# On the consistency and asymptotic normality of discrete-time LTI models identified from concatenated data sets

Vasquez Rodriguez, Sandra Paola; Kinnaert, Michel; Pintelon, Rik

*Published in:*
Automatica

*DOI:*
10.1016/j.automatica.2022.110209

*Publication date:*
2022

# On the consistency and asymptotic normality of discrete-time LTI models identified from concatenated data sets ⋆

Sandra Vásquez [a,b], Michel Kinnaert [a], Rik Pintelon [b]

[a] *Université Libre de Bruxelles, Avenue Franklin Roosevelt 50 - 1050 Brussels, Belgium*

[b] *Vrije Universiteit Brussel, Pleinlaan 2 - 1050 Brussels, Belgium*

**Abstract**

Even-though data concatenation is a well-known technique for identifying Linear Time-Invariant models from multiple records, the study of the asymptotic properties of the estimator continues to be limited. Therefore, we investigated consistency and asymptotic normality as the number or records tend to infinity, with focus on the identification of discrete-time parametric models for single-input single-output systems operating in open loop. This paper presents the results of a consistency and asymptotic normality study based on the analysis of the prediction error cost function and Monte Carlo simulations. We show that for persistently exciting input signals (filtered white noise), model structures such as Output-Error, AR and ARX are consistently estimated, and the estimated parameters are asymptotically normally distributed. On the other hand, ARMA, ARMAX and Box-Jenkins present a bias on the estimated parameters. However, this bias asymptotically disappears for longer records

*Key words:* Discrete-Time LTI models, Parametric models, Data concatenation, Consistency, Asymptotic normality

## 1 Introduction

For many industrial applications, data for system identification is available in the form of multiple records or has long stretches of missing data. This happens for several reasons like sensor or data transmission failure, multiple experiments, or data selection (e.g. selection of operating condition or selection of informative data). Examples of such applications include: the linear-modeling of helicopter flight-mechanics from multiple short flight data [1]; the identification of linear-models for Lithium-Ion batteries and tokamaks with data from multiple experiments [2,3]; and the modeling of Linear Parameter-Varying systems from a local approach where, after operating condition selection, multiple records are available to identify a local Linear Time-Invariant (LTI) model [4].

Applying algorithms for the identification of LTI models on individual data records can lead to poor quality models because these records might not be long enough. Therefore, one needs to exploit all the available data for the identification in order to improve the model quality.

For this, a first approach consists in treating missing data as unknown parameters to be estimated (together with the model parameters) during the identification procedure [5,6,7,8]. This becomes, however, unfeasible for applications where the amount of lost data is large or for multiple records with large or undetermined time gaps between them. Then, the best option is the identification from multiple records. The techniques mentioned in the literature are [1]: the superposition of records (i.e. averaging), the concatenation of records, or a multiple-cost function that combines costs dealing with the individual records (the definition is given in Section 2).

Clearly, the superposition of records does not handle different record lengths and it is not suitable for arbitrary excitation signals because – unless the experiments are synchronized – averaging leads to information loss. Data concatenation and multiple-cost function techniques do not have this limitation. Moreover, these techniques avoid modeling errors due to transient effects by introducing additional parameters that account for the system's initial/final conditions among records. A fundamental difference between these approaches is

⋆ This paper was not presented at any IFAC meeting. Corresponding author Sandra Vásquez.

*Email address:* `sandra.vasquez@vub.be` (Sandra Vásquez).

that, for data concatenation, there is the underlying assumption of records belonging to a system subject to the same operating conditions (otherwise modeling errors occur). We show later in this paper that the multiple-cost approach boils down to data concatenation when the operating conditions of all records are the same.

Even though system identification methods based on one record have been widely studied and well documented [7,8,9], few papers address the identification from multiple records and the analysis of the asymptotic properties of the estimates. [10] proposes a non-parametric estimation of the Frequency Response Function from concatenated data (with transients suppression), and provides an analysis of asymptotic properties such as bias and variance.

On the other hand, for parametric models the analysis of asymptotic properties is quite limited. For instance, [1,7] propose the multiple-cost function approach, and the well known System Identification Toolbox of MATLAB implements it [11]. Also, [8] presents a frequency-domain estimator for parametric models using concatenated data. However, none of these sources provides a discussion on asymptotic properties such as consistency and asymptotic normality when the number of concatenated records tends to infinity. An exception is [12], which presents a consistency analysis for system identification in the errors-in-variables setting departing from multiple-cost functions. This analysis is applicable to the Output Error (OE) model structure. Nevertheless, the analysis lacks from an important step in the proof of convergence of the estimator, which we include in this paper for OE (Lemma 2 in Section 3). Besides [12], consistency analysis for other model structures have not been carried out.

Therefore, we investigate the consistency and asymptotic normality for the identification of discrete-time parametric models with focus on single-input single-output systems and the data concatenation technique. Likewise the case of a single record, it is verified under ideal conditions (LTI system with known dynamic order) whether or not the true dynamics can be recovered by adding more data (concatenating more records). If so, then it makes sense to apply the concatenation technique to real life systems that do not perfectly match the ideal conditions. We consider the more general model structure Box-Jenkins (BJ), and analyze the specific conditions that lead to other model structures such as ARMAX and OE. We study the consistency and asymptotic normality by analyzing the frequency-domain counterpart of the prediction error cost function. Later, we illustrate the results through Monte Carlo simulations. Note that the results of this work are applicable to both time-domain and frequency-domain estimators.

This paper presents the main results as follows. Section 2 provides the problem statement and the definition of the estimator cost function to be analyzed. Next, the analysis of the consistency and the asymptotic normality is presented in Sections 3 and 4. The results are then illustrated by Monte Carlo simulations in Section 5. Finally, Section 6 presents the main conclusions.

## 2 Problem statement

This section first presents the definition of data concatenation. Next, the system to be identified is described. Finally, the prediction error cost function for concatenated records and its equivalent in the frequency-domain are presented.

### *2.1 Data concatenation*

A signal $x_c(t)$, result of the concatenation of $M$ data records $x_m(t)$ of length $N_m$, is defined as

$$x_c(t) = \begin{cases} x_0(t-K_0) & t = K_0, \dots, K_1 - 1 \\ x_1(t-K_1) & t = K_1, \dots, K_2 - 1 \\ \vdots & \vdots \\ x_{M-1}(t-K_{M-1}) & t = K_{M-1}, \dots, N-1 \end{cases}$$

where $N = \sum_{m=0}^{M-1} N_m$ is the length of $x_c(t)$, $K_m = \sum_{i=0}^{m-1} N_i$ for $m \geq 1$ quantifies the difference in samples between the beginning of records $x_m(t)$ and $x_0(t)$ in the concatenated data $x_c(t)$, and $K_0 = 0$.

### *2.2 System description*

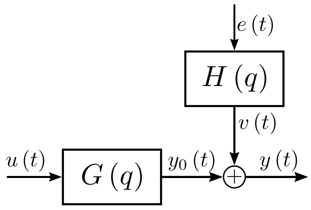



Fig. 1. Plant and noise model.

The system to be identified is represented in Fig. 1, with the input signal $u(t)$ and noise source $e(t)$ satisfying

**Assumption 1.** The input $u(t)$ is a persistently exciting signal (wide-sense stationary random process) that has the form of filtered white noise with finite fourth order moments.

**Assumption 2.** The signal $e(t)$ is zero mean white Gaussian noise with variance $\sigma_e^2$.

The output signal $y(t)$ is described by

$$y(t) = G(q,\theta)\, u(t) + H(q,\theta)\, e(t) \tag{1}$$

$G(q,\theta) = B(q,\theta)/A(q,\theta)$ and $H(q,\theta) = C(q,\theta)/D(q,\theta)$ are the plant and noise rational transfer function models respectively. The polynomial $A$ is defined as

$$A(q,\theta) = \sum_{r=0}^{n_A} a_r q^r \quad (2)$$

with $q$ the backward shift operator $[q^r x(t) = x(t-r)]$, $a_r$ the $r$th coefficient and $n_A$ the order. Same definition applies for polynomials $B$, $C$ and $D$, with $b_r$, $c_r$, $d_r$ the $r$th coefficients and $n_B$, $n_C$, $n_D$ the orders. Here, $\theta$ is the vector of the model parameters (i.e. all unknown polynomials coefficients), and $\theta_0$ denotes its true value. True transfer functions and polynomials are denoted as $A_0(q)$.

Depending on the parametrization of $G(q,\theta)$ and $H(q,\theta)$, one can define the different model structures of Table 1.

Table 1
Definition of the considered model structures (Str).

| Str. | Definition | Str. | Definition |
|---|---|---|---|
| BJ | $A \neq D$ | ARMAX | $A = D$ |
| OE | $C = D = 1$ | ARMA | $A = D$, $B = 0$ |
| FIR | $A = C = D = 1$ | ARX | $A = D$, $C = 1$ |
| | | AR | $A = D$, $B = 0$, $C = 1$ |

BJ: Box-Jenkins, OE: output error, FIR: finite impulse response, AR: auto-regressive, MA: moving average, X: exogenous input

The following assumptions are made.

**Assumption 3.** $G(q,\theta)$ and $H(q,\theta)$ have no zero/pole cancellation. With the plant/noise polynomial coefficients as defined in (2), $a_0 = 1$ for $G(q,\theta)$, and $H(q,\theta)$ is in a monic representation ($c_0 = d_0 = 1$). Thus, $\theta$ is uniquely identifiable. $G(q,\theta)$, $H(q,\theta)$ and $H^{-1}(q,\theta)$ are stable.

**Assumption 4.** The true model belongs to the considered model set: the right model structure and the appropriate model order are chosen.

Note that Assumptions 1 to 4 (with a more relaxed condition for $e(t)$, which is discussed later in Subsection 3.2) make part of the consistency analysis of the *one record* case in [7,8].

### *2.3 Prediction Error cost function*

The prediction error cost function for *one record* is given by [7]

$$V_N(\theta,\psi,\mathtt{z_t}) = \frac{1}{N}\sum_{t=0}^{N-1} \varepsilon^2(t,\theta,\mathtt{z_t})$$
$$\varepsilon(t,\theta,\mathtt{z_t}) = H^{-1}(q,\theta)\left[y(t) - G(q,\theta)u(t)\right]$$

with $\mathtt{z_t} = \left[y^T,\, u^T\right]^T$ the measurement vector in the time-domain ($y$, $u \in \mathbb{R}^{N\times 1}$), $N$ the number of samples, and $\varepsilon(t,\theta,\mathtt{z_t})$ the prediction error. The quadratic norm is a standard choice that is considered for the consistency analysis in [7,8]. The transient parameters $\psi$ (past values of $u(t)$ and $y(t)$) are hidden in the filter operations.

There are $M$ records of input/output signals $u_m(t)$ and $y_m(t)$ with a record length $N_m$ ($m \in [0, M-1]$). The following assumptions, specific for data concatenation, are made

**Assumption 5.** All $M$ records of the input/output signals are independent and are described by (1), with $u(t)$ and $e(t)$ satisfying Assumptions 1 and 2 (this implies a system subject to the same operating conditions). Besides, the system is operating in open loop, so that $e(t)$ and $u(t)$ are independent.

**Assumption 6.** For all $M$ records, the record length $N_m$ satisfies $N_m > n_{\psi r}$, with $n_{\psi r}$ the number of transient parameters $\psi$ associated to each record

$$n_{\psi r} = \begin{cases} \max(n_A, n_B) + \max(n_C, n_D) & \text{for } A \neq D \\ \max(n_B, n_C, n_D) & \text{for } A = D \end{cases} \quad (3)$$

This guaranties that each record will add information for the identification of the model parameters $\theta$ (if $N_m = n_{\psi r}$, then only the transient parameters associated to the $m$th record can be estimated).

Under Assumptions 2 and 5, the Conditional Maximum Likelihood solution consists in combining the data records via the sum of cost functions

$$V_N(\theta,\psi,\mathtt{z_t}) = \frac{1}{N}\sum_{m=0}^{M-1}\sum_{t=0}^{N_m-1} \varepsilon_m^2(t,\theta,\mathtt{z_t}) \quad (4)$$
$$\varepsilon_m(t,\theta,\mathtt{z_t}) = H^{-1}(q,\theta)\left[y_m(t) - G(q,\theta)u_m(t)\right] \quad (5)$$

**Theorem 1.** *Under Assumptions 2 and 5, the Conditional Maximum Likelihood solution* (4) *boils down to the concatenated data solution given by*

$$V_N(\theta,\psi,\mathtt{z_t}) = \frac{1}{N}\sum_{t=0}^{N-1}\left\{H^{-1}(q,\theta)\left[y_c(t) - G(q,\theta)u_c(t)\right]\right\}^2 \quad (6)$$

*Proof.* $\varepsilon_c = \left[\varepsilon_0^T, \cdots, \varepsilon_{M-1}^T\right]^T$ with $\varepsilon_m \in \mathbb{R}^{N_m\times 1}$ and $\varepsilon_m(t)$ given by (5). Then, $\varepsilon_c^T\varepsilon_c = \sum_{m=0}^{M-1}\varepsilon_m^T\varepsilon_m$. ☐

Note that Assumption 6 allows the concatenation of very short records. This is an important advantage in comparison with the non-parametric modeling proposed in [10], which is better suited for longer records (due to interpolation issues of the Local Polynomial Method).

The frequency-domain equivalent of the cost function in (6) is presented in the following subsection.

### 2.4 Frequency-Domain cost function

The Discrete Fourier Transform (DFT) of a signal $x_c(t)$ is given by

$$X_c(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_c(t)\, z_k^{-t}$$

with $k \in [0, N-1]$ the frequency bin, and $z_k = e^{j2\pi k/N}$.

The measurement vector in the frequency-domain is $\mathtt{z_f} = \left[Y_c^T, U_c^T\right]^T$, with $Y_c$ and $U_c$ the DFT of $y_c(t)$ and $u_c(t)$ ($Y_c, U_c \in \mathbb{C}^{N\times 1}$). By the Parseval's theorem, the cost function (6) has the following frequency-domain equivalent [8]

$$V_N(\theta, \psi, \mathtt{z_f}) = \frac{1}{N} \sum_{k=0}^{N-1} |\varepsilon_k(\theta, \psi, \mathtt{z_f})|^2 \tag{7}$$

$$\varepsilon_k(\theta,\psi,\mathtt{z_f}) = H^{-1}(z_k^{-1},\theta)\Big\{Y_c(k) - G(z_k^{-1},\theta)\, U_c(k) - \sum_{m=0}^{M-1} z_k^{-K_m}\left[T_{G\,m}(z_k^{-1},\theta,\psi) + T_{H\,m}(z_k^{-1},\theta,\psi)\right]\Big\} \tag{8}$$

with $T_{G\,m}(z^{-1},\theta,\psi)$ and $T_{H\,m}(z^{-1},\theta,\psi)$ the plant and noise transient terms associated to the $m$th record and $z^{-K_m}$ their corresponding delay term (with $K_m$ as defined in Subsection 2.1). The transient terms share the dynamics of the plant/noise model, and are defined as

$$T_{G\,m}(z^{-1},\theta,\psi) = I_m(z^{-1},\psi)/A(z^{-1},\theta)$$
$$T_{H\,m}(z^{-1},\theta,\psi) = J_m(z^{-1},\psi)/D(z^{-1},\theta)$$

with $I_m(z^{-1},\psi)$ and $J_m(z^{-1},\psi)$ polynomials of order $n_I$ and $n_J$ given by

$$n_I = \max(n_A, n_B) - 1; \quad n_J = \max(n_C, n_D) - 1$$

Here, $\psi \in \mathbb{R}^{n_\psi \times 1}$ is the vector of coefficients of the polynomials $I_m$ and $J_m$ for $m \in [0, M-1]$. Note that for $D = A$, the coefficients of $I_m$ and $J_m$ cannot be identified separately, so that $n_\psi = M\, n_{\psi r}$, with $n_{\psi r}$ as defined in (3).

The noise variance $\sigma_e^2$ can be estimated from (7) by replacing the model and transient parameters with their estimates $\hat{\theta}$ and $\hat{\psi}$, and accounting for the degrees of freedom as follows

$$\hat{\sigma}_e^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left|\varepsilon_k\left(\hat{\theta}, \hat{\psi}, \mathtt{z_f}\right)\right|^2 \frac{N}{N - n_\theta - n_\psi} \tag{9}$$

The cost function (7) is analyzed in the following section.

## 3 Consistency

This section presents the consistency analysis of the frequency-domain cost function for concatenated data. The analysis is performed as follows:

- First, the cost function $V_N(\theta,\psi,\mathtt{z_f})$ is expressed in terms of the model parameters $\theta$ only [$V_N(\theta,\mathtt{z_f})$]. To this end, the values of the transient parameters $\psi$ that minimize $V_N(\theta,\psi,\mathtt{z_f})$ are substituted in (7). Then, the expected value of this cost function is computed [$V_N(\theta) = \mathbb{E}\{V_N(\theta,\mathtt{z_f})\}$]. See Subsection 3.1.
- Next, one needs to prove the uniform convergence (w.r.t. $\theta$) of $V_N(\theta,\mathtt{z_f})$ to its expected value $V_N(\theta)$, and of $V_N(\theta)$ to the asymptotic cost function $V_*(\theta) = \lim_{N\to\infty} V_N(\theta)$. Then, the convergence of the global minimizers of $V_N(\theta,\mathtt{z_f})$, $V_N(\theta)$ and $V_*(\theta)$ is established from the uniform convergence of these cost functions. See Subsection 3.2.
- Finally, to establish the consistency, one needs to verify that the asymptotic cost function $V_*(\theta)$ is minimal in the true model parameters $\theta_0$. See Subsection 3.3.

This methodology is inspired in [8] (Chapter 17), with the results presented in this section (and its Appendices) original from this paper.

The analysis is carried out for the model structures of Table 1. Note that two cases lead to $N \to \infty$: finite number of records $M$ with record lengths $N_m \to \infty$, or finite $N_m$ with $M \to \infty$. According to (4), for $N_m \to \infty$, $V_N(\theta,\psi,\mathtt{z_t})$ is a sum of cost functions that yield consistent estimates. Therefore, the case of finite $N_m$ with $M \to \infty$ is the focus of this paper.

### 3.1 Definition of cost functions $V_N(\theta,\mathtt{z_f})$ and $V_N(\theta)$

This subsection provides the definitions of $V_N(\theta,\mathtt{z_f})$ and $V_N(\theta)$ for the more general model structure BJ. These definitions apply to all model structures of Table 1, with appropriate modifications for ARMAX and its special cases AR and ARX.

The cost function (7) and the error vector $\varepsilon(\theta,\psi,\mathtt{z_f}) \in \mathbb{C}^{N\times 1}$ [with the $k$th element given by (8)] can be expressed as

$$V_N(\theta,\psi,\mathtt{z_f}) = \frac{1}{N}\varepsilon^H(\theta,\psi,\mathtt{z_f})\,\varepsilon(\theta,\psi,\mathtt{z_f}) \tag{10}$$

$$\varepsilon(\theta,\psi,\mathtt{z_f}) = W(\theta)\,\mathtt{z_f} - L_P(\theta)\,\psi \tag{11}$$

with the superscript $H$ denoting the Hermitian transpose, $\mathtt{z_f} = \left[Y_c^T, U_c^T\right]^T \in \mathbb{C}^{2N\times 1}$ and $\psi \in \mathbb{R}^{n_\psi\times 1}$. $W \in \mathbb{C}^{N\times 2N}$ and $L_P \in \mathbb{C}^{N\times n_\psi}$ are matrices that depend on $\theta$. They are defined in Appendices A.1 and A.4 for, respectively, the BJ and ARMAX model structures.

The value of $\psi$ that minimizes (10) is obtained by solving $\partial V_N(\theta,\psi,\mathtt{z_f})/\partial\psi = 0$. Substitution of the solution $\psi =$

$\left(L_P^H L_P\right)^{-1} L_P^H W \mathbf{z_f}$ in (10) results in the cost function

$$V_N(\theta, \mathbf{z_f}) = \frac{1}{N} \mathbf{z_f}^H \mathsf{W}_N(\theta) \mathbf{z_f} \geq 0 \quad (12)$$

$$\mathsf{W}_N(\theta) = W(\theta)^H P(\theta) W(\theta) \quad (13)$$

with $\mathsf{W}_N \in \mathbb{C}^{2N \times 2N}$ a weighting matrix function of $\theta$, and $P \in \mathbb{C}^{N \times N}$ a projection matrix defined as

$$P(\theta) = I_N - L_P(\theta) \left(L_P^H(\theta) L_P(\theta)\right)^{-1} L_P^H(\theta) \quad (14)$$

The expected value of (12) is given by

$$V_N(\theta) = \frac{1}{N} \operatorname{tr}\left[\mathbb{E}\left\{W(\theta) \mathbf{z_f} \mathbf{z_f}^H W(\theta)^H\right\} P(\theta)\right] \quad (15)$$

with tr [ ] the trace operator. Computing $\mathbb{E}\left\{W \mathbf{z_f} \mathbf{z_f}^H W^H\right\}$ requires expressing the output $Y_c$ in terms of the input $U_c$, the noise source $E_c$, and the true model and transient parameters $\theta_0$ and $\psi_0$, resulting in the following expectations

$$\begin{aligned} S_{UU} &= \mathbb{E}\{U_c U_c^H\};\ S_{EE} = \mathbb{E}\{E_c E_c^H\};\ S_{\psi\psi} = \mathbb{E}\{\psi_0 \psi_0^H\} \\ S_{EU} &= \mathbb{E}\{E_c U_c^H\};\ S_{\psi U} = \mathbb{E}\{\psi_0 U_c^H\};\ S_{\psi E} = \mathbb{E}\{\psi_0 E_c^H\} \end{aligned} \quad (16)$$

The contribution of these terms to the cost function (15) can be expressed as (see Appendix A.2)

$$\begin{aligned} V_N(\theta) = &V_{UU}(\theta) + V_{EE}(\theta) + V_{\psi\psi}(\theta) \\ &+ V_{EU}(\theta) + V_{\psi U}(\theta) + V_{\psi E}(\theta) \end{aligned} \quad (17)$$

The derivative of $V_N(\theta)$ with respect to the model parameters $\theta$ and the expected values $S_{EE}$, $S_{EU}$ and $S_{\psi E}$ are computed in Appendices A.3 and B. Appendix C presents preliminary Lemmas required for the consistency analysis of the following subsections.

### 3.2 *Convergence of cost functions and minimizers*

The convergence analysis in this subsection refers to the cost function $V_N(\theta, \mathbf{z_f})$ in (12), its expected value $V_N(\theta) = \mathbb{E}\{V_N(\theta, \mathbf{z_f})\}$ in (17), and the asymptotic cost function $V_*(\theta) = \lim_{N\to\infty} V_N(\theta)$.

To prove the stochastic convergence of $V_N(\theta, \mathbf{z_f})$ to $V_N(\theta)$, we make the following assumption regarding the record lengths $N_m$. Later on, we will relax this condition.

**Assumption 7.** All $M$ records have the same length $N_r$ ($N_m = N_r$, $m \in [0, M-1]$) that satisfies Assumption 6.

**Lemma 2.** *The weighting matrix* $\mathsf{W}_N$ *in* (12) *is a positive definite Hermitian matrix that under Assumption 7 satisfies* $\|\mathsf{W}_N\|_1 < c < \infty$ *for* $N > N_*$, $\infty$ *included, where c is an* $N$*-independent constant.*

*Proof.* See Appendix D. □

**Lemma 3.** *Under Assumptions 1, 2, 3, 5 and 6,* $V_N(\theta, \mathbf{z_f})$ *is a continuous function of* $\theta$ *in a compact (i.e. closed and bounded) set of parameters* $\Theta$, *with* $\theta_0 \in \Theta$.

*Proof.* Under Assumptions 1, 2, 5 and 6, the projection matrix $P$ is of full rank. Under Assumption 3, $W$ and $L_P$ are continuous functions of $\theta$, and so is $(L_P^H L_P)$, its inverse $(L_P^H L_P)^{-1}$ and $P$. Therefore, $\mathsf{W}_N$ is a continuous function of $\theta$ and so is $V_N(\theta, \mathbf{z_f})$. The compactness of the parameter space should only be satisfied in the neighborhood of $\theta_0$. The construction of $\Theta$ is discussed in Chapter 17 of [8]. □

**Lemma 4.** *Under the Assumptions of Lemmas 2 and 3,* $V_N(\theta, \mathbf{z_f})$ *converges uniformly w.r.t.* $\theta$ *(in probability) to* $V_N(\theta)$ *in the compact set* $\Theta$.

*Proof.* Under Assumptions 1, 2, 3, 5, 6 and 7, the conditions of Lemma 17.3 in Chapter 17 of [8] are fulfilled (see Lemmas 2 and 3, and Appendix N). □

**Remark 5.** Relaxing Assumption 2 to $e(t)$ being white noise with finite fourth order moments requires another proof of convergence, since the conditions of Lemma 17.3 in Chapter 17 of [8] are not fulfilled for non-Gaussian $e(t)$. Though, simulations not included in this paper indicate there is convergence. Besides, the analysis of $V_N(\theta)$ carried out in the following subsection holds [because (B.1) remains true].

**Lemma 6.** $V_N(\theta)$ *converges uniformly w.r.t.* $\theta$ *to* $V_*(\theta)$ *in the compact set* $\Theta$.

*Proof.* This is the result of the convergence of the Riemann sum to the corresponding Riemann integral [with $\sum_{k=0}^{N-1} V_N(\theta, k) = \sum_{k=-N/2+1}^{N/2} V_N(\theta, k)$, and $f_s$ the sampling frequency]

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=-N/2+1}^{N/2} V_N(\theta, k) = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} V_N(\theta, f)\, df \quad (18)$$

□

In the presence of modeling errors, it may happen that $V_N(\theta, \mathbf{z_f})$, $V_N(\theta)$ and $V_*(\theta)$ do not have a unique global minimum [7]. Hence, we define the following sets of minimizing values in the compact set $\Theta$

$$\Theta_{N\,\mathbf{z_f}} = \arg\min_{\theta\in\Theta} V_N(\theta, \mathbf{z_f});\quad \Theta_N = \arg\min_{\theta\in\Theta} V_N(\theta)$$
$$\Theta_* = \arg\min_{\theta\in\Theta} V_*(\theta)$$

**Theorem 7.** *Under the Assumptions of Lemma 4,* $\Theta_N$ *converges to* $\Theta_*$ *and* $\Theta_{N\,\mathbf{z_f}}$ *converges in probability to* $\Theta_*$:

$$\lim_{N\to\infty} \Theta_N = \Theta_* \quad \textit{and} \quad \mathit{plim}_{N\to\infty} \Theta_{N\,\mathbf{z_f}} = \Theta_*$$

*Besides, without modeling errors, the set* $\Theta_*$ *contains the unique minimizer* $\theta_* = \theta_0$.

*Proof.* Under Assumptions 1, 2, 3, 5, 6 and 7, the conditions of Theorem 17.8 in Chapter 17 of [8] are fulfilled (see Lemmas 4 and 6). The proof for Theorem 17.8 is based on [13]. It suffices to replace *strong* by *weak convergence* (i.e. in probability).

The uniqueness of $\theta_*$ follows directly from Assumptions 1 and 3 and the non-existence of modeling errors. □

**Corollary 8.** *Theorem 7 can be generalized to the case of arbitrary record lengths $N_m$ satisfying Assumption 6.*

*Proof.* See Appendix E. □

This concludes the convergence proof for the cost functions and minimizers. The cost function analysis is presented in the following subsection.

### *3.3 Cost function analysis*

This subsection presents the analysis of the cost function $V_N(\theta)$ in (17) and the asymptotic value $V_*(\theta) = \lim_{N\to\infty} V_N(\theta)$, for finite $N_m$ with $M \to \infty$.

The first necessary condition for consistency is presented in Assumption 4. Note that for the consistency analysis, the model structure and order match exactly the true model. The second condition for consistency is $V_*(\theta)$ being minimal in the true model parameters $\theta_0$. For this, the following condition must be satisfied for each model parameter $\theta_r$

$$\left.\frac{\partial V_*(\theta)}{\partial \theta_r}\right|_{\theta_0} = \lim_{M\to\infty} \left.\frac{\partial V_N(\theta)}{\partial \theta_r}\right|_{\theta_0} = 0 \tag{19}$$

with $\theta_r$ meaning the $r$th coefficient of the plant ($A$, $B$) or noise ($C$, $D$) polynomials [i.e. $a_r$, $b_r$, $c_r$ or $d_r$, see (2)].

The following Lemmas refer to (19) considering $V_N(\theta)$ as the sum of terms in (17).

**Lemma 9.** *For any model structure of Table 1, each of the terms $V_{UU}(\theta)$, $V_{\psi\psi}(\theta)$, and $V_{\psi U}(\theta)$ satisfies condition* (19) *for all model parameters. Moreover, given Assumption 5, so does $V_{EU}(\theta)$.*

*Proof.* See Appendix F. □

**Lemma 10.** *For the BJ model structure, $V_{EE}(\theta)$ satisfies condition* (19) *for $a_r$ and $b_r$, but not for $c_r$ and $d_r$ when $N_m$ is finite and $M \to \infty$. However, as $N_m \to \infty$, $V_{EE}(\theta)$ satisfies condition* (19) *for $c_r$ and $d_r$.*

*Proof.* See Appendix G. □

**Lemma 11.** *For the BJ model structure, $V_{\psi E}(\theta)$ satisfies condition* (19) *for $a_r$, $b_r$ and $c_r$ for finite $N_m$ and $M \to \infty$. As $N_m \to \infty$, $V_{\psi E}(\theta)$ satisfies condition* (19) *for $d_r$.*

*Proof.* See Appendix H. □

**Lemma 12.** *For the BJ model structure, for $V_N(\theta)$ to satisfy condition* (19) *for $a_r$, it is necessary that $H(z^{-1}, \theta) = H_0(z^{-1})$. This implies that a bias on the noise model parameters ($c_r$ or $d_r$) introduces a bias on the plant model parameters ($a_r$ and $b_r$).*

*Proof.* See Appendix I. □

**Lemma 13.** *The results of Lemmas 10 and 11 for parameters $b_r$, $c_r$ and $d_r$ can be extended to the ARMAX model structure, and the results for $c_r$ and $d_r$ to the ARMA model structure (since $D = A$, what is stated for $d_r$ applies to $a_r$).*

*Proof.* The proof follows the same procedure as for Lemmas 10 and 11, with the modifications provided in Appendix A.4. □

**Theorem 14.** *Under Assumptions 1 to 6, for the BJ, ARMAX and ARMA model structures the estimate of the model parameters $\hat{\theta}(\mathbf{z_f}) \in \Theta_{N\,\mathbf{z_f}}$ is not consistent for finite $N_m$ and $M \to \infty$. $\hat{\theta}(\mathbf{z_f})$ presents a bias that vanishes for increasing record lengths $N_m$.*

*Proof.* It follows directly from Corollary 8 and Lemmas 9 to 13. □

**Lemma 15.** *For the ARX, AR, OE and FIR model structures, $V_{EE}(\theta)$ and $V_{\psi E}(\theta)$ satisfy condition* (19) *for all model parameters for finite $N_m$ and $M \to \infty$.*

*Proof.* Because OE is a special case of BJ (with $n_C = n_D = 0$), the results of Lemmas 10 and 11 for parameters $a_r$ and $b_r$ apply to OE, and so do for FIR, which is a special case of OE (with $n_A = 0$). For ARX see the proof in Appendix J. The proof for AR follows the same procedure as for ARX. □

**Theorem 16.** *Under Assumptions 1 to 6, for the ARX, AR, OE and FIR model structures the estimate of the model parameters $\hat{\theta}(\mathbf{z_f}) \in \Theta_{N\,\mathbf{z_f}}$ is consistent for finite $N_m$ and $M \to \infty$: $\text{plim}_{N\to\infty}\hat{\theta}(\mathbf{z_f}) = \theta_0$*

*Proof.* It follows directly from Corollary 8 and Lemmas 9 and 15. □

**Corollary 17.** *Under Assumptions 1 to 6, for the ARX, AR, OE and FIR model structures the estimate of the noise variance $\hat{\sigma}_e^2$ in* (9) *is consistent for finite $N_m$ and $M \to \infty$. For BJ, ARMAX and ARMA, $\hat{\sigma}_e^2$ is not consistent, and it presents a bias that vanishes for increasing record lengths $N_m$.*

*Proof.* It follows from Theorems 14 and 16 and the proof in Appendix K. □

This concludes the consistency analysis. In the following section, the asymptotic normality is studied.

## 4 Asymptotic normality

This section presents the study of the asymptotic normality of the consistently estimated model structures (Theorem 16). To prove the asymptotic normality of $\hat{\theta}(\mathbf{z_f})$, we first make use of Assumption 7. Later on, we will relax this condition.

**Lemma 18.** *For the ARX, AR, OE and FIR model structures, the weighting matrix* $\mathsf{W}_N$ *in* (12) *has continuous first- second- and third-order derivatives w.r.t.* $\theta$*, that under Assumption 7, have bounded 1-norm.*

*Proof.* See Appendix L. □

**Theorem 19.** *Under Assumptions 1 to 7, for the ARX, AR, OE and FIR model structures the estimate of the model parameters* $\hat{\theta}(\mathbf{z_f}) \in \Theta_{N\,\mathbf{z_f}}$ *converges in law at the rate* $O(N^{-1/2})$*, for finite* $N_m$ *and* $M \to \infty$*, to a Gaussian random variable with zero mean and covariance*

$$\operatorname{Cov}(\hat{\theta}) = V_N^{''-1}(\theta_0)\,\mathbb{E}\Big\{V_N^{'T}(\theta_0, \mathbf{z_f})\,V_N^{'}(\theta_0, \mathbf{z_f})\Big\}V_N^{''-1}(\theta_0) \tag{20}$$

*Proof.* See Appendix M. □

**Corollary 20.** *Theorem 19 can be generalized to the case of arbitrary record lengths* $N_m$ *satisfying Assumption 6.*

*Proof.* The same reasoning as for the proof for Corollary 8 can be followed. □

This concludes the study of the asymptotic properties of the estimator. In the following section, the results are illustrated by Monte Carlo simulations.

## 5 Monte Carlo simulation

To illustrate the results of the previous sections for the BJ and AR model structures, Monte Carlo simulations (1000 runs) were performed to evaluate the mean value and the mean squared error (MSE) of the estimated plant/noise models for increasing $M$. Note that the confidence bounds presented in this section are constructed based on these Monte Carlo experiments. The following cases are compared: the concatenation of records of the same length for minimal [$N_{r1} = n_{\psi r} + 1$, with $n_{\psi r}$ given by (3)] and longer ($N_{r2} > N_{r1}$) records, and one full record ($N$) as reference case. To compare the MSE, the total number of samples for all cases is the same ($N_{r1}M_1 = N_{r2}M_2 = N$).

The following true models are considered

$$G_{BJ\,0}(z^{-1}) = 1/(1-0.15z^{-1})$$
$$H_{BJ\,0}(z^{-1}) = (1+0.5z^{-1})/(1-1.13z^{-1}+0.64z^{-2})$$
$$H_{AR\,0}(z^{-1}) = 1/(1-0.15z^{-1})$$

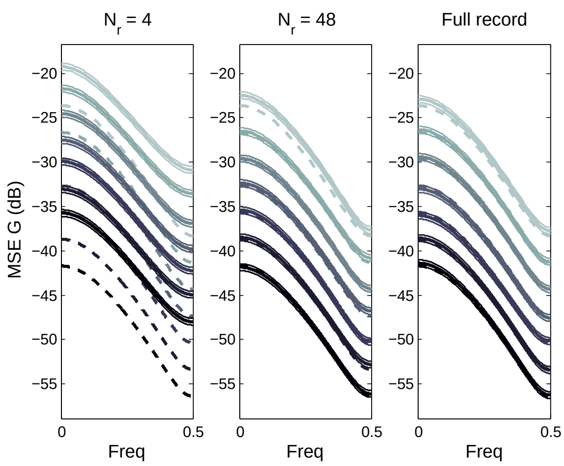
with transients polynomials $I_{m\,BJ}$, $J_{m\,BJ}$ of order $n_I = 0$, $n_J = 1$, and $J_{m\,AR}$ of order $n_J = 0$. The excitation and noise signals $[u(t), e(t)]$ are independent, zero mean, normally distributed random variables with variances $\sigma_u^2 = 1$ and $\sigma_e^2 = 0.3$ for BJ, and $\sigma_e^2 = 1$ for AR.

Figure 2 presents the MSE of the estimated transfer functions $G_{BJ}(z_k^{-1}, \hat{\theta})$, $\hat{\sigma}_e H_{BJ}(z_k^{-1}, \hat{\theta})$ and $\hat{\sigma}_e H_{AR}(z_k^{-1}, \hat{\theta})$ for $N$ doubled each time ($N \in \{96, \ldots, 6144\}$ for BJ, $N \in \{48, \ldots, 3072\}$ for AR). The Cramér–Rao Lower Bound (CRLB) of the full record case is included to provide a lower limit for the MSE of all cases. The 95% confidence interval of the MSE is provided to facilitate the comparison of the MSE with the CRLB.

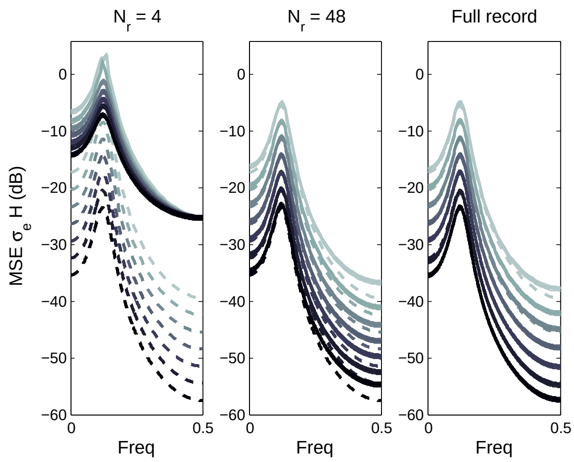
From Fig. 2a, 2b and 2c, it can be seen that for the full record case the MSE quickly reaches the CRLB for increasing values of $N$. In contrast, for the data concatenation cases the MSE approaches the CRLB but does not reach it. The difference between the MSE and the CRLB reflects the information loss for the data concatenation cases, since $n_\psi$ out of $N$ samples serve to estimate the transient parameters, with $n_\psi \propto M$. Hence, for longer record lengths $N_r$ the MSE becomes closer to the CRLB.

For a consistent estimation from concatenated data, the MSE should decrease with $N$ at the same rate as the full record case ($\sqrt{\text{MSE}} \propto 1/\sqrt{N}$). This seems to be the case for $G_{BJ}(z_k^{-1}, \hat{\theta})$ and $\hat{\sigma}_e H_{AR}(z_k^{-1}, \hat{\theta})$ (Fig. 2a and 2c): doubling $N$ leads to a decrease of $\sqrt{\text{MSE}}$ by 3dB [3dB $\approx$ dB($\sqrt{2}$)]. In contrast, $\hat{\sigma}_e H_{BJ}(z_k^{-1}, \hat{\theta})$ (Fig. 2b) clearly presents a bias: for increasing values of $N$ the MSE first decreases but eventually reaches a lower limit. However, this bias diminishes for longer record lengths $N_r$. Indeed, considering the time constant $\tau$ of $H_{BJ\,0}$ (with $\tau = -1/\operatorname{Re}\{\ln(z_p)\}$[samples], and $z_p$ the dominant pole), a record length $N_r = 48$ ($\sim 10.7\tau$) makes the bias negligible ($\sqrt{\text{MSE}}$ keeps decreasing by 3dB).

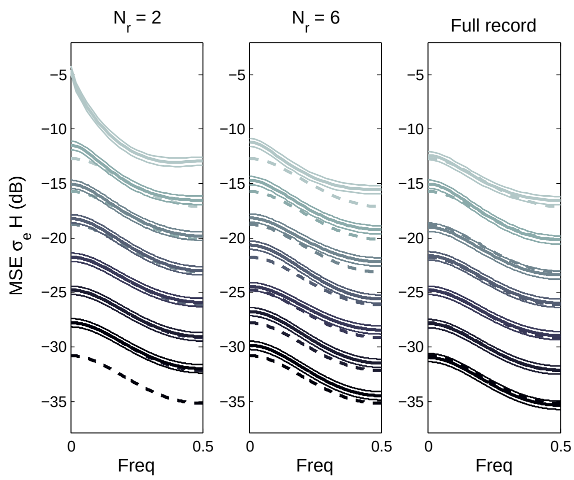
Figure 3 presents, for the minimal records case and increasing $N$, the mean value of the estimated parameters $\hat{a}_1$ (for BJ) and $\hat{d}_1$ (for AR) and the noise variance $\hat{\sigma}_e^2$ (for AR), with the 95% confidence intervals of the mean values. Because the true parameter $a_1$ is not contained in the confidence intervals (for $N \to \infty$), it is clear that a bias is present for $G_{BJ}(z_k^{-1}, \hat{\theta})$ (this




(a) BJ Plant model $G_{BJ}(z_k^{-1}, \hat{\theta})$.




(b) BJ Noise model $\hat{\sigma}_e H_{BJ}(z_k^{-1}, \hat{\theta})$.




(c) AR Noise model $\hat{\sigma}_e H_{AR}(z_k^{-1}, \hat{\theta})$.

Fig. 2. Monte Carlo simulation (1000 runs) for BJ and AR model structures. Cases: minimal record length (left), longer records (middle), and one full record (right). $N$ is doubled each time (from light to dark). Magnitudes presented in dB: $\sqrt{\text{MSE}}$ of estimated model (thick solid line) and its 95% confidence interval (solid line), Cramér–Rao Lower Bound of the full record case (dashed line).

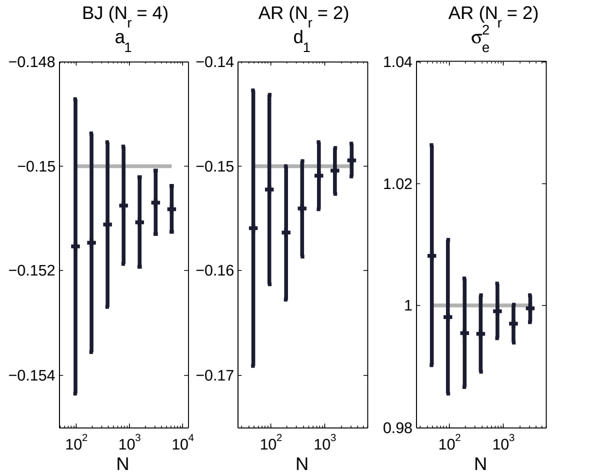



Fig. 3. Monte Carlo simulation (1000 runs) of the minimal record length case. Mean of estimated model parameters $\hat{a}_1$ and $\hat{d}_1$ (for BJ and AR respectively) and noise variance $\hat{\sigma}_e^2$ (for AR), with the 95% confidence intervals of the mean values (black line), true parameters (gray line). $N$ is doubled each time.

is explained by Lemma 12). This bias is however small for this example, so that it could not be easily detected in Fig. 2a. On the other hand, $d_1$ and $\sigma_e^2$ are contained in their respective confidence intervals, which indicates a consistent estimation of $\hat{\sigma}_e H_{AR}(z_k^{-1}, \hat{\theta})$. Besides, simulation results confirmed the asymptotic normality for $\hat{d}_1$.

The results of these Monte Carlo simulations are in accordance with Theorems 14, 16 and 19, and Corollary 17.

## 6 Conclusions

In this paper we studied the estimator's asymptotic properties (i.e. consistency and asymptotic normality) for the identification of single-input single-output systems from concatenated data. We showed that the multiple-cost solution for data records belonging to a system subject to the same operating conditions boils down to the data concatenation solution expressed in the time and frequency domain by (6) and (7).

Moreover, under Assumptions 1 to 6, the concatenation of multiple records of finite length leads asymptotically, for increasing number of records, to a consistent parameter estimation of discrete-time LTI models for the structures AR, ARX, OE, FIR. In contrast, the parameter estimation for BJ, ARMAX and ARMA is not consistent. Besides, for the consistently estimated model structures, the estimator is asymptotically normally distributed.

Therefore, the data concatenation technique is a suitable solution for the identification from multiple records because of the improved model quality and increased frequency resolution, as compared with a model estimated from a single record. The downside with regard to an estimation departing from a full long record is the increased uncertainty, and a bias in the case of BJ, ARMAX and ARMA model structures. However, this bias asymptotically disappears for longer records. Simulation results suggest that record lengths of at least some few times (e.g. 10) the dominant time constant of the plant and noise models result in a negligibly small bias.

Further research includes: the development of a convergence proof for a relaxed assumption regarding the normality of the noise source $e(t)$ (see Remark 5); the study of consistency for data concatenation in the case of multiple-input multiple-output systems; and the consistency analysis for the estimation of the Best Linear Approximation with concatenated data.


## Acknowledgements

This work was supported by the Fonds de la Recherche Scientifique – FNRS (research fellow grant).


## Appendix A Definitions for the cost functions in Subsection 3.1

First, the definitions are provided for BJ. Next, modifications for ARMAX are presented in Appendix A.4.

### *A.1 Definitions for $V_N(\theta, \psi, \mathtt{z_f})$ and $V_N(\theta, \mathtt{z_f})$*

To express the error vector $\varepsilon(\theta, \psi, \mathtt{z_f})$ as (11), the summations of transient polynomials in (8) are rewritten as

$$\sum_{m=0}^{M-1} z_k^{-K_m} I_m(z_k^{-1}, \psi) = L_{I_k}\psi_I; \sum_{m=0}^{M-1} z_k^{-K_m} J_m(z_k^{-1}, \psi) = L_{J_k}\psi_J \tag{A.1}$$

where $\psi_I \in \mathbb{R}^{M(n_I+1)\times 1}$ and $\psi_J \in \mathbb{R}^{M(n_J+1)\times 1}$ are vectors grouping the plant and noise transient parameters

$$\psi_I = \left[\psi_{I\,0}^T, \cdots, \psi_{I\,M-1}^T\right]^T; \quad \psi_J = \left[\psi_{J\,0}^T, \cdots, \psi_{J\,M-1}^T\right]^T$$

with $\psi_{I\,m}$ and $\psi_{J\,m}$ corresponding to the polynomials $I_m$ and $J_m$. $L_{I_k} \in \mathbb{C}^{1\times M(n_I+1)}$ and $L_{J_k} \in \mathbb{C}^{1\times M(n_J+1)}$ are vectors of powers of $z_k^{-1}$ that include the delay terms $z_k^{-K_m}$

$$L_{I_k} = \left[\eta_{I_k} z_k^{-K_0}, \cdots, \eta_{I_k} z_k^{-K_{M-1}}\right] \tag{A.2}$$

$$L_{J_k} = \left[\eta_{J_k} z_k^{-K_0}, \cdots, \eta_{J_k} z_k^{-K_{M-1}}\right] \tag{A.3}$$

$$\eta_{I_k} = \left[1,\, z_k^{-1}, \cdots,\, z_k^{-n_I}\right]; \quad \eta_{J_k} = \left[1,\, z_k^{-1}, \cdots,\, z_k^{-n_J}\right] \tag{A.4}$$

Hence, in (11), $\psi \in \mathbb{R}^{n_\psi \times 1}$, $L_P \in \mathbb{C}^{N\times n_\psi}$, and $W \in \mathbb{C}^{N\times 2N}$ are defined as

$$\psi = \left[\psi_I^T, \psi_J^T\right]^T \tag{A.5}$$

$$L_P = \left[D_{PI}L_I,\, D_{PJ}L_J\right] \tag{A.6}$$

$$W = \left[D_{WY},\, D_{WU}\right] \tag{A.7}$$

with $L_I \in \mathbb{C}^{N\times M(n_I+1)}$, $L_J \in \mathbb{C}^{N\times M(n_J+1)}$, and $D_{PI}$, $D_{PJ}$, $D_{WY}$ and $D_{WU}$ diagonal matrices belonging to $\mathbb{C}^{N\times N}$ given by

$$L_I = \left[L_{I_0}^T, \cdots, L_{I_{N-1}}^T\right]^T; \; L_J = \left[L_{J_0}^T, \cdots, L_{J_{N-1}}^T\right]^T \tag{A.8}$$

$$D_{PI} = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \frac{1}{A(z_k^{-1}, \theta)}, \cdots\right) \tag{A.9}$$

$$D_{PJ} = \operatorname{diag}\left(\cdots, \frac{1}{C(z_k^{-1}, \theta)}, \cdots\right) \tag{A.10}$$

$$D_{WY} = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)}, \cdots\right) \tag{A.11}$$

$$D_{WU} = \operatorname{diag}\left(\cdots, -\frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \frac{B(z_k^{-1}, \theta)}{A(z_k^{-1}, \theta)}, \cdots\right) \tag{A.12}$$

### *A.2 Definitions for $V_N(\theta)$*

In (11), expressing $Y_c$ in terms of $U_c$, $E_c$ and the true model and transient parameters $\theta_0$ and $\psi_0$ results in $W\mathtt{z_f}$ given by

$$W\mathtt{z_f} = D_U U_c + D_E E_c + L_\psi \psi_0$$

with $\psi_0 \in \mathbb{R}^{n_\psi \times 1}$, $L_\psi \in \mathbb{C}^{N\times n_\psi}$, and $D_U$, $D_E$, $D_{\psi I}$ and $D_{\psi J}$ diagonal matrices belonging to $\mathbb{C}^{N\times N}$ defined as

$$\psi_0 = \left[\psi_{0I}^T, \psi_{0J}^T\right]^T; \quad L_\psi = \left[D_{\psi I}L_I, D_{\psi J}L_J\right] \tag{A.13}$$

$$D_U = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \left[\frac{B_0(z_k^{-1})}{A_0(z_k^{-1})} - \frac{B(z_k^{-1}, \theta)}{A(z_k^{-1}, \theta)}\right], \cdots\right) \tag{A.14}$$

$$D_E = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \frac{C_0(z_k^{-1})}{D_0(z_k^{-1})}, \cdots\right) \tag{A.15}$$

$$D_{\psi I} = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \frac{1}{A_0(z_k^{-1})}, \cdots\right) \tag{A.16}$$

$$D_{\psi J} = \operatorname{diag}\left(\cdots, \frac{D(z_k^{-1}, \theta)}{C(z_k^{-1}, \theta)} \frac{1}{D_0(z_k^{-1})}, \cdots\right) \tag{A.17}$$

Computing $\mathbb{E}\left\{W\mathtt{z_f}\mathtt{z_f}^H W^H\right\}$ in (15), given the definitions of (16), results in the cost function (17), with

$$
\begin{aligned}
V_{UU}(\theta) &= \frac{1}{N}\operatorname{tr}\left[D_U S_{UU} D_U^H P\right]\\
V_{EE}(\theta) &= \frac{1}{N}\operatorname{tr}\left[D_E S_{EE} D_E^H P\right]\\
V_{\psi\psi}(\theta) &= \frac{1}{N}\operatorname{tr}\left[L_\psi S_{\psi\psi} L_\psi^H P\right]\\
V_{EU}(\theta) &= \frac{1}{N}\operatorname{tr}\left[2\,\mathrm{herm}\{D_E S_{EU} D_U^H\}\ P\right]\\
V_{\psi U}(\theta) &= \frac{1}{N}\operatorname{tr}\left[2\,\mathrm{herm}\{L_\psi S_{\psi U} D_U^H\}\ P\right]\\
V_{\psi E}(\theta) &= \frac{1}{N}\operatorname{tr}\left[2\,\mathrm{herm}\{L_\psi S_{\psi E} D_E^H\}\ P\right]\\
&\text{where}\quad \mathrm{herm}\{X\} = \left(X + X^H\right)/2 \qquad \text{(A.18)}
\end{aligned}
$$

### *A.3 Computation of $\partial V_N(\theta)/\partial\theta_r$*

The following equations present the derivative of each term of (17) with respect to the $r$th model parameter $\theta_r$, with herm{} defined by (A.18)

$$
\frac{\partial V_{UU}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\left(2\,\mathrm{herm}\left\{\frac{\partial D_U}{\partial\theta_r} S_{UU} D_U^H\right\} P + D_U S_{UU} D_U^H \frac{\partial P}{\partial\theta_r}\right) \qquad \text{(A.19)}
$$

$$
\frac{\partial V_{EE}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\left(2\,\mathrm{herm}\left\{\frac{\partial D_E}{\partial\theta_r} S_{EE} D_E^H\right\} P + D_E S_{EE} D_E^H \frac{\partial P}{\partial\theta_r}\right) \qquad \text{(A.20)}
$$

$$
\frac{\partial V_{\psi\psi}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\left(2\,\mathrm{herm}\left\{\frac{\partial L_\psi}{\partial\theta_r} S_{\psi\psi} L_\psi^H\right\} P + L_\psi S_{\psi\psi} L_\psi^H \frac{\partial P}{\partial\theta_r}\right) \qquad \text{(A.21)}
$$

$$
\begin{aligned}
\frac{\partial V_{EU}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\Bigg(&2\,\mathrm{herm}\left\{\frac{\partial D_E}{\partial\theta_r} S_{EU} D_U^H + D_E S_{EU}\frac{\partial D_U^H}{\partial\theta_r}\right\} P\\
&+2\,\mathrm{herm}\left\{D_E S_{EU} D_U^H\right\}\frac{\partial P}{\partial\theta_r}\Bigg) \qquad \text{(A.22)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial V_{\psi U}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\Bigg(&2\,\mathrm{herm}\left\{\frac{\partial L_\psi}{\partial\theta_r} S_{\psi U} D_U^H + L_\psi S_{\psi U}\frac{\partial D_U^H}{\partial\theta_r}\right\} P\\
&+2\,\mathrm{herm}\left\{L_\psi S_{\psi U} D_U^H\right\}\frac{\partial P}{\partial\theta_r}\Bigg) \qquad \text{(A.23)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial V_{\psi E}}{\partial\theta_r} = \frac{1}{N}\operatorname{tr}\Bigg(&2\,\mathrm{herm}\left\{\frac{\partial L_\psi}{\partial\theta_r} S_{\psi E} D_E^H + L_\psi S_{\psi E}\frac{\partial D_E^H}{\partial\theta_r}\right\} P\\
&+2\,\mathrm{herm}\left\{L_\psi S_{\psi E} D_E^H\right\}\frac{\partial P}{\partial\theta_r}\Bigg) \qquad \text{(A.24)}
\end{aligned}
$$

$$
\text{with}\quad \frac{\partial P}{\partial\theta_r} = -2\,\mathrm{herm}\left\{P\frac{\partial L_P}{\partial\theta_r}\left(L_P^H L_P\right)^{-1} L_P^H\right\} \qquad \text{(A.25)}
$$

### *A.4 Modifications for ARMAX*

Because $D = A$ for ARMAX, (A.1) is replaced by

$$
\sum_{m=0}^{M-1} z_k^{-K_m}\left\{I_m\left(z_k^{-1},\,\psi\right) + J_m\left(z_k^{-1},\,\psi\right)\right\} = L_{IJ_k}\psi_{IJ} \qquad \text{(A.26)}
$$

$$
\begin{aligned}
\text{with}\quad \psi_{IJ} &= \left[\psi_{IJ\,0}^T, \cdots, \psi_{IJ\,M-1}^T\right]^T\\
L_{IJ_k} &= \left[\eta_{IJ_k} z_k^{-K_0}, \cdots, \eta_{IJ_k} z_k^{-K_{M-1}}\right] \qquad \text{(A.27)}\\
\eta_{IJ_k} &= \left[1,\, z_k^{-1}, \cdots,\, z_k^{-n_{IJ}}\right];\ n_{IJ} = \max(n_I, n_J)
\end{aligned}
$$

Hence, the definitions of $\psi$, $L_P$, $\psi_0$ and $L_\psi$ in (A.5), (A.6) and (A.13) become

$$
\begin{aligned}
\psi = \psi_{IJ};&\quad \psi_0 = \psi_{0IJ}\\
L_P = D_{PJ} L_{IJ};&\quad L_\psi = D_{\psi J} L_{IJ}\\
\text{with}\quad L_{IJ} &= \left[L_{IJ_0}^T, \cdots, L_{IJ_{N-1}}^T\right]^T \qquad \text{(A.28)}
\end{aligned}
$$

## Appendix B Expected values in Subsection 3.1

Given the definitions in (16), the expectations $S_{EE}$, $S_{EU}$ and $S_{\psi E}$ are computed.

### *B.1 Computation of $S_{EE}$ and $S_{EU}$*

Under Assumptions 2 and 5

$$
S_{EE} = \sigma_e^2 I_N \qquad \text{(B.1)}
$$

$$
S_{EU} = 0_N \qquad \text{(B.2)}
$$

### *B.2 Computation of $S_{\psi E}$*

Given (A.13), for the BJ model structure the expectation $S_{\psi E}$ is structured as

$$
S_{\psi E} = \begin{bmatrix} S_{\psi EI}\\ S_{\psi EJ}\end{bmatrix} \qquad \text{(B.3)}
$$

$$
\text{with}\quad S_{\psi EI} = \mathbb{E}\left\{\psi_{0I} E^H\right\};\ S_{\psi EJ} = \mathbb{E}\left\{\psi_{0J} E^H\right\} \qquad \text{(B.4)}
$$

Under Assumption 5, $\psi_{0I}$ is function of $u(t)$ and $y_0(t)$ (true output), so that

$$
S_{\psi EI} = 0_{M(n_I+1)\times N} \qquad \text{(B.5)}
$$

In contrast, $\psi_{0J}$ is function of $e(t)$ and $v(t)$ (filtered noise). Thus, first the definition of $\psi_{0J}$ is presented. Next, $S_{\psi EJ}$ is computed.

#### *B.2.1 Definition of $\psi_{0J}$*

For a single data record, the definition of the true transient polynomial $J_{m\,0}\left(z^{-1}\right)$ is provided in Appendix 6.B of [8]. Extending it for $M$ records gives

$$
\begin{aligned}
J_{m\,0}(z_k^{-1}) = \Bigg\{&\sum_{i=1}^{n_C}\sum_{\mathsf{t}=1}^{i} c_i\left[e_m(-\mathsf{t}) - e_{m-1}(N_{m-1} - \mathsf{t})\right] z_k^{\mathsf{t}-i}\\
&-\sum_{\tilde{i}=1}^{n_D}\sum_{\tilde{\mathsf{t}}=1}^{\tilde{i}} d_{\tilde{i}}\left[v_m(-\tilde{\mathsf{t}}) - v_{m-1}(N_{m-1} - \tilde{\mathsf{t}})\right] z_k^{\tilde{\mathsf{t}}-\tilde{i}}\Bigg\}\frac{1}{\sqrt{N}}
\end{aligned}
$$

with $c_i$ and $d_{\tilde{i}}$ the $i$th and $\tilde{i}$th coefficients of $C_0\left(z^{-1}\right)$ and $D_0\left(z^{-1}\right)$, $e_m(-\mathsf{t})$ and $v_m(-\mathsf{t})$ the initial conditions of record $m$th, and $e_{m-1}\left(N_{m-1}-\mathsf{t}\right)$ and $v_{m-1}\left(N_{m-1}-\mathsf{t}\right)$ the final conditions of record $(m-1)$th. To obtain $\psi_{0J}$ explicitly, the previous equation is rewritten as [$r=i-\mathsf{t}$; $\tilde{r}=\tilde{i}-\tilde{\mathsf{t}}$; $x_{m-1}\left(N_{m-1}-\mathsf{t}\right)=x_c\left(K_m-\mathsf{t}\right)$]

$$J_{m\,0}\left(z_k^{-1}\right)=\left\{\sum_{r=0}^{n_C-1} z_k^{-r}\sum_{\mathsf{t}=1}^{n_C-r} c_{r+\mathsf{t}}\left[e_m\left(-\mathsf{t}\right)-e_c\left(K_m-\mathsf{t}\right)\right]\right.$$
$$\left.-\sum_{\tilde{r}=0}^{n_D-1} z_k^{-\tilde{r}}\sum_{\tilde{\mathsf{t}}=1}^{n_D-\tilde{r}} d_{\tilde{r}+\tilde{\mathsf{t}}}\left[v_m\left(-\tilde{\mathsf{t}}\right)-v_c\left(K_m-\tilde{\mathsf{t}}\right)\right]\right\}\frac{1}{\sqrt{N}}$$

Hence, $J_{m\,0}\left(z_k^{-1}\right)$ can be expressed as

$$J_{m\,0}\left(z_k^{-1}\right)=\eta_{J_k}\left(\frac{1}{\sqrt{N}}T_e\chi_{e\,m}-\frac{1}{\sqrt{N}}T_v\chi_{v\,m}\right) \quad \text{(B.6)}$$

with $\eta_{J_k}$ given by (A.4), $\chi_{e\,m}$ and $\chi_{v\,m}\in\mathbb{R}^{(n_J+1)\times 1}$ vectors of the initial/final conditions of $e_c\left(t\right)$ and $v_c\left(t\right)$. $T_e$ and $T_v\in\mathbb{R}^{(n_J+1)\times(n_J+1)}$ are upper-triangular matrices with coefficients of $C_0\left(z^{-1}\right)$ and $D_0\left(z^{-1}\right)$ respectively. $\chi_{e\,m}$ and $T_e$ are presented below ($\chi_{v\,m}$ and $T_v$ have the same structure)

$$\chi_{e\,m}=\begin{bmatrix} e_m\left(-\left(n_J+1\right)\right)-e_c\left(K_m-\left(n_J+1\right)\right)\\ \vdots\\ e_m\left(-1\right)-e_c\left(K_m-1\right)\end{bmatrix} \quad \text{(B.7)}$$

$$T_e=\begin{bmatrix} c_{n_J+1} & c_{n_J} & \cdots & c_1\\ 0 & c_{n_J+1} & \cdots & c_2\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & c_{n_J+1}\end{bmatrix} \quad \text{(B.8)}$$

Finally, $\psi_{0J}$ can be expressed as

$$\psi_{0J}=\frac{1}{\sqrt{N}}\left\{Q_e\chi_e-Q_v\chi_v\right\} \quad \text{(B.9)}$$

with $\chi_e$ and $\chi_v\in\mathbb{R}^{M(n_J+1)\times 1}$ vectors that group $\chi_{e\,m}$ and $\chi_{v\,m}$ for all $M$ records. $Q_e$ and $Q_v\in\mathbb{R}^{M(n_J+1)\times M(n_J+1)}$ are block diagonal matrices of $T_e$ and $T_v$ respectively. $\chi_e$ and $Q_e$ are presented below ($\chi_v$ and $Q_v$ have the same structure)

$$Q_e=I_M\otimes T_e;\quad \chi_e=\left[\chi_{e\,0}^T,\cdots,\chi_{e\,M-1}^T\right]^T \quad \text{(B.10)}$$

#### *B.2.2 Computation of $S_{\psi EJ}$*

Given (B.9) and (B.10), $S_{\psi EJ}$ has two components

$$S_{\psi EJ}=Q_e\mathbb{E}\left\{\frac{\chi_e}{\sqrt{N}}E_c^H\right\}-Q_v\mathbb{E}\left\{\frac{\chi_v}{\sqrt{N}}E_c^H\right\} \quad \text{(B.11)}$$

First, we compute the expectation of $\overline{E_c}\left(k\right)$ with the elements of $\chi_{e\,m}$ and $\chi_{v\,m}$ renamed as $e_{\tilde{c}}\left(K_m-\mathsf{t}\right)$ and $v_{\tilde{c}}\left(K_m-\mathsf{t}\right)$ to highlight the fact that $e_{\tilde{c}}\left(t\right)=e_c\left(t\right)$ only for the samples corresponding to the record $(m-1)$th [see (B.7)].

Since $v_{\tilde{c}}\left(t\right)=\frac{1}{\sqrt{N}}\sum_{\tilde{\tau}=-\infty}^{t}h\left(t-\tilde{\tau}\right)e_{\tilde{c}}\left(\tilde{\tau}\right)$ with $h\left(t\right)$ the impulse response of $H_0\left(z^{-1}\right)$, $\overline{E_c}\left(k\right)=\frac{1}{\sqrt{N}}\sum_{\tau=0}^{N-1}e_c\left(\tau\right)z_k^{\tau}$, and $\mathbb{E}\left\{e_{\tilde{c}}\left(\tilde{\tau}\right)\,e_c\left(\tau\right)\right\}=\sigma_e^2\delta\left(\tilde{\tau}-\tau\right)$ for $\tau\in\left[K_{m-1},K_m-1\right]$

$$\mathbb{E}\left\{\frac{e_{\tilde{c}}\left(K_m-\mathsf{t}\right)\overline{E_c}\left(k\right)}{\sqrt{N}}\right\}=\frac{\sigma_e^2}{N}z_k^{K_m-\mathsf{t}}$$
$$\mathbb{E}\left\{\frac{v_{\tilde{c}}\left(K_m-\mathsf{t}\right)\overline{E_c}\left(k\right)}{\sqrt{N}}\right\}=\frac{\sigma_e^2}{N}\sum_{\tilde{\tau}=K_{m-1}}^{K_m-\mathsf{t}}z_k^{\tilde{\tau}}\,h\left(K_m-\mathsf{t}-\tilde{\tau}\right)$$
$$=\frac{\sigma_e^2}{N}z_k^{K_m}\,\varSigma_{h\,m}\left(\mathsf{t}\right)$$
$$\text{with}\quad \varSigma_{h\,m}\left(\mathsf{t}\right)=z_k^{-\mathsf{t}}\sum_{T=0}^{N_{m-1}-\mathsf{t}}z_k^{-T}\,h\left(T\right) \quad \text{(B.12)}$$

Finally, the expectation of $\overline{E_c}\left(k\right)$ with $\chi_{e\,m}$ and $\chi_{v\,m}$ is [with $\eta_{J_k}$ given by (A.4)]

$$\mathbb{E}\left\{\frac{\chi_{e\,m}}{\sqrt{N}}\overline{E_c}\left(k\right)\right\}=-\frac{\sigma_e^2}{N}z_k^{K_m}\,z_k^{-(n_J+1)}\eta_{J_k}^H \quad \text{(B.13)}$$

$$\mathbb{E}\left\{\frac{\chi_{v\,m}}{\sqrt{N}}\overline{E_c}\left(k\right)\right\}=-\frac{\sigma_e^2}{N}z_k^{K_m}\begin{bmatrix}\varSigma_{h\,m}\left(n_J+1\right)\\ \vdots\\ \varSigma_{h\,m}\left(1\right)\end{bmatrix} \quad \text{(B.14)}$$

### *B.3 Remark for ARMAX*

By definition $L_{IJ}\psi_{0IJ}=L_I\psi_{0I}+L_J\psi_{0J}$ [see (A.1) and (A.26)]. Hence,

$$L_{IJ}S_{\psi E}=L_IS_{\psi EI}+L_JS_{\psi EJ}$$

with $S_{\psi EI}$ and $S_{\psi EJ}$ as defined in (B.4). Thus, the equations of Appendices B.2.1 and B.2.2 apply to the ARMAX model structure.

## Appendix C Preliminary Lemmas

Here, $H_a\left(z^{-1}\right)$ and $H_b\left(z^{-1}\right)$ are rational transfer functions with impulse responses $h_a\left(t\right)$ and $h_b\left(t\right)$.

**Lemma 21.** *Given that $z_k=e^{j2\pi k/N}$*

$$\sum_{k=0}^{N-1}z_k^{\alpha}=\begin{cases}0 & \text{for } \alpha\neq nN\\ N & \text{for } \alpha=nN\end{cases}\quad \text{with } n\in\mathbb{Z}$$

*Proof.* If $\alpha=nN$, $z_k^{\alpha}=1$. If $\alpha\neq nN$, $z_k^{\alpha}=z_{\alpha}^k$ and $\sum_{k=0}^{N-1}z_{\alpha}^k=\left(1-z_{\alpha}^N\right)/\left(1-z_{\alpha}\right)=0$ ☐

**Lemma 22.** *For $H_a(z^{-1})$ stable, $\alpha \in \mathbb{Z}$, $|\alpha| < N-1$, and $N \to \infty$*

$$\frac{1}{N}\sum_{k=0}^{N-1} H_a\left(z_k^{-1}\right) z_k^{\alpha} = O\left(|\lambda|^{N\varphi+\alpha}\right)$$

*with $\lambda$ the pole of $H_a(z^{-1})$ that is the closest to the unit circle ($|\lambda| < 1$) and $\varphi = \lceil \frac{-\alpha}{N} \rceil$.*

*Proof.* $H_a\left(z_k^{-1}\right) = \sum_{i=0}^{\infty} h_a(i) z_k^{-i}$. By Lemma 21

$$\frac{1}{N}\sum_{i=0}^{\infty} h_a(i) \sum_{k=0}^{N-1} z_k^{\alpha-i} = \sum_{n=\varphi}^{\infty} h_a(nN+\alpha)$$

with $\varphi = \lceil \frac{-\alpha}{N} \rceil$. If $H_a(z^{-1})$ is stable, its impulse response $h_a(t)$ can be bounded by $c_\lambda |\lambda|^t$ (with $c_\lambda \in \mathbb{R}_0^+$). Applying the result of geometric series and $N \to \infty$

$$\left|\sum_{n=\varphi}^{\infty} h_a(nN+\alpha)\right| \leq c_\lambda \sum_{n=\varphi}^{\infty} |\lambda|^{nN+\alpha} = O\left(|\lambda|^{N\varphi+\alpha}\right)$$

□

**Lemma 23.** *For $H_a(z^{-1})$ and $H_b(z^{-1})$ stable, $\alpha \in \mathbb{Z}$, $|\alpha| < N-1$, and $N \to \infty$*

$$\frac{1}{N}\sum_{k=0}^{N-1} H_a\left(z_k^{-1}\right) H_b(z_k) z_k^{\alpha} = O\left(|\lambda|^{N-|\alpha|} + |\lambda|^{|\alpha|}\right)$$

*with $\lambda$ the pole of $H_a(z^{-1})$ and $H_b(z^{-1})$ that is the closest to the unit circle ($|\lambda| < 1$).*

*Proof.* $H_a\left(z_k^{-1}\right) = \sum_{i=0}^{\infty} h_a(i) z_k^{-i}$, $H_b(z_k) = \sum_{l=0}^{\infty} h_b(l) z_k^{l}$. By Lemma 21

$$\frac{1}{N}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} h_a(i) h_b(l) \sum_{k=0}^{N-1} z_k^{-i+l+\alpha} = \sum_{i=0}^{\infty}\sum_{n=\lceil \frac{\alpha-i}{N} \rceil}^{\infty} h_a(i) h_b(nN+i-\alpha)$$

If $H_a(z^{-1})$ and $H_b(z^{-1})$ are stable, their impulse responses $h_a(t)$ and $h_b(t)$ can be bounded by $c_\lambda |\lambda|^t$ (with $c_\lambda \in \mathbb{R}_0^+$).

$$\left|\sum_{i=0}^{\infty}\sum_{n=\lceil \frac{\alpha-i}{N} \rceil}^{\infty} h_a(i) h_b(nN+i-\alpha)\right| \leq c_\lambda^2 \sum_{i=0}^{\infty}\sum_{n=\lceil \frac{\alpha-i}{N} \rceil}^{\infty} |\lambda|^{nN+2i-\alpha}$$

First, we split the summation over $i$ as: $i \in [0, N+\alpha-1]$ and $i \in [N+\alpha, \infty]$. Next, for $i \in [N+\alpha, \infty]$, we change variables as: $i = sN + q$ (with $s \in [1, \infty]$ and $q \in [\alpha, N+\alpha-1]$, so that $\lceil \frac{\alpha-i}{N} \rceil = -s$)

$$c_\lambda^2 \left[\sum_{i=0}^{N+\alpha-1}\sum_{n=\lceil \frac{\alpha-i}{N} \rceil}^{\infty} |\lambda|^{nN+2i-\alpha} + \sum_{s=1}^{\infty}\sum_{q=\alpha}^{N+\alpha-1}\sum_{n=-s}^{\infty} |\lambda|^{nN+2sN+2q-\alpha}\right]$$

After applying the result of geometric series and $N \to \infty$, we obtain the following results (with $\varphi = \lceil \frac{\alpha}{N} \rceil$)

$$c_\lambda^2 \left[\left(\frac{|\lambda|^{N\varphi-\alpha} - |\lambda|^{N\varphi+\alpha} + |\lambda|^{\alpha}}{1-|\lambda|^2}\right) + \left(\frac{|\lambda|^{N+\alpha}}{1-|\lambda|^2}\right)\right]$$

□

**Lemma 24.** *For $H_a(z^{-1})$ and $H_b(z^{-1})$ stable, $h_a(t)$ the impulse response of $H_a(z^{-1})$, $0 < c_a < N$, $-N < \alpha \leq -1$, and $N \to \infty$*

$$\frac{1}{N}\sum_{T=0}^{c_a} h_a(T) \sum_{k=0}^{N-1} H_b\left(z_k^{-1}\right) z_k^{\alpha-T} = O\left((c_a+1)|\lambda|^{N+\alpha}\right)$$

*with $\lambda$ the pole of $H_a(z^{-1})$ and $H_b(z^{-1})$ that is the closest to the unit circle ($|\lambda| < 1$) .*

*Proof.* $H_b\left(z_k^{-1}\right) = \sum_{i=0}^{\infty} h_b(i) z_k^{-i}$. By Lemma 21

$$\frac{1}{N}\sum_{T=0}^{c_a} h_a(T) \sum_{i=0}^{\infty} h_b(i) \sum_{k=0}^{N-1} z_k^{-i-T+\alpha} = \sum_{T=0}^{c_a} h_a(T) \sum_{n=\lceil \frac{T-\alpha}{N} \rceil}^{\infty} h_b(nN-T+\alpha)$$

If $H_a(z^{-1})$ and $H_b(z^{-1})$ are stable, their impulse responses $h_a(t)$ and $h_b(t)$ can be bounded by $c_\lambda |\lambda|^t$ (with $c_\lambda \in \mathbb{R}_0^+$). Besides, if $-N < \alpha \leq -1$ and $0 < c_a < N$, $1 = \lceil \frac{-\alpha}{N} \rceil \leq \lceil \frac{T-\alpha}{N} \rceil$, so that

$$\left|\sum_{T=0}^{c_a}\sum_{n=\lceil \frac{T-\alpha}{N} \rceil}^{\infty} h_a(T) h_b(nN-T+\alpha)\right| \leq c_\lambda^2 \sum_{T=0}^{c_a}\sum_{n=\lceil \frac{T-\alpha}{N} \rceil}^{\infty} |\lambda|^{nN+\alpha} \leq c_\lambda^2 \sum_{T=0}^{c_a}\sum_{n=1}^{\infty} |\lambda|^{nN+\alpha}$$

Applying the result of geometric series and $N \to \infty$

$$c_\lambda^2 \sum_{T=0}^{c_a}\sum_{n=1}^{\infty} |\lambda|^{nN+\alpha} = c_\lambda^2 \sum_{T=0}^{c_a} \frac{|\lambda|^{N+\alpha}}{1-|\lambda|^N} = O\left(c_\lambda^2 (c_a+1)|\lambda|^{N+\alpha}\right)$$

□

**Lemma 25.** *For $H_a(z^{-1})$ stable and $h_a(t)$ its the impulse response, $0 < c_a < N$, $|\alpha| < N-1$, and $N \to \infty$*

$$\frac{1}{N}\sum_{T=0}^{c_a} h_a(T) \sum_{k=0}^{N-1} z_k^{\alpha-T} = \begin{cases} O\left(|\lambda|^{\varphi N+\alpha}\right) & \text{if } 0 \leq \varphi N+\alpha \leq c_a \\ 0 & \text{otherwise} \end{cases}$$

*with $\lambda$ the pole of $H_a(z^{-1})$ that is the closest to the unit circle ($|\lambda| < 1$) and $\varphi = \lceil \frac{-\alpha}{N} \rceil$.*

*Proof.* By Lemma 21, and given $|\alpha| < N-1$ and $0 < c_a < N$

$$\frac{1}{N}\sum_{T=0}^{c_a} h_a(T) \sum_{k=0}^{N-1} z_k^{\alpha-T} = \sum_{T=0}^{c_a} h_a(T)\, \delta(\varphi N - T + \alpha)$$

with $\varphi = \left\lceil \frac{-\alpha}{N} \right\rceil$. If $H_a\left(z^{-1}\right)$ is stable, its impulse response $h_a(t)$ can be bounded by $c_\lambda |\lambda|^t$ (with $c_\lambda \in \mathbb{R}_0^+$). If $0 \le \varphi N + \alpha \le c_a$ (otherwise $\delta(\varphi N - T + \alpha) = 0$)

$$\left|\sum_{T=0}^{c_a} h_a(T)\delta(nN - T + \alpha)\right| \le c_\lambda \sum_{T=0}^{c_a} |\lambda|^T \delta(\varphi N - T + \alpha) = O(|\lambda|^{\varphi N + \alpha})$$

□

**Lemma 26.** *For $D_{H_a} = \operatorname{diag}\left(\cdots, H_a\left(z_k^{-1}\right), \cdots\right)$ and $D_{H_b} = \operatorname{diag}\left(\cdots, H_b\left(z_k^{-1}\right), \cdots\right)$, with $H_a\left(z^{-1}\right)$ and $H_b\left(z^{-1}\right)$ stable*

$$\frac{1}{N} L_X^H D_{H_a}^H D_{H_b} L_Y$$

*with $L_X$ and $L_Y$ equal to $L_I$ or $L_J$ as defined by (A.8), is a block diagonal matrix ($M \times M$ blocks) for record lengths $N_m \to \infty$. The diagonal blocks are of order $O(|\lambda|^{|\alpha|})$ with $\alpha \approx 0$ and $\lambda$ the pole of $H_a\left(z^{-1}\right)$ and $H_b\left(z^{-1}\right)$ that is the closest to the unit circle ($|\lambda| < 1$).*

*Proof.* The $\langle p, q \rangle$ block of $\frac{1}{N} L_I^H D_{H_a}^H D_{H_b} L_J$ ($p, q \in [0, M-1]$) is [see (A.2) to (A.4)]

$$\left(\frac{1}{N} L_I^H D_{H_a}^H D_{H_b} L_J\right)_{\langle p,q \rangle} = \frac{1}{N} \sum_{k=0}^{N-1} H_a(z_k) H_b\left(z_k^{-1}\right) z_k^{K_p - K_q} \eta_{I_k}^H \eta_{J_k} \tag{C.1}$$

By Lemma 23, (C.1) is an $O(|\lambda|^{N-|\alpha|} + |\lambda|^{|\alpha|})$, with $\alpha \approx K_p - K_q$. For $p = q$, $\alpha \approx 0$. For $p \neq q$, $N_m \le |\alpha| \le N - N_m$. Thus, $\frac{1}{N} L_I^H D_{H_a}^H D_{H_b} L_J$ is a block diagonal matrix for $N_m \to \infty$. □

**Lemma 27.** *With $L_P$ defined by (A.6),*

$$R = \left(\frac{1}{N} L_P^H L_P\right)^{-1} = \begin{bmatrix} R_A & R_B \\ R_B^H & R_C \end{bmatrix}$$

*For record lengths $N_m \to \infty$, $R_A$, $R_B$ and $R_C$ are block diagonal matrices with diagonal blocks $R_{A\,m}$, $R_{B\,m}$ and $R_{C\,m}$ ($m \in [0, M-1]$) of order $O\left(N^0\right)$.*

*Proof.*

$$Q = \frac{1}{N} L_P^H L_P = \begin{bmatrix} Q_A & Q_B \\ Q_B^H & Q_C \end{bmatrix}$$

By Lemma 26, $Q_A$, $Q_B$ and $Q_C$ are block diagonal matrices with diagonal blocks $Q_{A\,m}$, $Q_{B\,m}$ and $Q_{C\,m}$ of order $O\left(N^0\right)$ for $N_m \to \infty$. $Q$ can be inverted blockwise as

$$R_A = Q_A^{-1} + Q_A^{-1} Q_B \left(Q_C - Q_B^H Q_A^{-1} Q_B\right)^{-1} Q_B^H Q_A^{-1}$$
$$R_B = -Q_A^{-1} Q_B \left(Q_C - Q_B^H Q_A^{-1} Q_B\right)^{-1}$$
$$R_C = \left(Q_C - Q_B^H Q_A^{-1} Q_B\right)^{-1}$$

resulting in $R_A$, $R_B$ and $R_C$ as block diagonal matrices with diagonal blocks of order $O\left(N^0\right)$ for $N_m \to \infty$. □

**Lemma 28.** *For the OE model structure, and all $M$ records of equal length $N_r$ ($N = MN_r$), $Q = \frac{1}{N} L_P^H L_P$, with $L_P$ defined by (A.6), is a Hermitian block Toeplitz matrix. Hence, $R = Q^{-1}$ is given by [14]*

$$R = T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H - T_{\mathsf{L}}^H \triangle_{\mathsf{L}}^{-1} T_{\mathsf{L}}$$

$$\text{with} \quad T_{\mathsf{M}} = \begin{bmatrix} \mathsf{M}_{M-1} & \mathsf{M}_{M-2} & \cdots & \mathsf{M}_0 \\ 0 & \mathsf{M}_{M-1} & \cdots & \mathsf{M}_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathsf{M}_{M-1} \end{bmatrix} \tag{C.2}$$

$$T_{\mathsf{L}} = \begin{bmatrix} \mathsf{L}_0 & 0 & \cdots & 0 \\ \mathsf{L}_1 & \mathsf{L}_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \mathsf{L}_{M-1} & \mathsf{L}_{M-2} & \cdots & \mathsf{L}_0 \end{bmatrix} \tag{C.3}$$

$$\triangle_{\mathsf{M}} = \operatorname{diag}\left(\triangle_{\mathsf{M}\,r} \cdots \triangle_{\mathsf{M}\,r}\right); \ \triangle_{\mathsf{L}} = \operatorname{diag}\left(\triangle_{\mathsf{L}\,r} \cdots \triangle_{\mathsf{L}\,r}\right) \tag{C.4}$$

*For $M \to \infty$, $T_{\mathsf{M}}$ and $T_{\mathsf{L}}$ are sparse matrices: blocks $\mathsf{M}_p$ and $\mathsf{L}_p$ are only non-zero for $p$ belonging to*

$$[0, \cdots, \mathsf{F} - 1, M - \mathsf{F}, \cdots, M - 1] \tag{C.5}$$

*with $\mathsf{F}$ finite and determined by the slowest pole of $G_0\left(z^{-1}\right)$.*

*Proof.* For OE, $L_P = D_{PI} L_I$ [see (A.9), (A.8), (A.2) and (A.4)]. The $\langle p, q \rangle$ block of $Q$ ($p, q \in [0, M-1]$) is given by

$$Q_{\langle p,q \rangle} = \frac{1}{N} \sum_{k=0}^{N-1} H_{PI}(z_k) H_{PI}\left(z_k^{-1}\right) z_k^{K_p - K_q} \eta_{I_k}^H \eta_{I_k}$$

with $H_{PI}\left(z_k^{-1}\right)$ the $k$th diagonal element of $D_{PI}$. For equal record lengths $K_p - K_q = N_r(p - q)$. Hence, $Q$ is structured as

$$Q_{\langle p,q \rangle} = \mathsf{Q}_i \in \mathbb{C}^{\left(n_I+1\right) \times \left(n_I+1\right)}$$

with $i = p - q$, $\mathsf{Q}_{-i} = \mathsf{Q}_i^H$ and $\mathsf{Q}_i$ Toeplitz. By Lemma 23, $\mathsf{Q}_i = O(|\lambda|^{N-|\alpha|} + |\lambda|^{|\alpha|})$ with $\alpha \approx K_p - K_q$ and $\lambda$ the pole of $H_{PI}\left(z_k^{-1}\right)$ that is the closest to the unit circle ($|\lambda| < 1$). Thus, $Q$ is sparse for $M \to \infty$

$$\begin{cases} \mathsf{Q}_i \neq 0 & \text{for } i = \left[0, \cdots, \tilde{\mathsf{F}}, M-\tilde{\mathsf{F}}, \cdots, M-1\right] \\ \mathsf{Q}_i = 0 & \text{otherwise} \end{cases} \tag{C.6}$$

with $\tilde{\mathsf{F}}$ finite and determined by $|\lambda|$. The structure of $T_{\mathsf{M}}$ and $T_{\mathsf{L}}$ reflects the sparse nature of $Q$ as follows. Being $\tilde{Q}$ the leading principal submatrix of $Q$ of order $(M-1)(n_I+1) \times (M-1)(n_I+1)$, the blocks $\mathsf{M}_p$ and $\mathsf{L}_p$ on (C.2) and (C.3) are defined by [14]

$$-\left[\mathsf{M}_0^H\, \mathsf{M}_1^H \cdots \mathsf{M}_{M-2}^H\right] \tilde{Q}_l = \mathsf{Q}_{M-1-l} \tag{C.7}$$

$$-\left[\mathsf{L}_{M-1} \cdots \mathsf{L}_2\, \mathsf{L}_1\right] \tilde{Q}_l = \mathsf{Q}_{l+1}^H \tag{C.8}$$

with $\tilde{Q}_l = \left[\tilde{Q}^T_{\langle 0,l\rangle} \cdots \tilde{Q}^T_{\langle M-2,l\rangle}\right]^T$, $l \in [0, M-2]$, $\mathsf{M}_{M-1} = I$ and $\mathsf{L}_0 = 0$.

By (C.6) and for $l \in \left[3\tilde{\mathsf{F}}, M-2-3\tilde{\mathsf{F}}\right]$, (C.7) results in the following set of equations

$$\left[\mathsf{M}_{\mathsf{X}}^H \cdots \mathsf{M}_{\mathsf{X}+2\tilde{\mathsf{F}}}^H\right]\left[\mathsf{Q}_{-\tilde{\mathsf{F}}}^T \cdots \mathsf{Q}_{\tilde{\mathsf{F}}}^T\right]^T = \mathsf{Q}_{M-\tilde{\mathsf{F}}-\mathsf{X}-1}$$

with $\mathsf{X} \in \left[2\tilde{\mathsf{F}}, M-4\tilde{\mathsf{F}}-2\right]$, so that $\mathsf{Q}_{M-\tilde{\mathsf{F}}-\mathsf{X}-1} = 0$. Because $\mathsf{Q}_{-\tilde{\mathsf{F}}}$ to $\mathsf{Q}_{\tilde{\mathsf{F}}}$ are of full rank, and $\mathsf{M}_{\mathsf{X}}^H$ to $\mathsf{M}_{\mathsf{X}+2\tilde{\mathsf{F}}}^H$ only appear in this set of equations, we can conclude that $\mathsf{M}_p = 0$ for $p \in \left[2\tilde{\mathsf{F}}, M-2\tilde{\mathsf{F}}-2\right]$. So we choose $\mathsf{F} = 2\tilde{\mathsf{F}}+1$ for (C.5). A similar proof can be derived for $\mathsf{L}_p$. □

## Appendix D Proof of Lemma 2

Given (13), $\|\mathsf{W}_N\|_1$ can be bounded by

$$\|\mathsf{W}_N\|_1 \le \|W^H\|_1 \|W\|_1 \|P\|_1$$

with $\|W^H\|_1$ and $\|W\|_1$ of $O(N^0)$ according to (A.7), (A.11) and (A.12). From (14) and the definition of the 1-norm

$$\|P\|_1 \le 1 + \left\|\frac{L_P R L_P^H}{N}\right\|_1; \text{ with } R = \left(\frac{1}{N} L_P^H L_P\right)^{-1} \tag{D.1}$$

$$\left\|\frac{L_P R L_P^H}{N}\right\|_1 = \max_{l=0,\cdots,N-1} \sum_{k=0}^{N-1} \left|\left[\frac{L_P R L_P^H}{N}\right]_{k,l}\right| \tag{D.2}$$

The proof that (D.2) is $O(N^0)$ will be shown below for BJ and OE. For ARMAX the proof is similar than for OE because $L_P = D_{PI} L_I$ for OE and $L_P = D_{PJ} L_{IJ}$ for ARMAX [see (A.6) and Appendix A.4]. According to Theorem 14, BJ is consistent only for $N_m \to \infty$. Hence, for BJ it is sufficient to prove the case of long records.

### *D.1 $\|\mathsf{W}_N\|_1$ for Box-Jenkins*

For the BJ model structure, $L_P$ is a block matrix defined by (A.6), (A.8), (A.9) and (A.10). Hence, $R$ in (D.1) is a block matrix structured as

$$R = \begin{bmatrix} R_A & R_B \\ R_B^H & R_C \end{bmatrix}$$

so that

$$\left\|\frac{L_P R L_P^H}{N}\right\|_1 \le \left\|\frac{D_{PI} L_I R_A L_I^H D_{PI}^H}{N}\right\|_1 + \left\|\frac{D_{PJ} L_J R_C L_J^H D_{PJ}^H}{N}\right\|_1 + \left\|\frac{D_{PJ} L_J R_B^H L_I^H D_{PI}^H}{N}\right\|_1 + \left\|\frac{D_{PI} L_I R_B L_J^H D_{PJ}^H}{N}\right\|_1 \tag{D.3}$$

By Lemma 27 (in Appendix C), for sufficiently large record lengths $N_m$, $R_A$, $R_B$, $R_C$ are block diagonal matrices with diagonal blocks $R_{A\,m}$, $R_{B\,m}$ and $R_{C\,m}$ of order $O(N^0)$ ($m \in [0, M-1]$). Assuming $M$ records of equal length $N_r$ ($N = MN_r$, and $R_{A\,m} = R_{A\,r}$, $R_{B\,m} = R_{B\,r}$, $R_{C\,m} = R_{C\,r}$ for all $m$), the first term of (D.3) gives

$$\sum_{k=0}^{N-1} \left|\left[\frac{D_{PI} L_I R_A L_I^H D_{PI}^H}{N}\right]_{k,l}\right| = |H_{PI}(z_l)| \cdot \sum_{k=0}^{N-1} \left|\frac{1}{N} H_{PI}\left(z_k^{-1}\right) \eta_{I_k} R_{A\,r} \eta_{I_l}^H \sum_{m=0}^{M-1} z_{l-k}^{K_m}\right|$$

where $H_{PI}\left(z_k^{-1}\right)$ is the $k$th diagonal element of $D_{PI}$ and

$$\sum_{m=0}^{M-1} z_{l-k}^{K_m} = \sum_{m=0}^{M-1} e^{j2\pi m N_r (l-k)/N} = M\,\delta\left(\{k-l\} - nM\right) \tag{D.4}$$

with $n \in [0,\, N_r - 1]$. Hence,

$$\sum_{k=0}^{N-1} \left|\left[\frac{D_{PI} L_I R_A L_I^H D_{PI}^H}{N}\right]_{k,l}\right| = \frac{MN_r}{N} O\left(N^0\right) = O\left(N^0\right)$$

and the same result can be derived for the other terms of (D.3).

### *D.2 $\|\mathsf{W}_N\|_1$ for OE*

For the OE model structure, $L_P = D_{PI} L_I$ [see (A.8) and (A.9)]. Assuming $M$ records of equal length $N_r$ ($N = MN_r$), Lemma 28 (in Appendix C) applies for $R$ in (D.1), so that

$$\left\|\frac{L_P R L_P^H}{N}\right\|_1 \le \left\|\frac{L_P T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H L_P^H}{N}\right\|_1 + \left\|\frac{L_P T_{\mathsf{L}}^H \triangle_{\mathsf{L}}^{-1} T_{\mathsf{L}} L_P^H}{N}\right\|_1 \tag{D.5}$$

with $T_{\mathsf{M}}$, $T_{\mathsf{L}}$, $\triangle_{\mathsf{M}}$ and $\triangle_{\mathsf{L}}$, given by (C.2) to (C.4). Moreover, for $M \to \infty$, $T_{\mathsf{M}}$ and $T_{\mathsf{L}}$ are sparse matrices (blocks $\mathsf{M}_p$, $\mathsf{L}_p$ are only non-zero for $p$ belonging to (C.5), where $\mathsf{F}$ is finite).

We analyze the first term of (D.5), for which

$$\left[\tfrac{1}{N} L_P T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H L_P^H\right]_{k,l} = \tfrac{1}{N} L_{P\,k} T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H \left(L_{P\,l}\right)^H \tag{D.6}$$

with $L_{P\,k}$ and $L_{P\,l}$ the $k$th and $l$th row of $L_P$.

Being $T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H$ a product of block matrices, one can expand (D.6) and regroup the terms that have in common $\mathsf{M}_p \triangle_{\mathsf{M}\,r}^{-1} \mathsf{M}_q^H$. Thus, (D.6) can be expressed as the sum of $4\mathsf{F}^2$ terms that have this form

$$\tfrac{1}{N} H_{PI}(z_l)\, H_{PI}\left(z_k^{-1}\right) \eta_{I_k} \mathsf{M}_p \triangle_{\mathsf{M}\,r}^{-1} \mathsf{M}_q^H \eta_{I_l}^H \left( z_k^{-K_{\tilde{p}}} z_l^{K_{\tilde{q}}} \sum_{m=0}^{m_f} z_{l-k}^{K_m} \right) \tag{D.7}$$

with $p$, $q$, $\tilde{p}$, $\tilde{q}$ and $m_f$ belonging to (C.5). The contribution of (D.7) to $\left\| \frac{1}{N} L_P T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H L_P^H \right\|_1$ can be bounded above by

$$\left| H_{PI}\left(z_l\right) \right| \sum_{k=0}^{N-1} \left| \tfrac{1}{N} H_{PI}\left(z_k^{-1}\right) \eta_{I_k} \mathsf{M}_p \triangle_{\mathsf{M}\,r}^{-1} \mathsf{M}_q^H \eta_{I_l}^H \right| \left| \sum_{m=0}^{m_f} z_{l-k}^{K_m} \right|$$

where

$$\left| \sum_{m=0}^{m_f} z_{l-k}^{K_m} \right| = \left| \sum_{m=0}^{m_f} e^{j 2\pi m N_r (l-k)/N} \right| \leq \begin{cases} \mathsf{F} & \text{for } l-k \neq nM \\ M & \text{for } l-k = nM \end{cases}$$

with $n \in [0, \cdots, N_r - 1]$. Hence,

$$\left\| \frac{L_P T_{\mathsf{M}} \triangle_{\mathsf{M}}^{-1} T_{\mathsf{M}}^H L_P^H}{N} \right\|_1 = 4\mathsf{F}^2 O\left( \frac{M N_r}{N} + \frac{\mathsf{F}(N - N_r)}{N} \right) = O(N^0)$$

and the same result can be derived for $\left\| L_P T_{\mathsf{L}}^H \triangle_{\mathsf{L}}^{-1} T_{\mathsf{L}} L_P^H \right\|_1$.

## Appendix E Proof of Corollary 8

From the $M$ records of arbitrary but finite length $N_m$, one can form a finite number of sets $\mathcal{S}$ that group records of the same length. Then, $V_N(\theta, \psi, \mathsf{z_t})$ is expressed as a weighted sum of cost functions dealing with each set of records.

$$V_N(\theta, \psi, \mathsf{z_t}) = \sum_{s=1}^{\mathcal{S}} \frac{N_s}{N} V_{N_s}(\theta, \psi, \mathsf{z_t}) \tag{E.1}$$

with $N_s$ the number of samples in each set. Because $M \to \infty$, for at least one set $N_s \to \infty$. From (E.1) its clear that only sets where $N_s \to \infty$ contribute asymptotically to $V_N(\theta, \psi, \mathsf{z_t})$ [$\mathcal{S}$ is finite and $V_{N_s}(\theta, \mathsf{z_t})$ is finite for $\theta \in \Theta$]. Because the convergence applies to $V_{N_s}(\theta, \mathsf{z_t})$ for all sets where $N_s \to \infty$ (by Theorem 7), so does for $V_N(\theta, \mathsf{z_t})$.

## Appendix F Proof of Lemma 9

Given (A.19), (A.21) to (A.23), and the definition of $P$ and $\frac{\partial P}{\partial \theta_r}$ in (14) and (A.25)

$$\left.\frac{\partial V_{UU}}{\partial \theta_r}\right|_{\theta_0} = 0; \quad \left.\frac{\partial V_{\psi\psi}}{\partial \theta_r}\right|_{\theta_0} = 0; \quad \left.\frac{\partial V_{EU}}{\partial \theta_r}\right|_{\theta_0} = 0; \quad \left.\frac{\partial V_{\psi U}}{\partial \theta_r}\right|_{\theta_0} = 0$$

by the properties of the trace, (B.2), and the fact that $D_U|_{\theta_0} = 0_N$ [see (A.14)] and $L_\psi|_{\theta_0} = L_P|_{\theta_0}$ [see (A.6) and (A.13)] so that

$$L_\psi^H P|_{\theta_0} = P L_\psi|_{\theta_0} = 0_N; \quad L_\psi^H \frac{\partial P}{\partial \theta_r} L_\psi|_{\theta_0} = 0_N \tag{F.1}$$

## Appendix G Proof of Lemma 10

Equation (A.20) can be simplified to

$$\left.\frac{\partial V_{EE}}{\partial \theta_r}\right|_{\theta_0} = \frac{\sigma_e^2}{N} \operatorname{tr}\left( 2 \operatorname{herm}\left\{ \frac{\partial D_E}{\partial \theta_r} P \right\} \right)\Bigg|_{\theta_0} \tag{G.1}$$

because of (B.1), $D_E|_{\theta_0} = I_N$ [see (A.15)], and $\operatorname{tr}\left(\frac{\partial P}{\partial \theta_r}\right) = 0$ [see (A.25)]. This expression is evaluated below for the different model parameters

### G.1 $\partial V_{EE}/\partial a_r$ and $\partial V_{EE}/\partial b_r$

From (A.15), we get for parameters $a_r$ and $b_r$

$$\frac{\partial D_E}{\partial \theta_r} = 0_N \quad \to \quad \left.\frac{\partial V_{EE}}{\partial \theta_r}\right|_{\theta_0} = 0 \tag{G.2}$$

### G.2 $\partial V_{EE}/\partial c_r$ and $\partial V_{EE}/\partial d_r$

From (A.15), we get for parameters $c_r$ and $d_r$ ($1 \leq r \leq n_D$ or $1 \leq r \leq n_C$ )

$$\left.\frac{\partial D_E}{\partial c_r}\right|_{\theta_0} = D_r\, D_{Ec}; \quad \left.\frac{\partial D_E}{\partial d_r}\right|_{\theta_0} = D_r\, D_{Ed}$$

$$\text{with} \quad D_{Ec} = \operatorname{diag}\left(\cdots, -1/C_0\left(z_k^{-1}\right), \cdots\right)$$

$$D_{Ed} = \operatorname{diag}\left(\cdots, 1/D_0\left(z_k^{-1}\right), \cdots\right) \tag{G.3}$$

$$D_r = \operatorname{diag}\left(\cdots, z_k^{-r}, \cdots\right) \tag{G.4}$$

We evaluate (G.1) for $d_r$. Given the definition of the trace and (14)

$$\operatorname{tr}\left( \frac{\sigma_e^2}{N} D_r\, D_{Ed} P \right)\Bigg|_{\theta_0} = \bar{\mathcal{B}} - \frac{\sigma_e^2}{N} \operatorname{tr}\left(R \bar{\mathcal{Q}}\right) \tag{G.5}$$

$$\text{with} \quad \bar{\mathcal{B}} = \frac{\sigma_e^2}{N} \sum_{k=0}^{N-1} H_{Ed}\left(z_k^{-1}\right) z_k^{-r}$$

$$R = \left(\tfrac{1}{N} L_P^H L_P\right)^{-1}\Big|_{\theta_0}; \quad \bar{\mathcal{Q}} = \tfrac{1}{N} L_P^H D_r\, D_{Ed} L_P\Big|_{\theta_0};$$

where $H_{Ed}\left(z_k^{-1}\right)$ is the $k$th diagonal element of $D_{Ed}$.

By Lemma 22 in Appendix C (with $\alpha < 0$, $N + \alpha \approx N$), and with $\lambda_{Ed}$ the dominant pole of $H_{Ed}\left(z^{-1}\right)$ and $|\lambda_{Ed}| < 1$

$$\bar{\mathcal{B}} = O\left(\sigma_e^2 \left|\lambda_{Ed}\right|^N\right)$$

Because $L_P$ is a block matrix defined by (A.6), $R$ and $\bar{\mathcal{Q}}$ are block matrices structured as

$$R = \begin{bmatrix} R_A & R_B \\ R_B^H & R_C \end{bmatrix}; \quad \bar{\mathcal{Q}} = \begin{bmatrix} \bar{\mathcal{Q}}_A & \bar{\mathcal{Q}}_B \\ \bar{\mathcal{Q}}_C & \bar{\mathcal{Q}}_D \end{bmatrix};$$

so that

$$\operatorname{tr}\left(R \bar{\mathcal{Q}}\right) = \operatorname{tr}\left(R_A \bar{\mathcal{Q}}_A + R_B \bar{\mathcal{Q}}_C + R_B^H \bar{\mathcal{Q}}_B + R_C \bar{\mathcal{Q}}_D\right)$$

By Lemmas 26 and 27 (in Appendix C), for sufficiently large record lengths $N_m$, $R_A$, ..., $R_C$, $\bar{\mathcal{Q}}_A$, ..., $\bar{\mathcal{Q}}_D$ are block diagonal matrices with their $m$th diagonal blocks of order $O\left(N^0\right)$. Thus, $\operatorname{tr}\left(R\bar{\mathcal{Q}}\right)$ reduces to

$$\sum_{m=0}^{M-1} \operatorname{tr}\Big(R_{A\,m}\bar{\mathcal{Q}}_{A\,m}+R_{B\,m}\bar{\mathcal{Q}}_{C\,m}+R_{B\,m}^H\bar{\mathcal{Q}}_{B\,m}+R_{C\,m}\bar{\mathcal{Q}}_{D\,m}\Big)$$

Hence,

$$\frac{\sigma_e^2}{N}\operatorname{tr}\left(R\bar{\mathcal{Q}}\right) = O\left(\sigma_e^2\frac{n_\psi}{N}\right)$$

with $n_\psi = M\left(n_I + n_J + 2\right)$. Because $N = \sum_{m=0}^{M-1} N_m$, $n_\psi/N$ converges to zero only when $N_m \to \infty$. Thus, for finite $N_m$ and $M \to \infty$, $\partial V_{EE}/\partial d_r|_{\theta_0}$ converges to constant different than zero. A similar result can be derived for $\partial V_{EE}/\partial c_r|_{\theta_0}$.

## Appendix H Proof of Lemma 11

Substituting (A.25) in (A.24) results in the simplified equation

$$\left.\frac{\partial V_{\psi E}}{\partial \theta_r}\right|_{\theta_0} = \frac{1}{N}\operatorname{tr}\left(2\,\mathrm{herm}\left\{\left[\frac{\partial L_\psi}{\partial\theta_r} - \frac{\partial L_P}{\partial\theta_r}\right]S_{\psi E}P\right\}\right)\Bigg|_{\theta_0}$$

because of (F.1), $\left(L_P^H L_P\right)^{-1} L_P^H L_\psi\big|_{\theta_0} = I_N$ (given $L_\psi\big|_{\theta_0} = L_P\big|_{\theta_0}$), and $D_E|_{\theta_0} = I_N$ [see (A.15)]. Given the definition of $L_P$, $L_\psi$ and $S_{\psi E}$ [see (A.6), (A.13), (B.3) and (B.5)]

$$\left.\frac{\partial V_{\psi E}}{\partial \theta_r}\right|_{\theta_0} = \frac{1}{N}\operatorname{tr}\Bigg(2\,\mathrm{herm}\left\{\left[\frac{\partial D_{\psi J}}{\partial\theta_r} - \frac{\partial D_{PJ}}{\partial\theta_r}\right]L_J S_{\psi EJ}P\right\}\Bigg)\Bigg|_{\theta_0} \tag{H.1}$$

This expression is evaluated below for the model parameters $a_r$, $b_r$, $c_r$ and $d_r$.

### *H.1* $\partial V_{\psi E}/\partial a_r$, $\partial V_{\psi E}/\partial b_r$ *and* $\partial V_{\psi E}/\partial c_r$

From (A.10) and (A.17), we get for parameters $a_r$, $b_r$ and $c_r$

$$\left(\frac{\partial D_{\psi J}}{\partial\theta_r} - \frac{\partial D_{PJ}}{\partial\theta_r}\right)\Bigg|_{\theta_0} = 0_N \quad \to \quad \left.\frac{\partial V_{\psi E}}{\partial\theta_r}\right|_{\theta_0} = 0 \tag{H.2}$$

### *H.2* $\partial V_{\psi E}/\partial d_r$

The proof that $\partial V_{\psi E}/\partial d_r = 0$ for $N_m \to \infty$ can be derived following a similar procedure as in Appendix J.3. Also, this result can be derived as follows. According to (4), for $N_m \to \infty$, $V_N\left(\theta, \psi, \mathtt{z_t}\right)$ is a sum of cost functions that yield consistent estimates. This implies that $V_N\left(\theta\right)$ satisfies condition (19) for all $\theta_r$ when $N_m \to \infty$. Because all terms in (17), apart from $V_{\psi E}\left(\theta\right)$, have been proved to satisfy (19) for $d_r$ when $N_m \to \infty$, so does $V_{\psi E}\left(\theta\right)$.

## Appendix I Proof of Lemma 12

If $H\left(z^{-1},\theta\right) \neq H_0\left(z^{-1}\right)$, some of the terms of (17) do not satisfy condition (19) for $a_r$. For instance, simplifying (A.20) gives

$$\left.\frac{\partial V_{EE}}{\partial a_r}\right|_{A_0(z^{-1}),B_0(z^{-1})} = \frac{\sigma_e^2}{N}\operatorname{tr}\left(D_E D_E^H\frac{\partial P}{\partial a_r}\right)\Bigg|_{A_0(z^{-1}),B_0(z^{-1})}$$

Since $D_E|_{A_0(z^{-1}),B_0(z^{-1})} \neq I_N$ [see (A.15)], this term does not converge to zero for $N \to \infty$. This is illustrated with a Monte Carlo simulation in Section 5.

Moreover, for $V_N\left(\theta\right)$ to satisfy condition (19) for $b_r$, it is necessary that $G\left(z^{-1},\theta\right) = G_0\left(z^{-1}\right)$. This implies that a bias on parameters $a_r$ introduces a bias on $b_r$.

## Appendix J Proof of Lemma 15

The modifications provided in Appendix A.4 and the remark in Appendix B.3 apply to the ARX model structure. Besides, $D_{PJ} = I_N$ and $L_{IJ}^H L_{IJ} = \sum_{k=0}^{N-1} L_{IJ_k}^H L_{IJ_k} = N\,I_{n_\psi}$ [see (A.10) and (A.28)], so that (14) results in

$$P = I_N - \frac{1}{N}L_{IJ}L_{IJ}^H \tag{J.1}$$

The analysis of Appendices G and H applies to ARX, with the modifications presented below. Equations (G.1) and (H.1) are evaluated for $b_r$ and for $d_r$ to keep the notation (since $D = A$).

### *J.1* $\partial V_{EE}/\partial b_r$ *and* $\partial V_{\psi E}/\partial b_r$

Given (A.10), (A.15) and (A.17), for the parameter $b_r$, (G.2) and (H.2) apply.

### *J.2* $\partial V_{EE}/\partial d_r \equiv \partial V_{EE}/\partial a_r$

Given (J.1), for parameter $d_r$ $(1 \le r \le n_D)$, (G.5) becomes

$$\operatorname{tr}\left(\frac{\sigma_e^2}{N}D_r\,D_{Ed}P\right)\Bigg|_{\theta_0} = \widetilde{\mathcal{B}}$$

$$\text{with} \quad \widetilde{\mathcal{B}} = \frac{\sigma_e^2}{N}\sum_{k=0}^{N-1} H_{Ed}\left(z_k^{-1}\right)z_k^{-r}\left(1 - \frac{1}{N}L_{IJ_k}L_{IJ_k}^H\right)$$

By Lemma 22 in Appendix C (with $\alpha < 0$, $N + \alpha \approx N$), and $L_{IJ_k}L_{IJ_k}^H = \sum_{l=0}^{n_\psi-1} L_{IJ_k}\left(l\right)\overline{L_{IJ_k}}\left(l\right) = n_\psi$

$$\widetilde{\mathcal{B}} = \left(1 - \frac{n_\psi}{N}\right)O\Big(\sigma_e^2\left|\lambda_{Ed}\right|^N\Big)$$

with $\lambda_{Ed}$ the dominant pole of $H_{Ed}\left(z^{-1}\right)$ and $|\lambda_{Ed}| < 1$. Thus, $\partial V_{EE}/\partial d_r|_{\theta_0}$ converges to zero for finite $N_m$ and $M \to \infty$.

### J.3 $\partial V_{\psi E}/\partial d_r \equiv \partial V_{\psi E}/\partial a_r$

From (A.10) and (A.17), we get for parameter $d_r$ (with $1 \leq r \leq n_D$)

$$\left(\frac{\partial D_{\psi J}}{\partial d_r} - \frac{\partial D_{PJ}}{\partial d_r}\right)\Bigg|_{\theta_0} = D_r D_{\psi d}$$
$$\text{with} \quad D_{\psi d} = \text{diag}\left(\cdots, 1/D_0\left(z_k^{-1}\right), \cdots\right)$$
$$D_r = \text{diag}\left(\cdots, z_k^{-r}, \cdots\right)$$

We evaluate (H.1). Given the definition of the trace and (J.1)

$$\frac{1}{N}\,\text{tr}\left(D_r D_{\psi d} L_J S_{\psi EJ} P\right)\Big|_{\theta_0} = \tilde{\mathcal{B}} - \text{tr}\left(\frac{1}{N}\tilde{\mathcal{Q}}\tilde{F}\right)$$
$$\tilde{\mathcal{B}} = \frac{1}{N}\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right) z_k^{-r} L_{J_k} S_{\psi E J_k}\Bigg|_{\theta_0}$$
$$\tilde{\mathcal{Q}} = \frac{1}{N}\sum_{k=0}^{N-1} L_{IJ_k}^H H_{\psi d}\left(z_k^{-1}\right) z_k^{-r} L_{J_k}\Bigg|_{\theta_0};\quad \tilde{F} = \sum_{k=0}^{N-1} S_{\psi E J_k} L_{IJ_k}\Bigg|_{\theta_0}$$

with $S_{\psi EJ_k}$ the $k$th column of $S_{\psi EJ}$, $L_{J_k}$ and $L_{IJ_k}$ the $k$th rows of $L_J$ and $L_{IJ}$ [see (B.11), (A.3) and (A.27)], and $H_{\psi d}\left(z_k^{-1}\right)$ the $k$th diagonal element of $D_{\psi d}$.

First, we compute $\tilde{\mathcal{B}}$. Given (B.10) to (B.14), $\tilde{\mathcal{B}}$ has two components ($\tilde{\mathcal{B}} = \tilde{\mathcal{B}}_e + \tilde{\mathcal{B}}_v$) defined by

$$\tilde{\mathcal{B}}_e = \frac{1}{N}\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right) z_k^{-r} \sum_{m=0}^{M-1} z_k^{-K_m}\eta_{J_k} T_e \mathbb{E}\left\{\frac{\chi_{e\,m}}{\sqrt{N}}\overline{E_c}\,(k)\right\}$$
$$= -\frac{\sigma_e^2}{N^2}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right) z_k^{-1-r} p_e(z_k^{-1})$$
$$\tilde{\mathcal{B}}_v = -\frac{1}{N}\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right) z_k^{-r} \sum_{m=0}^{M-1} z_k^{-K_m}\eta_{J_k} T_v \mathbb{E}\left\{\frac{\chi_{v\,m}}{\sqrt{N}}\overline{E_c}\,(k)\right\}$$
$$= \frac{\sigma_e^2}{N^2}\sum_{m=0}^{M-1}\sum_{\rho=0}^{n_J}\sum_{T=0}^{N_{m-1}-\rho-1} h(T)\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right) z_k^{-1-r-T} p_{v_\rho}(z_k^{-1})$$

with $p_e(z_k^{-1})$ and $p_{v_\rho}(z_k^{-1})$ polynomials on $z_k^{-1}$ of order $n_J$. By Lemma 22 (with $\alpha < 0$, $N+\alpha \approx N$) and Lemma 24 (with $\alpha < 0$, $N+\alpha \approx N$, $N_{m-1}-\rho \approx N_{m-1}$) in C

$$\tilde{\mathcal{B}}_e = O\Big(\sigma_e^2 \frac{M}{N}\left|\lambda_{Be}\right|^N\Big);\quad \tilde{\mathcal{B}}_v = O\Big(\sigma_e^2\left|\lambda_{Bv}\right|^N\Big)$$

with $\lambda_{Be}$ the dominant pole of $H_{\psi d}\left(z^{-1}\right)$, and $\lambda_{Bv}$ the dominant pole of $H_{\psi d}\left(z^{-1}\right)$ and $H_0\left(z^{-1}\right)$.

Next, we compute $\text{tr}\left(\frac{1}{N}\tilde{\mathcal{Q}}\tilde{F}\right)$. Because $\tilde{\mathcal{Q}}$ and $\tilde{F}$ are block matrices ($M \times M$ blocks)

$$\text{tr}\left(\frac{1}{N}\tilde{\mathcal{Q}}\tilde{F}\right) = \sum_{m=0}^{M-1}\text{tr}\Big(\frac{1}{N}\sum_{i=0}^{M-1}\tilde{\mathcal{Q}}_{(m,i)}\tilde{F}_{(i,m)}\Big)$$

with $\tilde{F}_{(i,m)} = \tilde{F}_{e\,(i,m)} + \tilde{F}_{v\,(i,m)}$ and $\tilde{\mathcal{Q}}_{(m,i)}$ given by

$$\tilde{F}_{e\,(i,m)} = \sum_{k=0}^{N-1} T_e \mathbb{E}\left\{\frac{\chi_{e\,i}}{\sqrt{N}}\overline{E_c}\,(k)\right\} z_k^{-K_m}\eta_{IJ_k}$$
$$= -\frac{\sigma_e^2}{N} T_e \sum_{k=0}^{N-1}\eta_{J_k}^H \eta_{IJ_k} z_k^{-(n_J+1)} z_k^{K_i - K_m}$$
$$\tilde{F}_{v\,(i,m)} = -\sum_{k=0}^{N-1} T_v \mathbb{E}\left\{\frac{\chi_{v\,i}}{\sqrt{N}}\overline{E_c}\,(k)\right\} z_k^{-K_m}\eta_{IJ_k}$$
$$= \frac{\sigma_e^2}{N} T_v \sum_{k=0}^{N-1}\left[\varSigma_{h\,i}\left(n_J+1\right)\,\ldots\,\varSigma_{h\,i}\left(1\right)\right]^T \eta_{IJ_k} z_k^{K_i - K_m}$$
$$\tilde{\mathcal{Q}}_{(m,i)} = \frac{1}{N}\sum_{k=0}^{N-1} H_{\psi d}\left(z_k^{-1}\right)\eta_{IJ_k}^H \eta_{J_k} z_k^{-r} z_k^{K_m - K_i}$$

By Lemmas 21 and 25 in Appendix C, $\tilde{F}_{e\,(i,m)}$ and $\tilde{F}_{v\,(i,m)}$ are $O\left(\sigma_e^2\right)$ when $i = m+1$ (or $i = 0$ with $m = M-1$) and $0_{(n_J+1)\times(n_{IJ}+1)}$ elsewhere. By Lemma 22 (with $\alpha < 0$, $N+\alpha \approx N$) $\tilde{\mathcal{Q}}_{(m,m+1)}$ and $\tilde{\mathcal{Q}}_{(M-1,0)}$ are $O\big(\left|\lambda_{Be}\right|^N\big)$. Hence,

$$\text{tr}\left(\frac{1}{N}\tilde{\mathcal{Q}}\tilde{F}\right) = O\Big(\sigma_e^2\frac{M}{N}\left|\lambda_{Be}\right|^N\Big)$$

Thus, $\partial V_{\psi E}/\partial d_r|_{\theta_0}$ converges to zero for finite $N_m$ and $M \to \infty$.

## Appendix K Proof of Corollary 17

For a consistent estimator, $\theta_0$ is the minimum of $V_N\left(\theta\right)$. Evaluating (17) in $\theta_0$ gives

$$V_N\left(\theta_0\right) = V_{EE}\left(\theta_0\right) = \frac{1}{N}\,\text{tr}\left(S_{EE}P\right)|_{\theta_0}$$

because of (F.1), $D_U\big|_{\theta_0} = 0_N$ and $D_E\big|_{\theta_0} = I_N$ [see (A.14) and (A.15)]. Besides, given (B.1) and $\text{tr}\left(P\right) = N - n_\psi$ [see (14)]

$$V_N\left(\theta_0\right) = \sigma_e^2\frac{N - n_\psi}{N}$$

Therefore, a consistent $\hat{\theta}$ implies a consistent $\hat{\sigma}_e^2$ with (9), which accounts for the degrees of freedom.

## Appendix L Proof of Lemma 18

Given (13), $\left\|\frac{\partial \mathsf{W}_N}{\partial\theta_r}\right\|_1$ can be bounded by

$$\left\|\frac{\partial \mathsf{W}_N}{\partial\theta_r}\right\|_1 \leq \left\|\frac{\partial W^H}{\partial\theta_r}\right\|_1\left\|P\right\|_1\left\|W\right\|_1 + \left\|W^H\right\|_1\left\|P\right\|_1\left\|\frac{\partial W}{\partial\theta_r}\right\|_1 + \left\|W^H\right\|_1\left\|\frac{\partial P}{\partial\theta_r}\right\|_1\left\|W\right\|_1$$

where $\|\frac{\partial W^H}{\partial \theta_r}\|_1$, $\|\frac{\partial W}{\partial \theta_r}\|_1$, $\|W^H\|_1$ and $\|W\|_1$ are $O(N^0)$ according to (A.7), (A.11) and (A.12). For OE and ARX, $\|P\|_1$ is $O(N^0)$ according to Appendix D. Because $L_P = D_{PI} L_I$ for OE, and $L_P = D_{PJ} L_{IJ}$ for ARX [see (A.6), (A.8), (A.16), (A.17) and Appendix A.4]

$$\frac{\partial L_P}{\partial \theta_r} = D_{\partial \theta r} L_P \tag{L.1}$$

with $D_{\partial \theta r} \in \mathbb{C}^{N \times N}$ a diagonal matrix.

From (A.25) and (L.1), $\|\frac{\partial P}{\partial \theta_r}\|_1$ can be bounded by

$$\left\|\frac{\partial P}{\partial \theta_r}\right\|_1 \leq \left\|P\right\|_1 \left\|D_{\partial \theta r}\right\|_1 \left\|L_P \left(L_P^H L_P\right)^{-1} L_P^H\right\|_1 + \left\|L_P \left(L_P^H L_P\right)^{-1} L_P^H\right\|_1 \left\|D_{\partial \theta r}^H\right\|_1 \left\|P\right\|_1$$

where $\|D_{\partial \theta r}\|_1$ and $\|D_{\partial \theta r}^H\|_1$ are $O(N^0)$. According to Appendix D, $\|L_P \left(L_P^H L_P\right)^{-1} L_P^H\|_1$ is $O(N^0)$ for OE and ARX. Therefore, $\|\frac{\partial P}{\partial \theta_r}\|_1$ and $\|\frac{\partial \mathsf{W}_N}{\partial \theta_r}\|_1$ are $O(N^0)$. The proof that for OE and ARX $\|\frac{\partial^2 \mathsf{W}_N}{\partial \theta_r \partial \theta_q}\|_1$ and $\|\frac{\partial^3 \mathsf{W}_N}{\partial \theta_r \partial \theta_q \partial \theta_p}\|_1$ are $O(N^0)$ follows the same lines. Note that the proof here presented applies for FIR and AR.

## Appendix M Proof of Theorem 19

From the first order Taylor series expansion of $V_N^{'}(\theta, \mathtt{z_f}) \in \mathbb{R}^{1 \times n_\theta}$ around $\theta_0$, a standard reasoning leads to the following expression for a consistent estimator [7,8,13]

$$\hat{\theta}(\mathtt{z_f}) - \theta_0 \underset{N \to \infty}{\longrightarrow} -V_N^{''-1}(\theta_0) V_N^{'T}(\theta_0, \mathtt{z_f}) \quad \text{in prob.}$$

leading to (20). Under Assumptions 1 to 7, the conditions of Theorem 17.29 in Chapter 17 of [8] are fulfilled (see Lemma 18 and Appendix N).

## Appendix N Auxiliary proof for Lemma 4 and Theorem 19

The contribution of the noise on $Y_c(k)$ is

$$\mathcal{J}_k + H_0\left(z_k^{-1}\right) E_c(k) \tag{N.1}$$

$$\text{with} \quad \mathcal{J}_k = \sum_{m=0}^{M-1} z_k^{-K_m} \frac{J_{m\,0}\left(z_k^{-1}\right)}{D_0\left(z_k^{-1}\right)}$$

The aim is to prove that, under Assumption 2, (N.1) is independent over the frequency. $H_0\left(z_k^{-1}\right) E_c(k)$ is independent over $k$ (see [8]). To prove that $\mathcal{J}_k$ is independent over $k$, it suffices to show that $\mathcal{J}_k$ and $\mathcal{J}_l$ are uncorrelated for $k \neq l$, since $\mathcal{J}_k$ and $\mathcal{J}_l$ are jointly normally distributed [as they are function of $e_c(t)$]. Similarly, to prove that $\mathcal{J}_k$ and $H_0\left(z_k^{-1}\right) E_c(k)$ are independent over $k$, we will show that $\mathcal{J}_k$ and $H_0\left(z_l^{-1}\right) E_c(l)$ are uncorrelated for $k \neq l$.

### *N.1* $\mathbb{E}\{\mathcal{J}_k \mathcal{J}_l^H\}$

From (B.6)

$$\mathcal{J}_k = \mathcal{J}_{e\,k} - \mathcal{J}_{v\,k} \tag{N.2}$$

$$\text{with} \quad \mathcal{J}_{v\,k} = \frac{\eta_{J_k} T_v}{\sqrt{N} D_0\left(z_k^{-1}\right)} \mathcal{X}_{v\,k}; \quad \mathcal{X}_{v\,k} = \sum_{m=0}^{M-1} z_k^{-K_m} \chi_{v\,m} \tag{N.3}$$

and $\mathcal{J}_{e\,k}$ defined in a similar way as $\mathcal{J}_{v\,k}$, replacing $T_v$ by $T_e$, and $\mathcal{X}_{v\,k}$ by $\mathcal{X}_{e\,k}$ (with $\mathcal{X}_{e\,k}$ same as $\mathcal{X}_{v\,k}$, but $\chi_{v\,m}$ is replaced by $\chi_{e\,m}$) [see (B.7), (B.8) and (A.4)].

From the covariance $\mathbb{E}\{(\mathcal{J}_{e\,k} - \mathcal{J}_{v\,k})\left(\mathcal{J}_{e\,l}^H - \mathcal{J}_{v\,l}^H\right)\}$, we consider the term

$$\mathbb{E}\{\mathcal{J}_{v\,k} \mathcal{J}_{v\,l}^H\} = \frac{1}{D_0(z_k^{-1}) D_0(z_l)} \eta_{J_k} T_v \, \mathbb{E}\left\{\frac{\mathcal{X}_{v\,k} \mathcal{X}_{v\,l}^H}{N}\right\} T_v^H \eta_{J_l}^H \tag{N.4}$$

The elements of $\mathcal{X}_{v\,k} \in \mathbb{C}^{(n_J+1) \times 1}$ are of the form

$$\sum_{m=0}^{M-1} z_k^{-K_m} \{v_m(-\mathtt{t}) - v_{m-1}(N_{m-1} - \mathtt{t})\} = \eta_{M_k} \nu_{\mathtt{t}}$$

$$\eta_{M_k} = \begin{bmatrix} z_k^0 \\ \vdots \\ z_k^{-K_{M-1}} \end{bmatrix}^T; \; \nu_{\mathtt{t}} = \begin{bmatrix} v_0(-\mathtt{t}) - v_{-1}(N_{-1} - \mathtt{t}) \\ \vdots \\ v_{M-1}(-\mathtt{t}) - v_{M-2}(N_{M-2} - \mathtt{t}) \end{bmatrix}$$

with $\mathtt{t} \in [1, n_J + 1]$ (and record index $-1 \equiv M - 1$).

Therefore, under Assumptions 5 and 7, the elements of $\mathbb{E}\{\frac{1}{N} \mathcal{X}_{v\,k} \mathcal{X}_{v\,l}^H\}$ in (N.4) are of the form

$$\eta_{M_k} \mathbb{E}\left\{\frac{\nu_{\mathtt{t}} \nu_{\bar{\mathtt{t}}}^H}{N}\right\} \eta_{M_l}^H = \frac{1}{N} \eta_{M_k} \begin{bmatrix} \mathsf{b} & \mathsf{c} & 0 & \cdots & \mathsf{a} \\ \mathsf{a} & \mathsf{b} & \mathsf{c} & \cdots & 0 \\ 0 & \mathsf{a} & \mathsf{b} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathsf{c} & 0 & 0 & \cdots & \mathsf{b} \end{bmatrix} \eta_{M_l}^H \tag{N.5}$$

with $\mathsf{b} = \mathbb{E}\{v_r(-\mathtt{t}) v_r(-\bar{\mathtt{t}}) + v_r(N_r - \mathtt{t}) v_r(N_r - \bar{\mathtt{t}})\}$, $\mathsf{a} = \mathbb{E}\{-v_r(N_r - \mathtt{t}) v_r(-\bar{\mathtt{t}})\}$ and $\mathsf{c} = \mathbb{E}\{-v_r(-\mathtt{t}) v_r(N_r - \bar{\mathtt{t}})\}$, where $r$ denotes that $\mathsf{a}$, $\mathsf{b}$ and $\mathsf{c}$ do not depend on the record index $m$. Then, (N.5) reduces to (because $K_{m+1} = K_m + N_r$, and $z_k^{-K_m} z_l^{K_m} = z_{l-k}^{K_m}$)

$$\eta_{M_k} \mathbb{E}\left\{\frac{\nu_{\mathtt{t}} \nu_{\bar{\mathtt{t}}}^H}{N}\right\} \eta_{M_l}^H = \left(\frac{\mathsf{b}}{N} + \frac{\mathsf{a}}{N} z_k^{-N_r} + \frac{\mathsf{c}}{N} z_k^{N_r}\right) \sum_{m=0}^{M-1} z_{l-k}^{K_m} \tag{N.6}$$

Considering (D.4), (N.6) equals zero for $k - l \neq nM$, with $n \in [0, N_r - 1]$. Hence, for a given $k$, for $N - N_r$ frequencies $l$

$$\mathbb{E}\{\mathcal{J}_{v\,k} \mathcal{J}_{v\,l}^H\} = 0$$

For $k - l = nM$ ($N_r$ frequencies $l$), $\mathbb{E}\{\mathcal{J}_{v\,k} \mathcal{J}_{v\,l}^H\} =$

$\frac{M}{N}O(N^0)$. Asymptotically, the contribution of this finite number frequencies to the cost function can be neglected. For the other terms of $\mathbb{E}\left\{\mathcal{J}_k\mathcal{J}_l^H\right\}$ the result is the same, with the proof following the same procedure. A similar proof can be derived for $\mathbb{E}\left\{\mathcal{J}_k\mathcal{J}_l\right\}$.

### *N.2* $\mathbb{E}\left\{\mathcal{J}_k H_0\left(z_l\right)\overline{E_c}\left(l\right)\right\}$

From the covariance $\mathbb{E}\left\{\mathcal{J}_k H_0\left(z_l\right)\overline{E_c}\left(l\right)\right\}$, we consider the term [see (N.2) and (N.3)]

$$\mathbb{E}\left\{\mathcal{J}_{v\,k} H_0(z_l)\overline{E_c}(l)\right\} = \frac{H_0(z_l)\,\eta_{J_k} T_v}{D_0\left(z_k^{-1}\right)} \sum_{m=0}^{M-1} z_k^{-K_m}\mathbb{E}\left\{\frac{\chi_{v\,m}}{\sqrt{N}}\overline{E_c}(l)\right\}$$

For records of equal length $N_r$, the term $\Sigma_{h\,m}\left(\mathtt{t}\right)$ (with $\mathtt{t} \in [1, n_J+1]$ ) in (B.14) does not depend on the record index $m$ [see (B.12), where $N_{m-1} \equiv N_r$], and we denote this by $\Sigma_{h\,r}\left(\mathtt{t}\right)$. Therefore,

$$\sum_{m=0}^{M-1} z_k^{-K_m}\mathbb{E}\left\{\frac{\chi_{v\,m}}{\sqrt{N}}\overline{E_c}(l)\right\} = -\frac{\sigma_e^2}{N}\begin{bmatrix}\Sigma_{h\,r}\left(n_J+1\right)\\ \vdots \\ \Sigma_{h\,r}\left(1\right)\end{bmatrix}\sum_{m=0}^{M-1} z_{l-k}^{K_m} \tag{N.7}$$

Considering (D.4), (N.6) equals zero for $k-l \neq nM$, with $n \in [0,\, N_r-1]$. Hence, for a given $k$, for $N-N_r$ frequencies $l$

$$\mathbb{E}\left\{\mathcal{J}_{v\,k} H_0\left(z_l\right)\overline{E_c}\left(l\right)\right\} = 0$$

For $k-l = nM$ ($N_r$ frequencies $l$), $\mathbb{E}\left\{\mathcal{J}_{v\,k} H_0\left(z_l\right)\overline{E_c}\left(l\right)\right\}$ $= \frac{M}{N}O(N^0)$. Asymptotically, the contribution of this finite number frequencies to the cost function can be neglected. For the other terms of $\mathbb{E}\left\{\mathcal{J}_k H_0\left(z_l\right)\overline{E_c}\left(l\right)\right\}$ the result is the same, with the proof following the same procedure [see (B.13)]. A similar proof can be derived for $\mathbb{E}\left\{\mathcal{J}_k H_0\left(z_l\right)E_c\left(l\right)\right\}$.

## References


[1] D. Leith, D. Murray-Smith, R. Bradley, Combination of data sets for system identification, IEE Proceedings D - Control Theory and Applications 140 (1) (1993) 11–18.

[2] R. Relan, K. Tiels, J. Timmermans, J. Schoukens, A local polynomial approach to nonparametric estimation of the best linear approximation of lithium-ion battery from multiple datasets, IEEE Control Systems Letters 1 (1) (July 2017).

[3] B. Mavkov, E. Witrant, C. Prieur, D. Moreau, Multi-experiment state-space identification of coupled magnetic and kinetic parameters in tokamak plasmas, Control Engineering Practice 60 (March 2017).

[4] Q. Zhang, L. Ljung, LPV system common state basis estimation from independent local LTI models, IFAC-PapersOnLine 48 (28) (2015) 190–195.

[5] R. J. A. Little, D. B. Rubin, Statistical analysis with missing data, 3rd Edition, Wiley, 2019.

[6] A. Isaksson, Identification of models subject to missing data, IEEE Transactions on Automatic Control 38 (5) (1993) 813 – 819.

[7] L. Ljung, System identification: Theory for the user, 2nd Edition, 1999.

[8] R. Pintelon, J. Schoukens, System identification: A frequency domain approach, 2nd Edition, IEEE Press, 2012.

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis. Forecasting and control, 5th Edition, Wiley, 2015.

[10] J. Schoukens, G. Vandersteen, Y. Rolain, R. Pintelon, Frequency response function measurements using concatenated subrecords with arbitrary length, IEEE Transactions on Instrumentation and Measurement (10) (2012) 2682 – 2688.

[11] L. Ljung, System identification toolbox: User's guide., MathWorks (2014).

[12] I. Markovsky, R. Pintelon, Identification of linear time-invariant systems from multiple experiments, IEEE Transactions on Signal Processing 63 (13) (July 2015).

[13] T. Söderström, Convergence properties of the generalised least squares identification method, Automatica 10 (1974) 617–626.

[14] M. Tismenetsky, On inversion of hermitian block toeplitz matrices, Linear and Multilinear Algebra 35 (2) (1993) 165–171.