligned}
\hat{E}_{\|Z_{[s,T]}} [\hat{\mathbf{e}}_N(s+k) | Z_{[t_0,s]}] &= \hat{E}_{\|Z_{[s,s+k-1]}} [\hat{\mathbf{e}}_N(s+k) | Z_{[t_0,s]}] = \\
&= \hat{E}_{\|Z_{[s,s+k-1]}} [\mathbf{e}_N(s+k) | Z_{[t_0,s]}].
\end{aligned} \tag{B.92}$$

Let define the shorthands  $\mathbf{z}^k := \mathbf{z}_{[t,t+k]}$  and  $\bar{\mathbf{z}}^k := \mathbf{z}_{[t+1,t+k]}$ .

Then (B.92) is equivalent to the equation for the coefficients

$$\hat{\Sigma}_{\hat{\mathbf{e}}_{t+k}\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^+}^{-1} = \hat{\Sigma}_{\hat{\mathbf{e}}_{t+k}\mathbf{z}|\mathbf{z}^k} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^k}^{-1} = \hat{\Sigma}_{\mathbf{e}_{t+k}\mathbf{z}|\mathbf{z}^k} \hat{\Sigma}_{\mathbf{zz}|\mathbf{z}^k}^{-1} \tag{B.93}$$

and

$$\hat{\Sigma}_{\hat{\mathbf{e}}_{t+k}\bar{\mathbf{z}}|\mathbf{z}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^+}^{-1} = \hat{\Sigma}_{\hat{\mathbf{e}}_{t+k}\bar{\mathbf{z}}|\mathbf{z}^k} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^k}^{-1} = \hat{\Sigma}_{\mathbf{e}_{t+k}\bar{\mathbf{z}}|\mathbf{z}^k} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\mathbf{z}^k}^{-1} \tag{B.94}$$

*Proof:* The proof follows easily from the fact that  $\hat{\epsilon}_N(t+k) = \hat{E}[\mathbf{e}_N(t+k) \mid Z_{[t_0, t+k]}]$  ■

It follows from (B.93) and (B.94) and the discussion after formula (B.91) that the asymptotic variance we are seeking can be expressed in terms of variances and covariances of terms of the form

$$\sqrt{N}\hat{R}_k^{wf} := \sqrt{N}\hat{\Sigma}_{\mathbf{e}_{t+k}|\mathbf{z}^k}\Sigma_{\mathbf{z}\mathbf{z}^k}^{-1} \doteq \sqrt{N}\hat{\Sigma}_{\mathbf{e}_{t+k}|\mathbf{z}^k}\hat{\Sigma}_{\mathbf{z}\mathbf{z}^k}^{-1} \quad k = 0, \dots, T-t-1$$

and

$$\sqrt{N}\hat{R}_k^{wf} := \sqrt{N}\hat{\Sigma}_{\mathbf{e}_{t+k}|\bar{\mathbf{z}}^k}\Sigma_{\bar{\mathbf{z}}\bar{\mathbf{z}}^k}^{-1} \doteq \sqrt{N}\hat{\Sigma}_{\mathbf{e}_{t+k}|\bar{\mathbf{z}}^k}\hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}^k}^{-1} \quad k = 1, \dots, T-t$$

which we compute in the next proposition.

*Proposition B.7:* Let us assume that the assumptions A1 and A2 hold. Let us define <sup>22</sup>

$$\begin{aligned} \mathbf{z}_{[t_0+\tau, t+\tau]}^{zk} &:= E \left[ \mathbf{z}_{[t_0+\tau, t+\tau]} \mid \mathcal{Z}_{[t+\tau, t+\tau+k]}^\perp \right], \\ \bar{\mathbf{z}}_{[t_0+\tau, t+1+\tau]}^{zk} &:= E \left[ \mathbf{z}_{[t_0+\tau, t+1+\tau]} \mid \mathcal{Z}_{[t+1+\tau, t+\tau+k]}^\perp \right], \\ \Sigma_{\mathbf{z}^k\mathbf{z}^{zh}}(\tau) &:= E \left[ \mathbf{z}_{[t_0+\tau, t+\tau]}^{zk} \left( \mathbf{z}_{[t_0, t]}^{zh} \right)^\top \right]. \end{aligned}$$

Similar definitions hold for  $\Sigma_{\bar{\mathbf{z}}^k\bar{\mathbf{z}}^{zh}}(\tau)$  and  $\Sigma_{\mathbf{z}^k\bar{\mathbf{z}}^{zh}}(\tau)$ . Let also<sup>23</sup>  $Q_k^{wf} := \left( \Sigma_{\mathbf{z}\mathbf{z}^k}^{-1} \otimes I \right)$ ,  $\bar{Q}_k^{wf} := \left( \Sigma_{\bar{\mathbf{z}}\bar{\mathbf{z}}^k}^{-1} \otimes I \right)$  and  $\tau := h - k$ , then:

$$\Sigma_{11}^{wf}(k, h) := \text{AsCov} \left\{ \sqrt{N}\text{vec} \left( \hat{R}_k^{wf} \right), \sqrt{N}\text{vec} \left( \hat{R}_h^{wf} \right) \right\} = Q_k^{wf} \left( \Sigma_{\mathbf{z}^k\mathbf{z}^{zh}}(\tau) \otimes \Lambda_e \right) \left( Q_h^{wf} \right)^\top \quad (\text{B.95})$$

Similarly

$$\Sigma_{22}^{wf}(k, h) := \text{AsCov} \left\{ \sqrt{N}\text{vec} \left( \hat{R}_k^{wf} \right), \sqrt{N}\text{vec} \left( \hat{R}_h^{wf} \right) \right\} = \bar{Q}_k^{wf} \left( \Sigma_{\bar{\mathbf{z}}^k\bar{\mathbf{z}}^{zh}}(\tau) \otimes \Lambda_e \right) \left( \bar{Q}_h^{wf} \right)^\top \quad (\text{B.96})$$

As far as the cross terms are concerned, we have

$$\Sigma_{12}^{wf}(k, h) := \text{AsCov} \left\{ \sqrt{N}\text{vec} \left( \hat{R}_k^{wf} \right), \sqrt{N}\text{vec} \left( \hat{R}_h^{wf} \right) \right\} = Q_k^{wf} \left( \Sigma_{\mathbf{z}^k\bar{\mathbf{z}}^{zh}}(\tau) \otimes \Lambda_e \right) \left( \bar{Q}_h^{wf} \right)^\top \quad (\text{B.97})$$

*Proof:* It is essentially analogous to the proof of Proposition 8.5 and for reasons of space will be omitted. ■

Defining

$$\Sigma_{11}^{wf} := \begin{bmatrix} \Sigma_{11}^{wf}(0, 0) & \dots & \Sigma_{11}^{wf}(0, T-t-1) \\ \vdots & \ddots & \vdots \\ \Sigma_{11}^{wf}(T-t-1, 0) & \dots & \Sigma_{11}^{wf}(T-t-1, T-t-1) \end{bmatrix}, \quad (\text{B.98})$$

<sup>22</sup>The superscript  $^{zk}$  is meant to remind that the process  $\mathbf{z}$  is taken conditionally on the  $k$  lags of the joint future.

<sup>23</sup>Note that  $\Sigma_{\mathbf{z}\mathbf{z}^k} = \Sigma_{\mathbf{z}^k\mathbf{z}^k}(0)$ .

$$\Sigma_{22}^{wf} := \begin{bmatrix} \Sigma_{22}^{wf}(1, 1) & \dots & \Sigma_{22}^{wf}(1, T-t) \\ \vdots & \ddots & \vdots \\ \Sigma_{22}^{wf}(T-t, 1) & \dots & \Sigma_{22}^{wf}(T-t, T-t) \end{bmatrix}, \quad (\text{B.99})$$

and

$$\Sigma_{12}^{wf} := \begin{bmatrix} \Sigma_{12}^{wf}(0, 1) & \dots & \Sigma_{12}^{wf}(0, T-t) \\ \vdots & \ddots & \vdots \\ \Sigma_{12}^{wf}(T-t-1, 1) & \dots & \Sigma_{12}^{wf}(T-t-1, T-t) \end{bmatrix}, \quad (\text{B.100})$$

from Proposition B.7 it follows that

$$\text{AsVar} \left\{ \sqrt{N} \begin{bmatrix} \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}+\mathbf{z}|\mathbf{z}^+} \hat{\Sigma}_{\mathbf{z}\mathbf{z}|\mathbf{z}^+}^{-1} \right) \\ \text{vec} \left( \hat{\Sigma}_{\hat{\epsilon}+\bar{\mathbf{z}}|\bar{\mathbf{z}}^+} \hat{\Sigma}_{\bar{\mathbf{z}}\bar{\mathbf{z}}|\bar{\mathbf{z}}^+}^{-1} \right) \end{bmatrix} \right\} = \begin{pmatrix} P & 0 \\ 0 & \bar{P} \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{wf} & \Sigma_{12}^{wf} \\ (\Sigma_{12}^{wf})^\top & \Sigma_{22}^{wf} \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & \bar{P} \end{pmatrix}^\top \quad (\text{B.101})$$

Collecting the results obtained so far we have basically proven the main result of the paper.

*Proof of Theorem 8.1. [Whitening filter approach]*

For convenience we report, before starting the proof, the asymptotic variance formulas for  $\sqrt{N} \text{vec} \left( \tilde{A}_N^{wf} \right)$ ,  $\sqrt{N} \text{vec} \left( \tilde{B}_N^{wf} \right)$  and  $\sqrt{N} \text{vec} \left( \tilde{C}_N^{wf} \right)$ . The formulas are the same as those reported in Theorem 8.1 for the ‘‘innovation estimation’’ algorithm, provided we substitute the superscript  $^{ql}$  with the superscript  $^{wf}$ .

Letting

$$M^{wf} := \begin{pmatrix} M_{A1}^{wf} P & M_{A2}^{wf} \bar{P} \\ M_{B1}^{wf} P & M_{B2}^{wf} \bar{P} \\ M_C^{wf} P & 0 \end{pmatrix}$$

then

$$\text{AsVar} \left\{ \sqrt{N} \begin{bmatrix} \text{vec} \left( \tilde{A}_N^{wf} \right) \\ \text{vec} \left( \tilde{B}_N^{wf} \right) \\ \text{vec} \left( \tilde{C}_N^{wf} \right) \end{bmatrix} \right\} = M^{wf} \begin{pmatrix} \Sigma_{11}^{wf} & \Sigma_{12}^{wf} \\ (\Sigma_{12}^{wf})^\top & \Sigma_{22}^{wf} \end{pmatrix} (M^{wf})^\top \quad (\text{B.102})$$

Recalling now that under the assumption that  $\bar{A}^{t-t_0} = 0$

$$\begin{aligned} \tilde{A}_N^{wf} = \Delta \tilde{A}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \hat{\zeta} | \mathbf{g}} \Sigma_{\hat{\zeta} \hat{\zeta} | \mathbf{g}}^{-1} \\ \tilde{B}_N^{wf} = \Delta \tilde{B}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_x \mathbf{u} | (\hat{\zeta}, \hat{\epsilon})} \Sigma_{\mathbf{u}\mathbf{u} | (\hat{\zeta}, \hat{\epsilon})}^{-1} \\ \tilde{C}_N^{wf} = \Delta \tilde{C}_N^{wf} &\doteq \hat{\Sigma}_{\Delta \bar{\mathbf{v}}_y \hat{\zeta}} \Sigma_{\hat{\zeta} \hat{\zeta}}^{-1} \end{aligned}$$

asymptotic normality follows again from standard results on asymptotic properties of sample covariances [41][p. 550]. Using Propositions B.5, B.7 and equation (B.101) the asymptotic variance formula follows.



**Remark B.9** It is worth noting that under the stated assumption that  $\bar{A}^{t-t_0} = 0$  also  $\hat{\Theta}_s = 0 \forall s \geq t$ . This implies that  $\bar{H}_s(t, T-1) = \bar{H}_s(t+1, T) = I$  which simplifies significantly the variance expressions. The reason for this simplification is that, in the state recursion (48) the term due to  $\hat{\Theta}_t$  disappears. The same does not happen for the “innovation estimation” algorithm. We believe this may be the starting point for further analysis of these formulas. For reasons of space we postpone this discussion to future work.  $\diamond$

#### ACKNOWLEDGMENT

The author would like to thank Giorgio Picci for many discussions on the subject and for his continuing support and encouragement over these years and the colleagues in Padova for suggestions on how to improve the manuscript. The author is indebted with the editor (L. Ljung) of the special issue to which the paper was originally submitted and with the reviewers for their careful reading and many insightful comments.

#### REFERENCES

- [1] D. Bauer, “Some asymptotic theory for the estimation of linear systems using maximum likelihood methods or subspace algorithms,” Ph.D. dissertation, TU Wien, Austria, 1998.
- [2] —, “Comparing the cca subspace method to pseudo maximum likelihood methods in the case of no exogenous inputs,” *Journal of Time Series Analysis (submitted)*, 2002.
- [3] —, “Subspace algorithms,” 2003, slides of the Semiplenary Lecture given at SYSID, available at [www.sysid2003.nl/plenaries/sp4.pdf](http://www.sysid2003.nl/plenaries/sp4.pdf).
- [4] —, “Subspace methods,” in *Proc. of IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 2003, pp. 763–773.
- [5] —, “Asymptotic properties of subspace estimators,” *Automatica*, vol. 41, pp. 359–376, 2005.
- [6] D. Bauer, M. Deistler, and W. Scherrer, “On the impact of weighting matrices in subspace algorithms,” in *Proceedings of IFAC International Symposium on System Identification*, Santa Barbara, June 2000, p. N.A.
- [7] D. Bauer and M. Jansson, “Analysis of the asymptotic properties of the moesp type of subspace algorithms,” *Automatica*, vol. 36, pp. 497–509, 2000.
- [8] D. Bauer and L. Ljung, “Some facts about the choice of the weighting matrices in larimore type of subspace algorithm,” *Automatica*, vol. 36, pp. 763–773, 2001.
- [9] P. Caines and C. Chan, “Estimation, identification and feedback,” in *System Identification: Advances and Case Studies*, R. Mehra and D. Lainiotis, Eds. Academic, 1976, pp. 349–405.
- [10] A. Chiuso, “On the relation between cca and predictor based subspace identification,” in *submitted to CDC-ECC 2005*, 2005, available at <http://www.dei.unipd.it/~chiuso>.

- [11] A. Chiuso and G. Picci, "Constructing the state of random processes with feedback," in *Proc. of the IFAC Int. Symposium on System Identification (SYSID)*, Rotterdam, August 2003.
- [12] —, *Geometry of Oblique Splitting, Minimality and Hankel Operators*, ser. Lect. Notes in Control and Information Sciences. Springer, 2003, no. 286, pp. 85–124.
- [13] —, "The asymptotic variance of subspace estimates," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 257–291, 2004.
- [14] —, "Asymptotic variance of subspace methods by data orthogonalization and model decoupling: A comparative analysis," *Automatica*, vol. 40, no. 10, pp. 1705–1717, 2004.
- [15] —, "Numerical conditioning and asymptotic variance of subspace estimates," *Automatica*, vol. 40, no. 4, pp. 677–683, 2004.
- [16] —, "On the ill-conditioning of subspace identification with inputs," *Automatica*, vol. 40, no. 4, pp. 575–589, 2004.
- [17] —, "Consistency analysis of some closed-loop subspace identification methods," *Automatica*, vol. 41, no. 3, pp. 377–391, 2005.
- [18] C. T. Chou and M. Verhaegen, "Closed-loop identification using canonical correlation analysis," in *Proceedings of the European Control Conference*, Karlsruhe, 1999, pp. CD-ROM, F162.
- [19] T. Ferguson, *A Course in Large Sample Theory*. Chapman and Hall, 1996.
- [20] U. Forsell and L. Ljung, "Closed loop identification revisited," *Automatica*, vol. 35, pp. 1215–1242, 1999.
- [21] M. Gevers, L. Ljung, and P. V. den Hof, "Asymptotic variance expressions for closed-loop identification," *Automatica*, vol. 37, no. 5, pp. 781–786, 2001.
- [22] M. Gevers and B. Anderson, "Representation of jointly stationary feedback free processes," *Intern. Journal of Control*, vol. 33, pp. 777–809, 1981.
- [23] —, "On jointly stationary feedback-free stochastic processes," *IEEE Trans. Automat. Contr.*, vol. 27, pp. 431–436, 1982.
- [24] G. Golub and C. V. Loan, *Matrix Computation*, 2nd ed. The Johns Hopkins Univ. Press., 1989.
- [25] C. Granger, "Economic processes involving feedback," *Information and Control*, vol. 6, pp. 28–48, 1963.
- [26] T. Gustafsson, "Subspace-based system identification: weighting and pre-filtering of instruments," *Automatica*, vol. 38, pp. 433–443, 2002.
- [27] E. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. Wiley, 1988.
- [28] E. Hannan and D. Poskitt, "Unit canonical correlations between future and past," *The Annals of Statistics*, vol. 16, pp. 784–790, 1988.
- [29] M. Jansson, "Asymptotic variance analysis of subspace identification methods," in *Proceedings of SYSID2000*, S. Barbara Ca., 2000.
- [30] —, "Subspace identification and arx modelling," in *Proceedings of SYSID 2003*, Rotterdam, 2003.
- [31] M. Jansson and B. Wahlberg, "On consistency of subspace methods for system identification," *Automatica*, vol. 34, pp. 1507–1519, 1998.
- [32] T. Katayama, H. Kawauchi, and G. Picci, "Subspace identification of closed loop systems by orthogonal decomposition," *To appear, Automatica (brief paper)*, 2003.
- [33] W. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control," in *Proc. 29th IEEE Conf. Decision & Control*, Honolulu, 1990, pp. 596–604.
- [34] W. Lin, S. Qin, and L. Ljung, "A framework for closed-loop subspace identification with innovation estimation," Texas-Wisconsin Modeling and Control Consortium, Tech. Rep. 07, 2004, available at [www.che.wisc.edu/jbr-group/tech-reports/twmcc-2004-07.pdf](http://www.che.wisc.edu/jbr-group/tech-reports/twmcc-2004-07.pdf).

- [35] A. Lindquist and G. Picci, "Canonical correlation analysis, approximate covariance extension and identification of stationary time series," *Automatica*, vol. 32, pp. 709–733, 1996.
- [36] L. Ljung, *System Identification; Theory for the User*. Prentice Hall, 1997.
- [37] L. Ljung and T. McKelvey, "Subspace identification from closed loop data," *Signal Processing*, vol. 52, no. 2, pp. 209–216, 1996.
- [38] M. Moonen, B. De Moor, L. Vandeberghe, and J. Vandewalle, "On- and off-line identification of linear state-space models," *Int. J. of Control*, vol. 49, no. 1, pp. 219–232, 1989.
- [39] G. Picci and T. Katayama, "Stochastic realization with exogenous inputs and "subspace methods" identification," *Signal Processing*, vol. 52, pp. 145–160, 1996.
- [40] S. Qin and L. Ljung, "Closed-loop subspace identification with innovation estimation," in *Proceedings of SYSID 2003*, Rotterdam, 2003.
- [41] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall, 1989.
- [42] A. C. Van der Klauw, M. Verhaegen, and P. P. J. Van den Bosch, "State space identification of closed loop systems," in *Proc. 30th IEEE Conf. on Dec. & Control*, Brighton, U.K., 1991, pp. 1327–1332.
- [43] P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, pp. 649–660, 1993.
- [44] —, "N4SID: Subspace algorithms for the identification of combined deterministic– stochastic systems," *Automatica*, vol. 30, pp. 75–93, 1994.
- [45] —, *Subspace Identification for Linear Systems*. Kluwer Academic Publications, 1996.
- [46] —, "Closed loop subspace systems identification," in *Proceedings of 36th IEEE Conference on Decision & Control*, San Diego, CA, 1997, pp. pp. 1848–1853.
- [47] M. Verhaegen, "Application of a subspace model identification technique to identify lti systems operating in closed-loop," *Automatica*, vol. 29, pp. 1027–1040, 1993.
- [48] —, "Identification of the deterministic part of mimo state space models given in innovations form from input-output data," *Automatica*, vol. 30, pp. 61–74, 1994.
- [49] M. Verhaegen and P. Dewilde, "Subspace model identification, part 1. the output-error state-space model identification class of algorithms; part 2. analysis of the elementary output-error state-space model identification algorithm," *Int. J. Control*, vol. 56, pp. 1187–1210 & 1211–1241, 1992.
- [50] —, "Subspace model identification, part 3. analysis of the ordinary output-error state-space model identification algorithm," *Int. J. Control*, vol. 58, pp. 555–586, 1993.
- [51] M. Viberg, "Subspace-based methods for the identification of linear time-invariant systems," *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.
- [52] M. Viberg, B. Wahlberg, and B. Ottersten, "Analysis of state space system identification methods based on instrumental variables and subspace fitting," *Automatica*, vol. 33, no. 9, pp. 1603–1616, 1997.