Estimating Models with High-Order Noise Dynamics Using Semi-Parametric Weighted Null-Space Fitting

Miguel Galrinho [a], Cristian R. Rojas [a], Håkan Hjalmarsson [a]

[a] *Automatic Control Department, School of Electrical Engineering and Data Science,*
KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.

Abstract

Standard system identification methods often provide inconsistent estimates with closed-loop data. With the prediction error method (PEM), this issue is solved by using a noise model that is flexible enough to capture the noise spectrum. However, a too flexible noise model (i.e., too many parameters) increases the model complexity, which can cause additional numerical problems for PEM. In this paper, we consider the weighted null-space fitting (WNSF) method. With this method, the system is first modeled using a non-parametric ARX model, which is then reduced to a parametric model of interest using weighted least squares. In the reduction step, a parametric noise model does not need to be estimated if it is not of interest. Because the flexibility of the noise model is increased with the sample size, this will still provide consistent estimates in closed loop and asymptotically efficient estimates in open loop. In this paper, we prove these results, and we derive the asymptotic covariance for the estimation error obtained in closed loop, which is optimal for an infinite-order noise model. For this purpose, we also derive a new technical result for geometric variance analysis, instrumental to our end. Finally, we perform a simulation study to illustrate the benefits of the method when the noise model cannot be parametrized by a low-order model.

Key words: System identification; closed-loop identification; non-parametric identification; parameter identification; identification algorithms; least squares

1 Introduction

The prediction error method (PEM) is a benchmark for estimation of linear parametric models. If the model orders are correct and the noise is Gaussian, PEM with a quadratic cost function is asymptotically efficient (Ljung, 1999): the asymptotic covariance of the estimates coincides with the Cramér-Rao (CR) bound, the lowest covariance attainable by a consistent estimator.

Two models can typically be distinguished in a parametric model structure: the dynamic model and the noise model. Because the noise sequence is often the result of different noise contributions aggregated in a complex manner, the concept of a "correct order" for the noise model is often intractable in practice. While PEM is still

consistent in open loop when the noise model cannot capture the actual noise spectrum, this is not the case for data collected in closed loop.

This issue with closed-loop data is not exclusive of PEM. Instrumental variable methods (Söderström and Stoica, 1983) require the reference signal in order to construct the instruments in closed loop (Gilson and van den Hof, 2005; Gilson et al., 2009). Classical subspace methods (van Overschee and de Moor, 1994) also suffer from inconsistency in closed loop, although this issue has been overcome by more recent algorithms: for example, Verhaegen (1993) estimates the complete open-loop system followed by a model reduction step; Qin and Ljung (2003) estimate the innovations; Jansson (2003) uses a non-parametric ARX model to construct the Hankel matrices; Chiuso and Picci (2005) use a whitening filter.

With PEM, the inconsistency issue can in theory be solved by letting the noise model structure be arbitrarily flexible (i.e., letting the number of estimated parameters become arbitrarily large), guaranteeing that a correct noise spectrum can be captured by the model. If the global minimum of the PEM cost function is found, in open loop this will asymptotically not affect the statisti-

cal properties of the dynamic-model estimate; in closed loop, consistency is attained but not efficiency. The problem with this approach is that, because the noise model might require many parameters, the optimization problem to solve becomes computationally heavier, and the PEM cost function may have more local minima, thus aggravating the numerical search for the global minimum.

Some methods use a semi-parametric approach, estimating a non-parametric model in a first step, which is then used in a second step to estimate the dynamic model. In this case, an estimate of the noise model is no longer needed for consistency of the dynamic model. This type of approach is more user-friendly, because the user is not required to choose an appropriate parametric noise model: this has been highlighted by Schoukens et al. (2011), who propose a frequency-domain method using non-parametric noise models, based on the theory developed by Schoukens et al. (2009); Pintelon et al. (2010a,b). In the time domain, methods based on the same idea have also been proposed by Zhu (2001); Zhu and Hjalmarsson (2016); Everitt et al. (2017). The two latter ones have the advantage of not requiring numerical search algorithms, but they have only been considered for open loop data.

The weighted null-space fitting (WNSF) method (Galrinho et al., 2014) also first estimates a non-parametric model, and then reduces it to a parametric model, which may or may not include the noise model, and can be used with closed-loop data. Moreover, similarly to the methods by Zhu and Hjalmarsson (2016); Everitt et al. (2017), WNSF does not apply non-linear optimization techniques, but uses weighted least squares. In this sense, the method can be seen as belonging to a family of iterative least-squares methods with an intermediate non-parametric model estimate.

Precursors of the method have used an impulse response estimate to obtain a rational transfer function (Evans and Fischl, 1973; Shaw, 1994; Shaw et al., 1994; Lemmerling et al., 2001). The origin of this family of methods can be traced back to the field of time-series analysis. Durbin (1960) proposes two methods for auto-regressive moving-average (ARMA) time series estimation, using a high-order AR time series as intermediate step to obtain the ARMA parameters by least squares. The first method is non-iterative but does not attain the CR bound, while the second method remedies this by iterating between estimating the AR and MA polynomials, initialized with the first method. An alternative way to attain the CR bound from Durbin's first method as starting point, and which simplified the analysis, was proposed by Mayne and Firoozan (1982), using an additional filtering step.

An important theoretical challenge is how to establish consistency and asymptotic efficiency for this type of methods. The aforementioned papers for estimation of time series consider the non-parametric model order as tending to infinity, but "small" compared to the sample size. In practice, however, it is intuitive that the non-parametric model order is chosen depending on the available sample size. In this sense, the theoretical analysis should consider this relation formally, with the non-parametric model order tending to infinity as function of the sample size, according to some specified rate. This dependency is introduced by Hannan and Kavalieris (1983) to prove consistency of the method by Mayne and Firoozan (1982). Later applications include vector ARMA time-series (Hannan and Kavalieris, 1984; Reinsel et al., 1992; Dufour and Jouini, 2014), for which also asymptotic efficiency has been treated formally.

Galrinho et al. (2018) have derived the asymptotic properties of WNSF for a fully parametric model (i.e., dynamic and noise models), making use of the results by Ljung and Wahlberg (1992) on the statistical properties of non-parametric least-squares estimates when the model order tends to infinity as function of the sample size at particular rates. In particular, the fully parametric WNSF is shown to be consistent and asymptotically efficient with open- and closed-loop data when the correct model structure is chosen. In this paper, instead of estimating both the dynamic and noise models, we disregard the parametric noise model, reducing the non-parametric model estimate to obtain a parametric dynamic model only. The dynamic model estimate will then be asymptotically efficient in open loop, and consistent in closed loop, with optimal asymptotic covariance for an infinite-order noise model. The asymptotic properties of the proposed method correspond to the asymptotic properties of PEM with an infinite-order noise model (Forssell and Ljung, 1999) in both open and closed loop, but performed with a robust numerical procedure, and without using a numerical search algorithm.

The case addressed in this paper, where WNSF is used with no parametric noise-model estimate, will be denoted semi-parametric WNSF. Despite having been mentioned by Galrinho et al. (2018), a formal analysis of the semi-parametric case has not been considered. Technically, the theoretical analysis of this setting is significantly more challenging than the fully parametric case. The reason is that expressions with structure $(TR^{-1}T')^{-1}$, where T and R are matrices, appear in both variants of the method; however, T and R are both square and invertible in the fully parametric case, and the analysis can be done using $T^{-\prime}RT^{-1}$, while this is not the case for the semi-parametric case. This type of structure arises in the analysis by Hjalmarsson and Mårtensson (2011), whose geometric approach to variance analysis can be helpful here. There, however, the sizes of the matrices are constant, whereas here the dimensions grow unlimited, causing important technical issues that need to be resolved. This motivates that, in Section 4.2, we derive a new result based on the approach by Hjalmarsson and Mårtensson (2011) such

that it can be applied in the setting of our problem.

The paper is organized in the following way. In Section 2, we introduce definitions, assumptions, and background. In Section 3, we present the semi-parametric algorithm of WNSF. In Section 4, we provide the asymptotic analysis, which includes: 1) a new result for variance analysis that is instrumental for our problem; 2) consistency with open- and closed-loop data; 3) asymptotic efficiency with open-loop data and the corresponding asymptotic covariance matrix with closed-loop data. In Section 5, we perform an experimental analysis illustrating the potential of the method.

2   Preliminaries

Most of the notation, definitions, and assumptions used by Galrinho et al. (2018) apply to this paper. For completeness, we present them here.

2.1   Notation

- A' is the transpose of matrix A.
- A^* is the complex conjugate transpose of matrix A.
- $\|x\| = \sqrt{\sum_{k=1}^n |x_k|^2}$, with x an $n \times 1$ vector.
- $\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\|$, with A a matrix and x a vector of appropriate dimensions.
- $\|A\|_F = \sqrt{\text{Trace}(AA^*)}$ (i.e., the Frobenius norm), with A a matrix.
- $\left\|\bar{G}(q)\right\|_{\mathcal{H}_2} := \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Trace}\bar{G}(e^{i\omega})\bar{G}^*(e^{i\omega})d\omega}$, with $\bar{G}(q)$ a transfer matrix.
- $\left\|\bar{G}(q)\right\|_{\mathcal{H}_\infty} := \sup_\omega \|\bar{G}(e^{i\omega})\|$.
- C denotes any constant, which need not be the same in different expressions.
- $\Gamma_n(q) = [q^{-1} \quad \cdots \quad q^{-n}]'$, where q^{-1} is the backward time-shift operator.
- $\mathcal{T}_{n,m}(X(q))$ is the Toeplitz matrix of size $n \times m$ ($m \leq n$) with first column $[x_0 \quad \cdots \quad x_{n-1}]'$, where $X(q) = \sum_{k=0}^\infty x_k q^k$, and zeros above the main diagonal.
- $\mathbb{E}x$ denotes expectation of the random vector x.
- $\bar{\mathbb{E}}x_t := \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^N \mathbb{E}x_t$.
- $x_N = \mathcal{O}(f_N)$ means that the function x_N tends to zero at a rate not slower than f_N, as $N \to \infty$, w.p.1.
- $\langle X(q), Y(q) \rangle := \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{i\omega})Y^*(e^{i\omega})d\omega$, with $X(q)$ and $Y(q)$ transfer matrices of appropriate sizes.

2.2   Assumptions

Assumption 1 (True system and parametric model)
The system has scalar input $\{u_t\}$, scalar output $\{y_t\}$, and is subject to scalar noise $\{e_t\}$. These signals are related by

$$y_t = G_o(q)u_t + H_o(q)e_t, \qquad (1)$$

where $G_o(q)$ and $H_o(q)$ are rational functions given by

$$G_o(q) = \frac{L_o(q)}{F_o(q)} = \frac{l_1^o q^{-1} + \cdots + l_{m_l}^o q^{-m_l}}{1 + f_1^o q^{-1} + \cdots + f_{m_f}^o q^{-m_f}},$$
$$H_o(q) = \frac{C_o(q)}{D_o(q)} = \frac{1 + c_1^o q^{-1} + \cdots + c_{m_c}^o q^{-m_c}}{1 + d_1^o q^{-1} + \cdots + d_{m_d}^o q^{-m_d}}. \qquad (2)$$

The transfer functions G_o, H_o, and H_o^{-1} are assumed to be stable. The polynomials L_o and F_o—as well as C_o and D_o—do not share common factors.

We parametrize $G(q)$ as

$$G(q, \theta) = \frac{L(q, \theta)}{F(q, \theta)} = \frac{l_1 q^{-1} + \cdots + l_{m_l} q^{-m_l}}{1 + f_1 q^{-1} + \cdots + f_{m_f} q^{-m_f}}, \quad (3)$$

where

$$\theta = \begin{bmatrix} f_1 & \cdots & f_{m_f} & l_1 & \cdots & l_{m_l} \end{bmatrix}'$$

is the parameter vector, with known orders m_f and m_l. We assume that there is θ_o such that $G(q, \theta_o) = G_o(q)$. The orders m_c and m_d of the noise model numerator and denominator polynomials are not known.

Because we allow for data to be collected in closed loop, the input $\{u_t\}$ is allowed to have a stochastic part. Then, let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $\{e_s, u_s, s \leq t-1\}$. For the noise, the following assumption applies.

Assumption 2 (Noise) *The noise sequence $\{e_t\}$ is a stochastic process that satisfies*

$$\mathbb{E}[e_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[e_t^2|\mathcal{F}_{t-1}] = \sigma_o^2, \quad \mathbb{E}[|e_t|^{10}] \leq C, \forall t.$$

Before stating the assumption on the input, we introduce some definitions from Ljung and Wahlberg (1992).

Definition 1 (f_N-quasi-stationarity) *Let f_N be a decreasing sequence of positive scalars, with $f_N \to 0$ as $N \to \infty$, and*

$$R_{vv}^N(\tau) = \begin{cases} \frac{1}{N} \sum_{t=\tau+1}^N v_t v_{t-\tau}', & 0 \leq \tau < N, \\ \frac{1}{N} \sum_{t=1}^{N+\tau} v_t v_{t-\tau}', & -N < \tau \leq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The vector sequence $\{v_t\}$ is f_N-quasi-stationary if

(1) There exists $R_{vv}(\tau)$ such that
   $\sup_{|\tau| \leq N} \left\| R_{vv}^N(\tau) - R_{vv}(\tau) \right\| \leq C_1 f_N,$
(2) $\frac{1}{N} \sum_{t=-N}^N \|v_t\|^2 \leq C_2$

for all N large enough, where C_1 and C_2 are finite constants.

Definition 2 (f_N-stability) *A filter $G(q) = \sum_{k=0}^{\infty} g_k q^{-k}$ is f_N-stable if $\sum_{k=0}^{\infty} |g_k|/f_k < \infty$.*

Definition 3 (Power spectral density) *The power spectral density of an f_N-quasi-stationary sequence $\{v_t\}$ is given by $\Phi_v(z) = \sum_{\tau=-\infty}^{\infty} R_{vv}(\tau) z^{-\tau}$, if the sum exists for $|z| = 1$.*

Assumption 3 (Input) *The input sequence $\{u_t\}$ is defined by $u_t = -K(q)y_t + r_t$ under the following conditions.*

(1) *The sequence $\{r_t\}$ is independent of $\{e_t\}$, f_N-quasi-stationary with $f_N = \sqrt{\log N/N}$, and uniformly bounded.*
(2) *With $\Phi_r(z) = F_r(z)F_r(z^{-1})$ the spectral factorization of $\{r_t\}$ and $F_r(z)$ causal, $F_r(q)$ is BIBO stable.*
(3) *The closed loop system is f_N-stable with $f_N = 1/\sqrt{N}$.*
(4) *The transfer function $K(z)$ is bounded on the unit circle.*
(5) *The spectral density of $\{[r_t \ e_t]'\}$ is bounded from below by the matrix δI, for some $\delta > 0$.*

Operation in open loop is obtained by taking $K(q) = 0$.

Alternatively to (1), the true system can be written as

$$A_o(q)y_t = B_o(q)u_t + e_t,$$

where

$$
\begin{aligned}
A_o(q) &:= \frac{1}{H_o(q)} =: 1 + \sum_{k=1}^{\infty} a_k^o q^k, \\
B_o(q) &:= \frac{G_o(q)}{H_o(q)} =: \sum_{k=1}^{\infty} b_k^o q^k
\end{aligned}
\tag{4}
$$

are stable (Assumption 1). In a first step, WNSF estimates truncated versions of $A_o(q)$ and $B_o(q)$, using the ARX model

$$A(q, \eta^n)y_t = B(q, \eta^n)u_t + e_t, \tag{5}$$

where

$$\eta^n = \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_n \end{bmatrix}', \tag{6}$$

$$A(q, \eta^n) = 1 + \sum_{k=1}^{n} a_k q^{-k}, \quad B(q, \eta^n) = \sum_{k=1}^{n} b_k q^{-k}.$$

Because the order needs to be infinite for the system to be in the ARX model set, we make the model order n depend on the sample size N—denoted $n = n(N)$—according to the following assumption.

Assumption 4 (ARX-model order) *The ARX model order is selected according to:*

D1. $n(N) \to \infty$, *as* $N \to \infty$;
D2. $n^{4+\delta}(N)/N \to 0$, *for some* $\delta > 0$, *as* $N \to \infty$.

Compared with Galrinho et al. (2018), the only difference in the assumptions is on the model: therein also the noise model uses a parametric structure analogous to (3), whereas here it is not estimated.

2.3 Prediction Error Method

The prediction error method minimizes a cost function of the prediction errors

$$\varepsilon_t(\theta, \zeta) = H^{-1}(q, \zeta)\left(y_t - \frac{L(q, \theta)}{F(q, \theta)}u_t\right), \tag{7}$$

where $H(q, \zeta)$ is a noise model, parametrized by

$$H(q, \zeta) = \frac{C(q, \zeta)}{D(q, \zeta)} = \frac{1 + c_1 q^{-1} + \cdots + c_{m_c} q^{-m_c}}{1 + d_1 q^{-1} + \cdots + d_{m_d} q^{-m_d}},$$

with

$$\zeta = \begin{bmatrix} c_1 & \cdots & c_{m_c} & d_1 & \cdots & d_{m_d} \end{bmatrix}'$$

With a quadratic cost function, the PEM estimate of the parameters is obtained by minimizing

$$J(\theta, \zeta) = \frac{1}{N}\sum_{t=1}^{N} \frac{1}{2}\varepsilon_t^2(\theta, \zeta), \tag{8}$$

where N is the sample size. Using the quadratic cost function (8) can provide asymptotically optimal estimates when the noise is Gaussian.

Let $H(q, \zeta)$ be such that there exists $\zeta = \zeta_o$ for which $H(q, \zeta_o) = H_o(q)$. Denoting by $\hat{\theta}_N^{\text{PEM}}$ the parameter vector θ that (together with some ζ) minimizes (8), the estimate $\hat{\theta}_N^{\text{PEM}}$ is asymptotically distributed as (Ljung, 1999)

$$\sqrt{N}(\hat{\theta}_N^{\text{PEM}} - \theta_o) \sim As\mathcal{N}(0, \sigma_o^2 M_{\text{PEM}}^{-1}), \tag{9}$$

where $\mathcal{N}$ stands for the Gaussian distribution. Let Φ_u^r be the spectrum of

$$u_t^r = S_o(q)r_t, \tag{10}$$

with $S_o(q) = [1 + K(q)G_o(q)]^{-1}$ the sensitivity function, and Γ_{m_f} and Γ_{m_l} be according to the definition of Γ_n in Section 2.1 with $n = m_f$ and $n = m_l$, respectively. Then, the asymptotic covariance matrix in (9) satisfies

$$M_{\text{PEM}} \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \bar{\Omega}(e^{i\omega})\Phi_u^r(e^{i\omega})\bar{\Omega}^*(e^{i\omega})d\omega =: M, \tag{11}$$

where

$$\bar{\Omega}(e^{i\omega}) = \begin{bmatrix} -\frac{G_o}{H_o F_o} \Gamma_{m_f} \\ \frac{1}{H_o F_o} \Gamma_{m_l} \end{bmatrix}.$$

Inside the matrix, arguments were omitted for notational simplicity; however, for clarity, we point out that functions of q are evaluated at $q = e^{i\omega}$ when inside of integrals that have the frequency ω as variable of integration.

In open loop (in which case Φ_u^r is simply the input spectrum), $M_{\text{PEM}} = M$, and it corresponds to the CR bound under a Gaussian noise assumption. In closed loop, $M_{\text{PEM}} = M$ when the number of parameters in ζ tends to infinity. In this case, M does not correspond to the CR bound, but to the optimal covariance (from a prediction error perspective) with an infinite-order noise model (Forssell and Ljung, 1999). For additional discussion on open- and closed-loop accuracy, we refer to Bombois et al. (2011); Aguero and Goodwin (2007)

The interest of estimating a non-parametric noise model in closed loop is that even if the noise spectrum needs to be captured by a high-order model, it will still be possible to obtain a consistent estimate of the dynamic model $G(q, \theta)$. However, estimating a non-parametric noise model simultaneously with a parametric dynamic model with PEM is not realistic. The reason is that, as the number of parameters in $H(q, \zeta)$ increases, the prediction error (7) becomes a more complicated function of ζ, which makes the problem computationally heavier, and more difficult to find the global minimum of the non-convex cost function (8). Consequently, the theoretically attractive result that PEM with a non-parametric noise model provides estimates with covariance corresponding to M may not always be useful in practice. It turns out that this setting can be handled with WNSF without increasing the difficulty of the problem.

3 Semi-Parametric Weighted Null-Space Fitting Algorithm

The WNSF method consists of three steps (Galrinho et al., 2018). First, we estimate a non-parametric ARX model, with least squares. Second, we reduce this estimate to a parametric model, with least squares, providing a consistent estimate. Third, we re-estimate the parametric model, with weighted least squares, where an estimate of the optimal weighting is used to attain an asymptotically efficient estimate. We now consider the procedure for each step, without estimating a parametric noise model. Because WNSF has been already presented by Galrinho et al. (2014, 2018), we will not go into detail here on the motivation and derivation of each equation, but instead focus on the idea of the semi-parametric algorithm.

For the first step, consider (5) in the regression form

$$y_t = (\varphi_t^n)' \eta^n + e_t,$$
$$\varphi_t^n = \begin{bmatrix} -y_{t-1} \cdots -y_{t-n} & u_{t-1} \cdots u_{t-n} \end{bmatrix}'.$$

Then, the least-squares estimate of η^n is obtained by

$$\hat{\eta}_N^n = [R_N^n]^{-1} r_N^n, \tag{12}$$

where

$$R_N^n = \frac{1}{N} \sum_{t=n+1}^{N} \varphi_t^n (\varphi_t^n)', \quad r_N^n = \frac{1}{N} \sum_{t=n+1}^{N} \varphi_t^n y_t, \tag{13}$$

for which we have that (Ljung and Wahlberg, 1992)

$$R_N^n \to \bar{R}^n := \bar{\mathbb{E}}\left[\varphi_t^n (\varphi_t^n)'\right], \text{ as } N \to \infty, \text{ w.p.1,}$$
$$r_N^n \to \bar{r}^n := \bar{\mathbb{E}}\left[\varphi_t^n y_t\right], \quad \text{ as } N \to \infty, \text{ w.p.1,}$$
$$\hat{\eta}_N^n \to \bar{\eta}^n := \left[\bar{R}^n\right]^{-1} \bar{r}^n, \text{ as } N \to \infty, \text{ w.p.1.}$$

For the second step, we obtain an estimate of $G(q, \theta)$, from the non-parametric ARX-model estimate. For this purpose, we may use (2) and (4) to write

$$C_o(q) A_o(q) - D_o(q) = 0,$$
$$F_o(q) B_o(q) - L_o(q) A_o(q) = 0. \tag{14}$$

Because we are not interested in estimating a parametric noise model, the first equation in (14) is not relevant for our purposes. Then, we require only

$$F_o(q) B_o(q) - L_o(q) A_o(q) = 0. \tag{15}$$

By convolution, (15) can be written in matrix form as

$$b_o^n - Q_n(\eta_o^n) \theta_o = 0, \tag{16}$$

where η_o^n is given by (6) evaluated at the true coefficients of (4), b_o^n consists of $b^n = [b_1 \ \dots \ b_n]'$ also evaluated at the true coefficients, and

$$Q_n(\eta^n) = \begin{bmatrix} -Q_n^f(\eta^n) & Q_n^l(\eta^n) \end{bmatrix},$$
$$Q_n^l(\eta^n) = \mathcal{T}_{n,m_l}(A(q, \eta^n)), \ Q_n^f(\eta^n) = \mathcal{T}_{n,m_f}(B(q, \eta^n)), \tag{17}$$

where $\mathcal{T}_{n \times m_f}$ is according to the definition in Section 2.1 with $m = m_f$. Motivated by (16), we replace η_o^n by its estimate $\hat{\eta}_N^n$ (and the same for b_o^n, which is a part of η_o^n) and obtain an estimate of θ with least squares:

$$\hat{\theta}_N^{\text{LS}} = (Q_n'(\hat{\eta}_N^n) Q_n(\hat{\eta}_N^n))^{-1} Q_n'(\hat{\eta}_N^n) \hat{b}_N^n. \tag{18}$$

This estimate, as will be shown in Theorem 1, is consistent.

For the third step, we re-estimate θ to obtain an asymptotically efficient estimate. This is done by taking into account the statistical properties of the errors in $\hat{\eta}_N^n$. As η^n_o is replaced by $\hat{\eta}_N^n$ in (16), the residuals can be written as

$$\hat{b}_N^n - Q_n(\hat{\eta}_N^n)\theta_\mathrm{o} = T_n(\theta_\mathrm{o})(\hat{\eta}_N^n - \eta_\mathrm{o}^n), \qquad (19)$$

where

$$T_n(\theta) = \begin{bmatrix} -T_n^l(\theta) & T_n^f(\theta) \end{bmatrix},$$
$$T_n^l(\theta) = \mathcal{T}_{n,n}(L(q,\theta)), \ T_n^f(\theta) = \mathcal{T}_{n,n}(F(q,\theta)). \ (20)$$

For the term $\hat{\eta}_N^n - \eta_\mathrm{o}^n$ in (19), if we neglect the bias error originating from the truncation taking place in the ARX model (which should be close to zero for sufficiently large n) we have that, approximately,

$$\sqrt{N}\left(\hat{\eta}_N^n - \bar{\eta}^n\right) \sim As\mathcal{N}(0, \sigma_\mathrm{o}^2[\bar{R}^n]^{-1}).$$

This allows us to express the covariance of the residuals (19) as being proportional to

$$\bar{W}_n^{-1}(\theta_\mathrm{o}) = T_n(\theta_\mathrm{o})[\bar{R}^n]^{-1}T_n'(\theta_\mathrm{o}), \qquad (21)$$

Using the inverse of (21) as weighting, when solving (19) in a least-squares sense, minimizes the variance of the parameter estimate. Although this covariance is dependent on the true parameters, a consistent estimate is available from Step 2. Hence, we may use as weighting

$$W_n(\hat{\theta}_N^{\mathrm{LS}}) = \left(T_n(\hat{\theta}_N^{\mathrm{LS}})[R^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}})\right)^{-1}. \qquad (22)$$

and the estimate obtained in this step is thus given by

$$\hat{\theta}_N^{\mathrm{WLS}} = \left(Q_n'(\hat{\eta}_N^n)W_n(\hat{\theta}_N^{\mathrm{LS}})Q_n(\hat{\eta}_N^n)\right)^{-1}Q_n'(\hat{\eta}_N^n)W_n(\hat{\theta}_N^{\mathrm{LS}})\hat{b}_N^n. \qquad (23)$$

The algorithm may be summarized as follows.

Algorithm 1 *The semi-parametric WNSF method consists of the following steps:*

(1) *compute a non-parametric ARX-model estimate with (12);*
(2) *compute a parametric dynamic-model estimate with (18);*
(3) *re-compute a parametric dynamic-model estimate with (23).*

Optionally, we may continue to iterate, potentially improving the estimation quality for finite sample size. However, we show in the next section that Algorithm 1 has the same asymptotic properties as PEM with an infinite-order noise model. Nevertheless, the algorithm has advantages with respect to PEM with an arbitrarily flexible noise model. First, WNSF estimates the non-parametric noise model in a separate step, as part of

an ARX model, which is linear in the model parameters; thus, it does not make the problem computationally more difficult, unlike if (8) is minimized with an arbitrary large number of noise model parameters. Second, it overall does not use numerical search algorithms that can converge to non-global optima. The price to pay is that, even if the noise spectrum can be modeled parametrically, WNSF still requires a noise model whose order tends to infinity in order to satisfy the statistical properties that we proceed to show.

4 Theoretical Analysis

In this section, we perform a theoretical analysis of the semi-parametric WNSF method. In particular, we will show that Step 3 in Algorithm 1 provides a consistent estimate and derive its covariance matrix. For that, we will need some auxiliary results.

4.1 Results from Galrinho et al. (2018)

To show the aforementioned results, we will need that the estimate obtained in Step 2 of Algorithm 1 is consistent. For that, we have the following result.

Theorem 1 *Let Assumptions 1, 2, 3, and 4 hold, and let $\hat{\theta}_N^{LS}$ be given by (18). Then,*

$$\hat{\theta}_N^{LS} \to \theta_o, \ as \ N \to \infty, \ w.p.1.$$

PROOF. For the fully parametric case, where θ additionally contains the noise model parameters, the analogous result is shown by Galrinho et al. (2018) in Theorem 1. Therein, the idea of the proof is to consider separately the part of the expression in (18) providing the noise model estimates and the dynamic model estimates, as the two problems are separable when no weighting is used. The part corresponding to the dynamic model, which in turn corresponds to $Q(\eta^n)$ given by (17), are identical for both fully- and semi-parametric cases. Hence, the result follows from Theorem 1 in Galrinho et al. (2018). □

Although consistency of Step 2 with semi-parametric WNSF is a specific case of the results for the fully-parametric method, consistency and asymptotic covariance of Step 3 are technically more challenging to derive. In the following subsection, we provide further insight into why, and derive a result that will be instrumental for the remainder of our analysis.

4.2 Result for variance analysis with geometric approach

We begin by writing, for the estimate from Step 3,

$$
\hat{\theta}_N^{\mathrm{WLS}} - \theta_{\mathrm{o}}
= M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) Q_n'(\hat{\eta}_N) W_n(\hat{\theta}_N^{\mathrm{LS}}) T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \eta_{\mathrm{o}}^{n(N)}),
\tag{24}
$$

where we define $M(\eta^n, \theta) := Q_n'(\eta^n) W_n(\theta) Q_n(\eta^n)$ and $\hat{\eta}_N := \hat{\eta}_N^{n(N)}$, recalling that n is a function of N according to Assumption 4 (for notational simplicity, we use only n in matrix subscripts even when it is a function of N). To analyze the asymptotic properties of semi-parametric WNSF, the limit value of (24) and the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N^{\mathrm{WLS}} - \theta_{\mathrm{o}})$ will be considered. The technical challenge in this analysis, compared to Galrinho et al. (2018), comes from the matrix $W_n(\hat{\theta}_N^{\mathrm{LS}})$, and consequently also the matrix $M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$, which contains $W_n(\hat{\theta}_N^{\mathrm{LS}})$. Considering (20), we observe that the outer inverse in (22) cannot be computed by taking the individual inverses contained in $W_n(\hat{\theta}_N^{\mathrm{LS}})$, as consequence of $T_n(\hat{\theta}_N^{\mathrm{LS}})$ not being square. On the other hand, in the fully parametric case, the matrix $T_n(\hat{\theta}_N^{\mathrm{LS}})$ is square and converges to an invertible matrix (Galrinho et al., 2018). Hence, we may write $W_n(\hat{\theta}_N^{\mathrm{LS}}) = T_n^{-\prime}(\hat{\theta}_N^{\mathrm{LS}}) R_N^n T_n^{-1}(\hat{\theta}_N^{\mathrm{LS}})$ for the fully parametric case, which is used throughout the analysis by Galrinho et al. (2018), but not for the semi-parametric case considered in this paper.

To deal with this issue, we use the approach by Hjalmarsson and Mårtensson (2011), writing the aforementioned matrices as projections of the rows of some matrix onto the subspace spanned by the rows of another matrix. This will be applied to the limit value of the matrix $M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$, defined by

$$
\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) := \lim_{n \to \infty} Q_n'(\eta_{\mathrm{o}}^n)[T_n(\theta_{\mathrm{o}})(\bar{R}^n)^{-1} T_n'(\theta_{\mathrm{o}})]^{-1} Q_n(\eta_{\mathrm{o}}^n).
\tag{25}
$$

Writing $Q_n(\eta_{\mathrm{o}}^n)$, $T_n(\theta_{\mathrm{o}})$, and $\bar{R}^n$ (defined in (17), (20), and (13), respectively) in the frequency domain, we have

$$
Q_n'(\eta_{\mathrm{o}}^n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \begin{bmatrix} \Gamma_{m_f} & 0 \\ 0 & \Gamma_{m_l} \end{bmatrix} \begin{bmatrix} -B_{\mathrm{o}} \\ A_{\mathrm{o}} \end{bmatrix} \Gamma_n^* d\omega,
$$

$$
T_n(\theta_{\mathrm{o}}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \begin{bmatrix} \Gamma_n & \Gamma_n \end{bmatrix} \begin{bmatrix} -L_{\mathrm{o}}^* & 0 \\ 0 & F_{\mathrm{o}}^* \end{bmatrix} \begin{bmatrix} \Gamma_n^* & 0 \\ 0 & \Gamma_n^* \end{bmatrix} d\omega,
$$

$$
\bar{R}^n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \begin{bmatrix} \Gamma_n & 0 \\ 0 & \Gamma_n \end{bmatrix} \begin{bmatrix} -G_{\mathrm{o}} S_{\mathrm{o}} F_r & -H_{\mathrm{o}} S_{\mathrm{o}} \sigma_{\mathrm{o}} \\ S_{\mathrm{o}} F_r & -K H_{\mathrm{o}} S_{\mathrm{o}} \sigma_{\mathrm{o}} \end{bmatrix}
$$

$$
\cdot \begin{bmatrix} -G_{\mathrm{o}} S_{\mathrm{o}} F_r & -H_{\mathrm{o}} S_{\mathrm{o}} \sigma_{\mathrm{o}} \\ S_{\mathrm{o}} F_r & -K H_{\mathrm{o}} S_{\mathrm{o}} \sigma_{\mathrm{o}} \end{bmatrix}^* \begin{bmatrix} \Gamma_n^* & 0 \\ 0 & \Gamma_n^* \end{bmatrix} d\omega.
$$

Arguments are omitted for notational simplicity, but functions of q in the time domain should be evaluated at $q = e^{i\omega}$. Using these expressions, we may write $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}})$ as

$$
\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) =
$$
$$
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle]^{-1} \langle \Psi_n, \gamma \rangle,
\tag{26}
$$

where

$$
\Omega_n = \begin{bmatrix} -G_{\mathrm{o}}(q) S_{\mathrm{o}}(q) \Gamma_n(q) & -\sigma_{\mathrm{o}} H_{\mathrm{o}}(q) S_{\mathrm{o}}(q) \Gamma_n(q) \\ S_{\mathrm{o}}(q) F_r(q) \Gamma_n(q) & -\sigma_{\mathrm{o}} K(q) H_{\mathrm{o}}(q) S_{\mathrm{o}}(q) \Gamma_n(q) \end{bmatrix},
$$
$$
\Psi_n = \frac{F_{\mathrm{o}}^*(q)}{S_{\mathrm{o}}^*(q) F_r^*(q)} \begin{bmatrix} \Gamma_n(q) & 0 \end{bmatrix},
$$
$$
\gamma = \frac{S_{\mathrm{o}}(q) F_r(q)}{F_{\mathrm{o}}(q)} \begin{bmatrix} -B_{\mathrm{o}}(q) \Gamma_{m_f}(q) & 0 \\ A_{\mathrm{o}}(q) \Gamma_{m_l}(q) & 0 \end{bmatrix}.
\tag{27}
$$

Following the approach by Hjalmarsson and Mårtensson (2011), we recognize that the term

$$
\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle
$$

in (26) can be written as

$$
\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle = \langle \mathrm{Proj}_{\mathcal{S}_{\Omega_n}} \Psi_n, \mathrm{Proj}_{\mathcal{S}_{\Omega_n}} \Psi_n \rangle,
$$

where $\mathrm{Proj}_{\mathcal{S}_{\Omega_n}} \Psi_n$ denotes the projection of the rows of Ψ_n onto the subspace spanned by the rows of Ω_n. As $n \to \infty$, the dimensions of the matrix Ω_n increase, and the subspace spanned by its rows approaches $\mathcal{H}_2$. Then, the limit value of the projection will be the causal part of the projected matrix.

For a simplified case, suppose that Ψ_n were causal and that its dimension did not depend on n (i.e., $\Psi_n = \Psi$). In this case, we would have $\lim_{n \to \infty} \langle \mathrm{Proj}_{\mathcal{S}_{\Omega_n}} \Psi, \mathrm{Proj}_{\mathcal{S}_{\Omega_n}} \Psi \rangle = \langle \Psi, \Psi \rangle$. In turn, we would then have that $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) = \langle \gamma, \Psi \rangle \langle \Psi, \Psi \rangle \langle \Psi, \gamma \rangle = \langle \mathrm{Proj}_{\mathcal{S}_\Psi} \gamma, \mathrm{Proj}_{\mathcal{S}_\Psi} \gamma \rangle$. If we now reintroduce that the dimension of Ψ depends on n ($\Psi = \Psi_n$), and we assume that the rows of Ψ_n span $\mathcal{H}_2$ as $n \to \infty$, we have that $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) = \lim_{n \to \infty} \langle \mathrm{Proj}_{\mathcal{S}_{\Psi_n}} \gamma, \mathrm{Proj}_{\mathcal{S}_{\Psi_n}} \gamma \rangle = \langle \gamma, \gamma \rangle$. These arguments follow from results in Hjalmarsson and Mårtensson (2011). However, handling the dimensional increase of Ψ_n with n requires additional technical developments. One of the key results for the asymptotic analysis in this paper is that the aforementioned result (i.e., that $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) = \langle \gamma, \gamma \rangle$) still holds when the dimensions of Ψ_n increase with n. This is considered in the following theorem.

Theorem 2 *Let*

$$\Omega_n = \begin{bmatrix} F_1(e^{i\omega})\Gamma_n(e^{i\omega}) & F_2(e^{i\omega})\Gamma_n(e^{i\omega}) \\ F_3(e^{i\omega})\Gamma_n(e^{i\omega}) & F_2(e^{i\omega})F_4(e^{i\omega})\Gamma_n(e^{i\omega}) \end{bmatrix},$$

$$\Psi_n = \begin{bmatrix} F_5^*(e^{i\omega})\Gamma_n(e^{i\omega}) & 0 \end{bmatrix}, \quad \gamma = \begin{bmatrix} F_6(e^{i\omega})\Gamma_{m_f}(e^{i\omega}) & 0 \\ F_7(e^{i\omega})\Gamma_{m_l}(e^{i\omega}) & 0 \end{bmatrix},$$

where $F_j(q) = \sum_{k=0}^{\infty} f_k^{(j)} q^{-k}$ ($j = \{1,...,6\}$) are exponentially stable (i.e., $|f_k^{(j)}| < C\lambda^k \ \forall j, \ \lambda < 1$), $f_0^{(4)} \neq 0$, and $[F_3(q) - F_1(q)F_4(q)]^{-1}$ and $F_5^{-1}(q)$ are exponentially stable. Then, if there is $\bar{n}$ such that $\sigma_{\min}(\langle \Omega_n, \Omega_n \rangle) < C$ ($\sigma_{\min}$ denotes the smallest singular value) for all $n > \bar{n}$,

$$\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle]^{-1} \langle \Psi_n, \gamma \rangle$$
$$= \langle \gamma, \gamma \rangle. \quad (28)$$

PROOF. See Appendix B. □

4.3 Consistency and Asymptotic Covariance

Using the results derived above, we can show the asymptotic properties of semi-parametric WNSF. Regarding consistency of Step 3 in Algorithm 1, we have the following result.

Theorem 3 *Let Assumptions 1, 2, 3, and 4 hold. Then,*

$$\hat{\theta}_N^{WLS} \to \theta_o, \ as \ N \to \infty, \ w.p.1.$$

PROOF. See Appendix C. □

Theorem 3 implies that semi-parametric WNSF provides consistent estimates of θ_o.

Regarding the asymptotic distribution and covariance of Step 3 in Algorithm 1, we have the following result.

Theorem 4 *Let Assumptions 1, 2, 3, and 4 hold. Then,*

$$\sqrt{N}(\hat{\theta}_N^{WLS} - \theta_o) \sim As\mathcal{N}(0, \sigma_o^2 M^{-1}),$$

where M is given by (11).

PROOF. See Appendix D.

As consequence of Theorem 4, the semi-parametric WNSF method summarized in Algorithm 1 has the same asymptotic distribution and covariance as PEM with an infinite-order noise model (Ljung, 1999; Forssell and Ljung, 1999). In open loop and for Gaussian noise, this corresponds to an asymptotically efficient estimate.

5 Simulations

In this section, we perform three simulation examples. In the first, we illustrate the asymptotic properties of the method. In the second, we use a scenario where a low-order parametrization for the noise model does not capture the noise spectrum accurately, which is specially advantageous for semi-parametric WNSF compared with PEM with different choices of noise model. In the third, we motivate the advantage of semi-parametric WNSF compared to the fully parametric version when using the previous scenario.

5.1 Illustration of asymptotic properties

As consequence of the results in Section 4, semi-parametric WNSF is asymptotically efficient in open loop for Gaussian noise, with the asymptotic covariance of the dynamic-model estimates given by $\sigma_o^2 M^{-1}$. In closed loop, the asymptotic covariance is still given by $\sigma_o^2 M^{-1}$, but in this case it does not correspond to the CR bound, but to the optimal asymptotic covariance when the noise-model order tends to infinity.

To illustrate this, we perform open- and closed-loop simulations such that the closed-loop data are generated by

$$u_t = \frac{1}{1 + K(q)G_o(q)} r_t - \frac{K(q)H_o(q)}{1 + K(q)G_o(q)} e_t,$$
$$y_t = \frac{G_o(q)}{1 + K(q)G_o(q)} r_t + \frac{H_o(q)}{1 + K(q)G_o(q)} e_t,$$

and the open-loop data by

$$u_t = \frac{1}{1 + K(q)G_o(q)} r_t,$$
$$y_t = G_o(q)u_t + H_o(q)e_t,$$

where $\{r_t\}$ and $\{e_t\}$ are independent Gaussian white sequences with unit variance, $K(q) = 1$, and

$$G_o(q) = \frac{q^{-1} + 0.1q^{-2}}{1 - 0.5q^{-1} + 0.75q^{-2}}, \quad H_o(q) = \frac{1 + 0.7q^{-1}}{1 - 0.9q^{-1}}.$$

We perform 1000 Monte Carlo runs, with sample sizes $N \in \{300, 600, 1000, 3000, 6000, 10000\}$. We apply WNSF with an ARX model of order 50 to both the open- and closed-loop data. Performance is evaluated by the mean-squared error of the estimated parameter
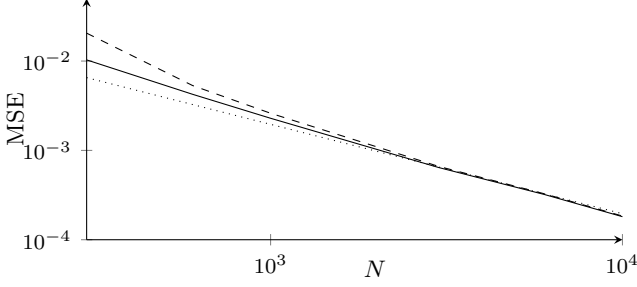
Fig. 1. Illustration of asymptotic properties: theoretical asymptotic MSE (dotted) and average MSE for the parameter estimates as function of sample size obtained with semi-parametric WNSF in closed loop (solid) and open loop (dashed).

vector of the dynamic model,

$$\text{MSE} = ||\hat{\theta}_N^{\text{WLS}} - \bar{\theta}_\text{o}||^2,$$

As this simulation has the purpose of illustrating asymptotic properties, initial conditions are assumed known— that is, the sums in (13) start at $t = 1$ instead of $t = n+1$.

The results are presented in Fig. 1, with the average MSE plotted as function of the sample size (closed loop in solid line, open loop in dashed line). We plot also $\sigma_\text{o}^2 \text{Trace}[M^{-1}]/N$ (dotted line), which the average MSE attains both in open and closed loop: because the data were generated such that Φ_u^r, the spectrum of (10), is the same for both data sets, both scenarios have the same asymptotic covariance, in accordance to our theoretical results.

5.2 Random noise model

When a low-order parametrization of the noise model is not enough to capture the noise spectrum, the noise model may require many parameters. With PEM, a simultaneous estimate of the dynamic model and a long noise model is not numerically robust due to the non-convexity of the cost function. The semi-parametric WNSF is appropriate to deal with this scenario, because the noise spectrum is captured beforehand with a non-parametric ARX model.

Modeling the correct noise spectrum is particularly important in closed loop, as the estimates of the dynamic model will be inconsistent if the noise model is not flexible enough to capture the noise spectrum. For this reason, we consider a closed-loop setting, where data are generated by

$$u_t = \frac{1}{1 + K(q)G_\text{o}(q)}r_t - \frac{K(q)H_\text{o}(q)}{1 + K(q)G_\text{o}(q)}e_t$$
$$y_t = \frac{G_\text{o}(q)}{1 + K(q)G_\text{o}(q)}r_t + \frac{H_\text{o}(q)}{1 + K(q)G_\text{o}(q)}e_t.$$

The signals $\{r_t\}$ and $\{e_t\}$ are Gaussian white noise sequences with variances 1 and 4, respectively. The system is given by

$$G_\text{o}(q) = \frac{1.0q^{-1} - 0.80q^{-2}}{1 - 0.95q^{-1} + 0.90q^{-2}},$$

the controller by $K(q) = 0.2$, and the true noise model by

$$H_\text{o}(q) = 1 + \sum_{k=1}^{N-1} \lambda_k q^{-k}$$

with $\lambda_k = w_k e^{-0.2k}$, where w_k is drawn from a Gaussian distribution with zero mean and unit variance. Here, differently than Assumption 1, stability of $H_\text{o}(q)$ is not ensured. However, this is not an issue if the noise is Gaussian, as there always exists an inversely stable $H_\text{o}(q)$ for which the noise sequence has the same spectrum.

To model the noise, one possibility is to try to find an appropriate low-order parametrization in the form

$$H(q, \zeta) = \frac{c_1 q^{-1} + \cdots + c_{m_h} q^{-m_h}}{d_1 q^{-1} + \cdots + d_{m_h} q^{-m_h}}, \qquad (29)$$

where $\zeta = [c_1 \ \dots \ c_{m_h} \ d_1 \ \dots \ d_{m_h}]'$. In this case, a Box-Jenkins model is estimated. The most appropriate noise model order may be chosen by using some information criterion, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Ljung, 1999).

The alternative is to use a high-order model for the noise model. For example,

$$H(q, \zeta^n) = 1 + \sum_{k=1}^{n} \zeta_k q^{-k}, \qquad (30)$$

or

$$H(q, \zeta^n) = \frac{1}{1 + \sum_{k=1}^{n} \zeta_k q^{-k}}, \qquad (31)$$

where $\zeta^n = [\zeta_1 \ \dots \ \zeta_n]'$. In particular, the choice (31) has the same structure as the noise model estimated by semi-parametric WNSF.

Motivated by these alternatives, we compare the following methods:

- semi-parametric WNSF, with non-parametric ARX model order $n = 200$ (denoted WNSF_sp);
- PEM, with default MATLAB initialization, and noise model (29) with $m_h \in \{1, 2, ...30\}$, where the order m_h is chosen using the AIC or BIC criterion (denoted PEM_aic and PEM_bic, respectively);
- PEM, with noise model (31) with $n = 200$, and MATLAB default initialization (denoted PEM_sp);

9

- PEM, with noise model (31) with $n = 200$, and initialized by WNSF$_{sp}$ (denoted PEM$_{spWNSF}$);

PEM uses the implementation in MATLAB2016b System Identification Toolbox. All the methods use a maximum of 100 iterations, but stop early upon convergence (default settings for PEM, 10^{-4} as tolerance for the normalized relative change in the parameter estimates for WNSF). The search algorithm used by PEM is chosen automatically. The noise model (30) was not used with PEM for computational reasons: the optimization becomes extremely slow as stability of the inverse of the noise model when computing the prediction errors (7) is difficult to fulfill with so many parameters, while the inverse of any estimate of (31) is always stable.

We use sample sizes $N \in \{1000, 5000, 10000\}$ and perform 100 Monte Carlo runs. Performance is evaluated by the FIT of the impulse response of the dynamic model, given by

$$\text{FIT} = 100 \left(1 - \frac{\|g_{o} - \hat{g}\|}{\|g_{o} - \text{mean}(g_{o})\|} \right), \qquad (32)$$

in percent, where g_{o} is a vector with the impulse response parameters of $G_{o}(q)$ (mean(g_{o}) is its average), and similarly for $\hat{g}$ but for the estimated model. In (32), sufficiently long impulse responses are taken to make sure that the truncation of their tails does not affect the FIT.

The FITs for the different sample sizes are shown in Fig. 2. For $N = 1000$, semi-parametric WNSF has better performance than PEM with default MATLAB initialization, both with low-order noise model chosen with AIC/BIC and with non-parametric noise model. Among the PEM alternatives initialized by default with MATLAB, an AIC/BIC criterion with a Box-Jenkins model with orders up to 30 performed better than using a non-parametric noise model. However, if initialized with semi-parametric WNSF, PEM with non-parametric noise model performs considerably better than with default MATLAB initialization. Nevertheless the initializing estimate WNSF$_{sp}$ may in sometimes be better than the resulting estimate PEM$_{spWNSF}$, which can be due to problems with over-fitting. For $N = 5000$, WNSF and PEM with non-parametric noise and default MATLAB initialization have similar median performance, but the algorithm for PEM failed more often. This can be remedied by initializing PEM with WNSF$_{sp}$. Here, PEM with AIC/BIC had no low-performance outliers, but the median performance was poorer than for the semi-parametric approaches. Similar conclusions can be drawn for $N = 10000$, where PEM does not necessarily provide better results with more data samples, potentially due to numerical problems.

Overall, WNSF showed more robust performance among the sample sizes used than the different variants of PEM
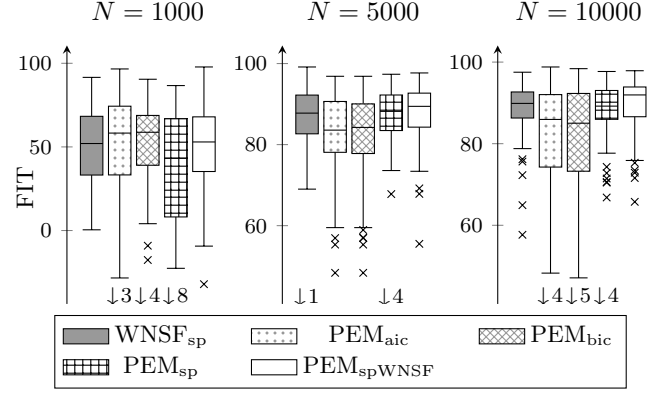

Fig. 2. FITs for given methods and sample sizes from 100 Monte Carlo runs.

Table 1
Average computational times in seconds for WNSF, the search among all orders required for PEM$_{aic}$ and PEM$_{bic}$, and PEM$_{np}$.

N	WNSF$_{sp}$	PEM$_{aic,bic}$	PEM$_{np}$	PEM$_{spWNSF}$
1000	1.0	30	641	83
5000	0.91	27	133	42
10000	1.3	38	236	82

with MATLAB initialization. However, a more evident advantage is the computational time. Table 1 shows the average times, in seconds, required for the identification using semi-parametric WNSF, PEM with all the orders computed for AIC and BIC, PEM$_{sp}$, and PEM$_{spWNSF}$, for the different sample sizes (all the computations were performed in the same computer). Here, we observe how WNSF requires much lower computational time than the alternatives. This is a consequence of PEM estimating the noise model in the non-linear optimization procedure, whereas in WNSF the high-order model is estimated in a previous least-squares step, which is numerically robust. Moreover, the time required for WNSF and PEM with AIC/BIC does not change significantly with N, unlike with PEM$_{sp}$. In this case, the time does not necessarily decrease for smaller N. The problem arising when using smaller N is that the cost function more likely becomes ill-conditioned at some parameter values during the optimization. The time required for PEM with non-parametric noise model decreases when semi-parametric WNSF is used for initialization. However, the times are still significantly larger than for semi-parametric WNSF, while the improvement in performance in not considerable (Fig. 2).

5.3 Comparison with fully-parametric WNSF

In the following, we motivate the advantage of using semi-parametric WNSF compared to the fully-parametric version when the noise cannot be modeled
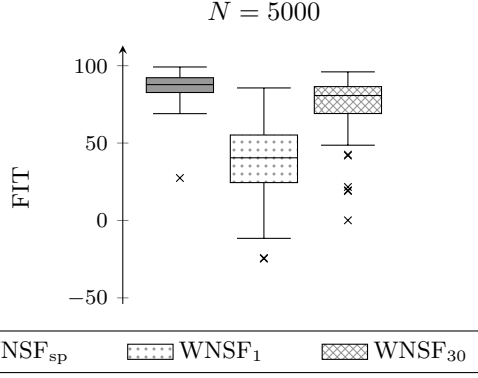
Fig. 3. FITs for semi-parametric and fully-parametric WNSF from 100 Monte Carlo runs.

with a low-order parametrization. In closed loop, and unlike PEM, fully-parametric WNSF should provide consistent estimates of the dynamic model even with an under-parametrized noise model (although this case is not covered in the analysis by Galrinho et al. (2018), it should follow from a similar approach). In this sense, to obtain consistent estimates, fully-parametric WNSF can be used with any choice of noise model. Nevertheless, this does not render semi-parametric WNSF useless, as we will see with the following illustration.

Consider, besides the WNSF$_{sp}$ result from the previous simulation, the following methods:

- fully-parametric WNSF with noise model (29) and order $m_h = 1$ (denoted WNSF$_1$);
- fully-parametric WNSF with noise model (29) and order $m_h = 30$ (denoted WNSF$_{30}$).

The number of iterations and stopping criterion is as in the previous simulation, and we use $N = 5000$.

The FIT results are presented in Fig. 3. The fully-parametric WNSF, both with noise-model order $m_h = 1$ and $m_h = 30$, performs worse than the semi-parametric version. In the case that $m_h = 1$, this may be due to wrong noise model coefficients being included in the weighting, which, despite not affecting consistency of the dynamic-model estimates, may affect the results for finite sample sizes. For $m_h = 30$, the same reasoning may still apply; in addition, because 30 estimated coefficients of the noise model are included in the weighting, which may have high variance, this may also deteriorate the dynamic-model estimates for finite sample sizes. These results suggest that when a high-order noise model is required, it may be better to use semi-parametric WNSF than the fully-parametric version.

6 Conclusion

Many standard system identification methods provide inconsistent estimates with closed-loop data. In the particular case of PEM, this issue is avoided by choosing

a noise-model order that is large enough to capture the noise spectrum. An appropriate order is often difficult to choose, and making it arbitrarily large increases the numerical problems of PEM. The WNSF method, analyzed by Galrinho et al. (2018) in the fully parametric setting, can also be used without a parametric noise-model estimate, which we named semi-parametric WNSF. In this paper, we deepened this discussion.

A simulation study illustrates the importance of separating the dynamic- and noise-model identification when a high-order noise model is required, both in terms of performance and computational time. With WNSF, this separation always occurs, as the method first estimates a non-parametric ARX model. Then, a parametric noise model does not need to be obtained, as the noise spectrum has been captured in the first step.

We also provide a theoretical analysis of the asymptotic properties. To this end, we extend the geometric approach by Hjalmarsson and Mårtensson (2011), deriving a more general result with the matrix dimensions in the inner projection tending to infinity. We show that semi-parametric WNSF provides consistent estimates of the dynamic model. With open-loop data, the estimates are also asymptotically efficient; with closed-loop data, the asymptotic covariance of the estimates corresponds to the best possible covariance with a non-parametric noise model. This gives WNSF attractive features in terms of flexibility of noise-model structure and asymptotic properties: if a correct parametric noise model is estimated, the dynamic-model estimates are asymptotically efficient; if not, they are consistent and optimal for a non-parametric noise model. We used a simulation study to illustrate how semi-parametric WNSF is an appropriate method for scenarios where the noise model cannot be accurately modeled with a low-order parametrization.

A   Auxiliary Results for the Proof of Theorem 2

The following results will be used to prove Theorem 2.

Cauchy-Schwarz inequality for transfer-matrix inner products

Let $X(q)$ and $Y(q)$ be transfer matrices and x and y be vectors of appropriate dimensions. Then, we have

$$
\begin{aligned}
||\langle X, Y \rangle||^2 &= \sup_{||x||=1, ||y||=1} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} x' X(e^{i\omega}) Y^*(e^{i\omega}) y \mathrm{d}\omega \right| \\
&\leq \sup_{||x||=1, ||y||=1} \sqrt{x' \langle X, X \rangle x \; y' \langle Y, Y \rangle y} \\
&= ||\langle X, X \rangle|| \; ||\langle Y, Y \rangle||.
\end{aligned}
\tag{A.1}
$$

Bound for spectral norm of transfer-matrix inner products

Let $X(q)$ be a transfer matrix. Then, we have

$$\begin{aligned}
||\langle X, X \rangle|| &\leq ||\langle X, X \rangle||_F = \sqrt{\text{Trace}(\langle X, X \rangle^2)} \\
&\leq \text{Trace}\langle X, X \rangle = ||X||_{\mathcal{H}_2}^2,
\end{aligned} \quad \text{(A.2)}$$

where the second inequality follows from $\text{Trace}(A^2) \leq [\text{Trace}(A)]^2$ for a positive semi-definite matrix A.

Bound for Toeplitz operators of stable filters

Let $X(q) := \sum_{k=-\infty}^{\infty} x_k q^{-k}$, with $||X(q)||_{\mathcal{H}_\infty} < C$. From Theorem 3.1 by Partington (1989), it follows that

$$\left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* X(e^{i\omega}) \right\| \leq ||X(q)||_{\mathcal{H}_\infty} \ \forall n \in \mathbb{N}. \quad \text{(A.3)}$$

B Proof of Theorem 2

In this appendix, we prove Theorem 2.

Inner Projection: $\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle$

Let

$$\Psi_n = \Psi_n^c + \Psi_n^a, \quad \text{(B.1)}$$

where

$$\Psi_n^c := \begin{bmatrix} f_0^{(5)} e^{-i\omega} & 0 \\ f_0^{(5)} e^{-2i\omega} + f_1^{(5)} e^{-i\omega} & 0 \\ \vdots & \vdots \\ f_0^{(5)} e^{-ni\omega} + \sum_{k=1}^{n-1} f_{n-k}^{(5)} e^{-ki\omega} & 0 \end{bmatrix},$$

$$\Psi_n^a := \begin{bmatrix} \sum_{k=0}^{\infty} f_{k+1}^{(5)} e^{i\omega k} & 0 \\ \sum_{k=0}^{\infty} f_{k+2}^{(5)} e^{i\omega k} & 0 \\ \vdots & \vdots \\ \sum_{k=0}^{\infty} f_{k+n}^{(5)} e^{i\omega k} & 0 \end{bmatrix}.$$

Alternatively, Ψ_n^c can be written as $\Psi_n^c = \begin{bmatrix} P_n \Gamma_n & 0 \end{bmatrix}$, where $P_n = \mathcal{T}_{n \times n}(F_5(q))$. Using (B.1), we can write

$$\begin{aligned}
&\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle \\
&= \langle \Psi_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c \rangle \\
&\quad + \langle \Psi_n^a, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c \rangle \\
&\quad + \langle \Psi_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^a \rangle \\
&\quad + \langle \Psi_n^a, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^a \rangle \\
&= \langle \Psi_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c \rangle
\end{aligned} \quad \text{(B.2)}$$

because $\langle \Psi_n^a, \Omega_n \rangle = 0 = \langle \Omega_n, \Psi_n^a \rangle \ \forall n$, as Ω_n is causal and Ψ_n^a is anti-causal.

Now, we will construct an approximation of Ψ_n^c whose rows can be written using a linear combination of the rows of Ω_n. To facilitate this, we define (arguments are omitted for notational simplicity)

$$\tilde{\Omega}_n := \begin{bmatrix} \Gamma_n F_1 F_4 & \Gamma_n F_3 F_4 \\ \Gamma_n (F_3 - F_1 F_4) & 0 \end{bmatrix},$$

which has been obtained by multiplying the first n rows of Ω_n by F_4, and subtracting the newly obtained first n rows from the last n rows. Then, if it is possible to find a linear combination of the rows of $\tilde{\Omega}_n$ to write Ψ_n^c, the same is possible for Ω_n.

Let $\bar{F}_3 := F_3 - F_1 F_4$, whose inverse is exponentially stable by assumption. In addition, let row i of $\bar{F}_3^{-1} \Psi_n^c$ be given by

$$\bar{F}_3^{-1} \Psi_n^c(i) =: \left[\sum_{k=1}^{\infty} \beta_k^{(i)} q^{-k} \ \ 0 \right],$$

where $|\beta_k^{(i)}| \leq C \lambda^k$ with $\lambda < 1$, and

$$\hat{\Psi}_n^c(i) := \left[\sum_{k=1}^{n} \beta_k^{(i)} \bar{F}_3 q^{-k} \ \ 0 \right]$$

be the row i of a matrix $\hat{\Psi}_n^c(i)$. We re-write the right side of (B.2) as

$$\langle \Psi_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c \rangle = \Lambda_n + \Delta_n^{(1)}, \quad \text{(B.3)}$$

where

$$\begin{aligned}
\Lambda_n =& \langle \hat{\Psi}_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \hat{\Psi}_n^c \rangle \\
\Delta_n^{(1)} =& \langle \Psi_n^c - \hat{\Psi}_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c \rangle \\
&+ \langle \Psi_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c - \hat{\Psi}_n^c \rangle \\
&+ \langle \Psi_n^c - \hat{\Psi}_n^c, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n^c - \hat{\Psi}_n^c \rangle
\end{aligned}$$

By construction, $\hat{\Psi}_n^c(i)$ is a linear combination of the rows of $\tilde{\Omega}_n$ (and hence, of Ω_n); therefore, $\hat{\Psi}_n^c \in \mathcal{S}_{\Omega_n}$ and

$$\Lambda_n = \langle \hat{\Psi}_n^c, \hat{\Psi}_n^c \rangle. \quad \text{(B.4)}$$

Using (B.2), (B.3), and (B.4), we write

$$\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle = \langle \hat{\Psi}_n^c, \hat{\Psi}_n^c \rangle + \Delta_n^{(1)}. \quad \text{(B.5)}$$

Moreover, we can re-write (B.5) as

$$\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle = \langle \Psi_n^c, \Psi_n^c \rangle + \Delta_n, \quad \text{(B.6)}$$

where $\Delta_n = \Delta_n^{(1)} + \Delta_n^{(2)}$ and

$$\Delta_n^{(2)} = \langle \Psi_n^c - \hat{\Psi}_n^c, \Psi_n^c \rangle + \langle \Psi_n^c, \Psi_n^c - \hat{\Psi}_n^c \rangle + \langle \Psi_n^c - \hat{\Psi}_n^c, \Psi_n^c - \hat{\Psi}_n^c \rangle.$$

Replacing (B.6) in (28), we obtain

$$\lim_{n\to\infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle]^{-1} \langle \Psi_n, \gamma \rangle$$
$$= \lim_{n\to\infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n^c, \Psi_n^c \rangle + \Delta_n]^{-1} \langle \Psi_n, \gamma \rangle, \quad \text{(B.7)}$$

and using the Sherman-Morrison-Woodbury formula, we re-write (B.7) as

$$\lim_{n\to\infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n^c, \Psi_n^c \rangle + \Delta_n]^{-1} \langle \Psi_n, \gamma \rangle$$
$$= \lim_{n\to\infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle$$
$$+ \lim_{n\to\infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \Delta_n$$
$$\cdot [I + \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \Delta_n]^{-1} \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle.$$
$$\text{(B.8)}$$

We want to show that the second term on the right-hand side of (B.8), for which we can write

$$||\langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \Delta_n [I + \langle \Psi_n^c, \Psi_n^c \rangle \Delta_n]^{-1}$$
$$\cdot \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle||$$
$$\leq ||\langle \gamma, \Psi_n \rangle||^2 \, ||\langle \Psi_n^c, \Psi_n^c \rangle^{-1}||^2$$
$$\cdot ||[I + \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \Delta_n]^{-1}|| \, ||\Delta_n||,$$
$$\text{(B.9)}$$

tends to zero as n tends to infinity. We start by considering the term Δ_n, for which we will need that

$$\langle \Psi_n^c, \Psi_n^c \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_n \Gamma_n \Gamma_n^* P_n' d\omega = P_n P_n'. \quad \text{(B.10)}$$

Using also (A.1) and (A.2), we can write

$$||\Delta_n|| \leq 2||\langle \Omega_n, \Omega_n \rangle|| \, ||\langle \Omega_n, \Omega_n \rangle^{-1}|| \, ||P_n|| \, ||\Psi_n^c - \hat{\Psi}_n^c||_{\mathcal{H}_2}$$
$$+ ||\langle \Omega_n, \Omega_n \rangle|| \, ||\langle \Omega_n, \Omega_n \rangle^{-1}|| \, ||\Psi_n^c - \hat{\Psi}_n^c||_{\mathcal{H}_2}^2$$
$$+ 2||P_n|| \, ||\Psi_n^c - \hat{\Psi}_n^c||_{\mathcal{H}_2} + ||\Psi_n^c - \hat{\Psi}_n^c||_{\mathcal{H}_2}^2. \quad \text{(B.11)}$$

For row i of $\Psi_n^c - \hat{\Psi}_n^c$, we have that

$$||\Psi_n^{c,i} - \hat{\Psi}_n^{c,i}|| = ||\bar{F}_3 \sum_{k=1}^{\infty} \beta_k^{(i)} q^{-k} - \sum_{k=1}^{n} \beta_k^{(i)} \bar{F}_3 q^{-k}||$$
$$\leq ||\bar{F}_3|| \, ||\sum_{k=n+1}^{\infty} \beta_k^{(i)} q^{-k}|| \leq C\lambda^n$$

Then,

$$||\Psi_n^c - \hat{\Psi}_n^c||_{\mathcal{H}_2} \leq C\sqrt{n}\lambda^n \to 0 \text{ as } n \to \infty. \quad \text{(B.12)}$$

Moreover, using (A.1), we have

$$||P_n|| = \left|\left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_5^* d\omega \right|\right| \leq ||F_5^*||_{\mathcal{H}_\infty} \leq C, \quad \text{(B.13)}$$

and, using (A.3) and

$$\langle \Omega_n, \Omega_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi}$$
$$\begin{bmatrix} \Gamma_n \Gamma_n^*(|F_1|^2 + |F_2|^2) & \Gamma_n \Gamma_n^*(F_1 F_3^* + |F_2|^2 F_4^*) \\ \Gamma_n \Gamma_n^*(F_3 F_1^* + |F_2|^2 F_4) & \Gamma_n \Gamma_n^*(|F_3|^2 + |F_2 F_4|^2) \end{bmatrix} d\omega,$$

we have

$$||\langle \Omega_n, \Omega_n \rangle|| \leq ||\,|F_1|^2 + |F_2|^2||_{\mathcal{H}_\infty}$$
$$+ 2||F_1 F_3^* + |F_2|^2 F_4^*||_{\mathcal{H}_\infty}$$
$$+ ||\,|F_3|^2 + |F_2 F_4|^2||_{\mathcal{H}_\infty}$$
$$\leq C.$$
$$\text{(B.14)}$$

By assumption, it follows from $\sigma_{\min}(\langle \Omega_n, \Omega_n \rangle) < C$ that $||\langle \Omega_n, \Omega_n \rangle^{-1}|| < C$. Together with (B.11), (B.12), and (B.14), we have

$$||\Delta_n|| \to 0 \text{ as } n \to \infty. \quad \text{(B.15)}$$

Then, if the remaining matrix norms in the right-hand side of (B.9) are bounded, this term will tend to zero as $n \to \infty$. For $\langle \gamma, \Psi_n \rangle$, we have

$$||\langle \gamma, \Psi_n \rangle|| = \left|\left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \begin{bmatrix} \Gamma_n \Gamma_n^* F_5 F_6 \\ \Gamma_n \Gamma_n^* F_5 F_7 \end{bmatrix} d\omega \right|\right|$$
$$\leq ||F_5 F_6||_{\mathcal{H}_\infty} + ||F_5 F_7||_{\mathcal{H}_\infty} \leq C. \quad \text{(B.16)}$$

Also, we have that

$$||\langle \Psi_n^c, \Psi_n^c \rangle^{-1}|| = P_n^{-'} P_n^{-1}$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_5^{-1} d\omega \, \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_5^{-*} d\omega$$
$$\leq ||F_5||_{\mathcal{H}_\infty}^2 \leq C.$$
$$\text{(B.17)}$$

Together with (B.8), (B.9), (B.15), and (B.16), we have

$$\lim_{n\to\infty} \langle \gamma, \Psi_n \rangle [\langle \Psi_n, \Omega_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle]^{-1} \langle \Psi_n, \gamma \rangle$$
$$= \lim_{n\to\infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle. \quad \text{(B.18)}$$

Outer projection: $\langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle$

Recalling that Ψ_n can be written as (B.1), we use that γ is causal and Ψ_n^a anti-causal to conclude that, analogously to (B.2),

$$\lim_{n\to\infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n, \gamma \rangle$$
$$= \lim_{n\to\infty} \langle \gamma, \Psi_n^c \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} \langle \Psi_n^c, \gamma \rangle. \quad \text{(B.19)}$$

Now, row i of γ can be written as

$$\gamma(i) =: \left[\begin{array}{cc} \sum_{k=1}^{\infty} s_k^{(i)} q^{-k} & 0 \end{array}\right],$$

where $|s_k^{(i)}| \leq C\lambda^k$, $\lambda < 1$ due to exponential stability. Let also $\gamma^n(i) := \left[\begin{array}{cc} \sum_{k=1}^{n} s_k^{(i)} q^{-k} & 0 \end{array}\right]$ be row i of a matrix γ^n. We re-write the right side of (B.19) as

$$
\begin{aligned}
&\lim_{n\to\infty} \langle\gamma, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma\rangle \\
&= \lim_{n\to\infty} \langle\gamma^n, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma^n\rangle + \lim_{n\to\infty} \Delta_n^{(3)},
\end{aligned}
$$
(B.20)

where

$$
\begin{aligned}
\Delta_n^{(3)} &= \langle\gamma - \gamma^n, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma\rangle \\
&\quad + \langle\gamma, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma - \gamma^n\rangle \\
&\quad + \langle\gamma - \gamma^n, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma - \gamma^n\rangle.
\end{aligned}
$$

Using a similar approach for $\Delta_n^{(3)}$ as we did for Δ_n, it can be shown that

$$||\Delta_n^{(3)}|| \leq C(||\gamma - \gamma^n||_{\mathcal{H}_2} + ||\gamma - \gamma^n||_{\mathcal{H}_2}^2) \to 0 \text{ as } n \to \infty.$$
(B.21)

Thus, (B.20) reduces to

$$
\begin{aligned}
&\lim_{n\to\infty} \langle\gamma, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma\rangle \\
&= \lim_{n\to\infty} \langle\gamma^n, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma^n\rangle.
\end{aligned}
$$
(B.22)

Finally, we have that

$$
\begin{aligned}
\langle\gamma^n, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma^n\rangle &= \langle\mathrm{Proj}_{\mathcal{S}_{\Psi_n^c}}\gamma^n, \mathrm{Proj}_{\mathcal{S}_{\Psi_n^c}}\gamma^n\rangle \\
&= \langle\gamma^n, \gamma^n\rangle,
\end{aligned}
$$
(B.23)

where the last equality follows from $\gamma^n \in \mathcal{S}_{\Psi_n^c}$, as consequence of P_n being invertible. Thus, replacing (B.23) in (B.22), we have

$$\lim_{n\to\infty} \langle\gamma, \Psi_n^c\rangle\langle\Psi_n^c, \Psi_n^c\rangle^{-1}\langle\Psi_n^c, \gamma\rangle = \lim_{n\to\infty} \langle\gamma^n, \gamma^n\rangle = \langle\gamma, \gamma\rangle,$$

which, together with (B.18), (B.19), (B.20), and (B.21) implies (28), as we wanted to show. □

C Proof of Theorem 3

We will show that

$$||\hat{\theta}_N^{\mathrm{WLS}} - \theta_o|| \to 0, \text{ as } N \to \infty, \text{ w.p.1.} \quad \text{(C.1)}$$

To do this, we use (24) to write

$$
\begin{aligned}
\left\|\hat{\theta}_N^{\mathrm{WLS}} - \theta_o\right\| &\leq \left\|M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})\right\| \|Q_n(\hat{\eta}_N)\| \\
&\quad \cdot \left\|W_n(\hat{\theta}_N^{\mathrm{LS}})\right\| \|T_n(\theta_o)\| \left\|\hat{\eta}_N - \eta_o^{n(N)}\right\|.
\end{aligned}
$$
(C.2)

From Galrinho et al. (2018), we have that

$$||\hat{\eta}_N - \eta_o^{n(N)}|| \to 0 \text{ as } N \to \infty, \text{ w.p.1}, \quad \text{(C.3)}$$
$$||T_n(\theta_o)|| \leq C \quad \forall n, \quad \text{(C.4)}$$

and that, w.p.1, there exists $\bar{N}$ such that

$$||Q_n(\hat{\eta}_N)|| \leq C \quad \forall N > \bar{N}. \quad \text{(C.5)}$$

Then, we have left to show that $W_n(\hat{\theta}_N^{\mathrm{LS}})$ is bounded and $M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ is invertible for sufficiently large N.

We begin by considering the inverse of $W_n(\hat{\theta}_N^{\mathrm{LS}})$, for which we have

$$
\begin{aligned}
||T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}})|| &\leq ||T_n(\theta_o)[\bar{R}^n]^{-1}T_n'(\theta_o)|| \\
&\quad + ||T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_o)[\bar{R}^n]^{-1}T_n'(\theta_o)||.
\end{aligned}
$$
(C.6)

In turn, it can be shown that

$$T_n(\theta_o)[\bar{R}^n]^{-1}T_n'(\theta_o) = \langle\Psi_n, \Omega_n\rangle\langle\Omega_n, \Omega_n\rangle^{-1}\langle\Omega_n, \Psi_n\rangle,$$

where Ψ_n and Ω_n are given by (27), which satisfy the conditions of Theorem 2 with

$$
\begin{aligned}
F_1(q) &= -G_o(q)S_o(q)F_r(q), &\quad F_2(q) &= -H_o(q)S_o(q)\sigma_o, \\
F_3(q) &= S_o(q)F_r(q), &\quad F_4(q) &= K(q).
\end{aligned}
$$

In particular, there is $\bar{n}$ such that $\langle\Omega_n, \Omega_n\rangle = \bar{R}^n$ is invertible for all $n > \bar{n}$ (Ljung and Wahlberg, 1992), and $F_3(q) - F_1(q)F_4(q) = F_r(q)$ can be chosen to have a stable inverse under the constraint $\Phi_r(e^{i\omega}) = |F_r(e^{i\omega})|^2$. From (B.6), (B.10), (B.13), and (B.15), we have that there is $\bar{n}$ such that

$$||T_n(\theta_o)[\bar{R}^n]^{-1}T_n'(\theta_o)|| = ||\langle\Psi_n^c, \Psi_n^c\rangle + \Delta_n|| \leq C \; \forall n > \bar{n}. \quad \text{(C.7)}$$

Then, and using also (B.17), we have

$$
\begin{aligned}
||\bar{W}(\theta_o)|| &:= ||[T_n(\theta_o)[\bar{R}^n]^{-1}T_n'(\theta_o)]^{-1}|| \\
&= ||(\langle\Psi_n^c, \Psi_n^c\rangle + \Delta_n)^{-1}|| \leq C \quad \forall n > \bar{n}.
\end{aligned}
$$
(C.8)

Concerning the second term on the right-hand side

of (C.6), we can write

$$||T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})[\bar{R}^n]^{-1}T_n'(\theta_{\mathrm{o}})||$$
$$\leq ||T_n(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})|| \, ||[R_N^n]^{-1}|| \, ||T_n(\hat{\theta}_N^{\mathrm{LS}})||$$
$$+ ||T_n(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})|| \, ||[R_N^n]^{-1}|| \, ||T_n(\theta_{\mathrm{o}})||$$
$$+ ||T_n(\theta_{\mathrm{o}})||^2 \, ||[R_N^n]^{-1} - [\bar{R}^n]^{-1}|| \qquad (\mathrm{C}.9)$$

Now, the results by Galrinho et al. (2018) apply to (C.9). In particular, there is $\bar{N}$ such that

$$||[R_N^n]^{-1}|| \leq C \quad \forall N > \bar{N}, \qquad ||[R_N^n]|| \leq C \quad \forall N > \bar{N},$$
$$||T_n(\hat{\theta}_N^{\mathrm{LS}})|| \leq C \quad \forall N > \bar{N} \qquad (\mathrm{C}.10)$$

w.p.1, and that

$$||T_n(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})|| \to 0 \text{ as } N \to \infty, \text{ w.p.1},$$
$$||[R_N^n]^{-1} - [\bar{R}^n]^{-1}|| \leq ||\bar{R}_N^n|| \, ||\bar{R}^n - R_N^n|| \, ||[R_N^n]^{-1}||$$
$$\to 0 \text{ as } N \to \infty, \text{ w.p.1}.$$

Together with (C.4), we conclude that

$$||T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})[\bar{R}^n]^{-1}T_n'(\theta_{\mathrm{o}})||$$
$$\to 0 \text{ as } N \to \infty, \text{ w.p.1}. \quad (\mathrm{C}.11)$$

Using (C.11), (C.7), and (C.6), there is $\bar{N}$ such that

$$||T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}})|| \leq C \quad \forall N > \bar{N} \quad \text{w.p.1}.$$

Because of (C.11) and invertibility of $T_n(\theta_{\mathrm{o}})[\bar{R}^n]^{-1}T_n'(\theta_{\mathrm{o}})$, by continuity of eigenvalues there is $\bar{N}$ such that $W(\hat{\theta}_N^{\mathrm{LS}}) = [T_n(\hat{\theta}_N^{\mathrm{LS}})[R_N^n]^{-1}T_n'(\hat{\theta}_N^{\mathrm{LS}})]^{-1}$ exists for all $N > \bar{N}$, and

$$||W(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}(\theta_{\mathrm{o}})|| \to 0, \text{ as } N \to \infty, \text{ w.p.1}. \quad (\mathrm{C}.12)$$

Moreover, (C.12) and (C.8) imply that, w.p.1,

$$||W(\hat{\theta}_N^{\mathrm{LS}})|| \leq C \quad \forall N > \bar{N}. \qquad (\mathrm{C}.13)$$

Having shown (C.13), we have left to show that $M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ is invertible for sufficiently large N, in order to show (C.1). Recall the definition (25), which can alternatively be written as (26), where γ is given by (27). Then, from Theorem 2 with

$$F_6(q) = -\frac{B_{\mathrm{o}}(q)S_{\mathrm{o}}(q)F_r(q)}{F_{\mathrm{o}}(q)}, \quad F_7(q) = \frac{A_{\mathrm{o}}(q)S_{\mathrm{o}}(q)F_r(q)}{F_{\mathrm{o}}(q)},$$

we have that
$$\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}}) = M. \qquad (\mathrm{C}.14)$$
where M is given by (11). Because the inverse of M corresponds to the CR bound for an open-loop problem

with input $u_t = S_{\mathrm{o}}(q)r_t$, and the CR bound exists for an informative experiment (Ljung, 1999), we conclude that $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}})$ is invertible. Then, we analyze the difference

$$||M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) - Q_n'(\eta_{\mathrm{o}}^n)\bar{W}_n(\theta_{\mathrm{o}})Q_n(\eta_{\mathrm{o}}^n)||$$
$$\leq ||Q_n(\hat{\eta}_N) - Q_n(\eta_N^n)|| \, ||W(\hat{\theta}_N^{\mathrm{LS}})|| \, ||Q_n(\hat{\eta}_N)||$$
$$+ ||Q_n(\hat{\eta}_N) - Q_n(\eta_N^n)|| \, ||W(\hat{\theta}_N^{\mathrm{LS}})|| \, ||Q_n(\eta_{\mathrm{o}}^n)||$$
$$+ ||Q_n(\eta_N^n)||^2 \, ||W(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}(\theta_{\mathrm{o}})||$$

From Galrinho et al. (2018), we have that

$$||Q_n(\hat{\eta}_N) - Q_n(\eta_N^n)|| \to 0, \text{ as } N \to \infty, \text{ w.p.1},$$
$$||Q_n(\eta_{\mathrm{o}}^n)|| \leq C \quad \forall n. \qquad (\mathrm{C}.15)$$

Together with (C.12) and (C.5), we conclude that

$$||M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) - Q_n'(\eta_{\mathrm{o}}^n)\bar{W}_n(\theta_{\mathrm{o}})Q_n(\eta_{\mathrm{o}}^n)||$$
$$\to 0, \text{ as } N \to \infty, \text{ w.p.1}. \quad (\mathrm{C}.16)$$

Using (C.16), invertibility of $\bar{M}(\eta_{\mathrm{o}}, \theta_{\mathrm{o}})$, and continuity of eigenvalues, we have that there is $\bar{N}$ such that $M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ is invertible for all $N > \bar{N}$,

$$||M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})|| \leq C \quad \forall N > \bar{N}, \qquad (\mathrm{C}.17)$$

and, using also (C.14) and (25),

$$M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) \to M^{-1} \text{ as } N \to \infty, \text{ w.p.1}. \quad (\mathrm{C}.18)$$

Finally, using (C.17), (C.13), (C.5), (C.4), (C.3), and (C.2), we conclude that (C.1) is satisfied, as we wanted to show. □

D   Proof of Theorem 4

In this appendix, we prove Theorem 4. We begin by reformulating (24) as

$$\sqrt{N}(\hat{\theta}_N^{\mathrm{WLS}} - \theta_{\mathrm{o}}) = M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}),$$

where

$$M(\hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) = Q_n'(\hat{\eta}_N)W_n(\hat{\theta}_N^{\mathrm{LS}})Q_n(\hat{\eta}_N),$$
$$x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) = \sqrt{N}Q_n'(\hat{\eta}_N)W_n(\hat{\theta}_N^{\mathrm{LS}})T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \eta_{\mathrm{o}}^{n(N)}).$$

If we assume that

$$x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}}) \sim As\mathcal{N}(0, X), \qquad (\mathrm{D}.1)$$

we have that, using (C.18) and Lemma B.4 by Söderström and Stoica (1989),

$$\sqrt{N}(\hat{\theta}_N^{\mathrm{WLS}} - \theta_{\mathrm{o}}) \sim As\mathcal{N}\left(0, M^{-1}XM^{-1}\right). \qquad (\mathrm{D}.2)$$

We will proceed to show that (D.1) is verified with

$$X = \sigma_{\mathrm{o}}^2 \lim_{n\to\infty} Q_n'(\eta_{\mathrm{o}}^n)\bar{W}_n(\theta_{\mathrm{o}})Q_n(\eta_{\mathrm{o}}^n) = \sigma_{\mathrm{o}}^2 M,$$

where the second equality follows directly from (25) and (C.14). We now proceed to show the first equality.

The idea is, as in Galrinho et al. (2018), to use Theorem 7.3 by Ljung and Wahlberg (1992). With this purpose, we first show that $x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ has the same asymptotic distribution and covariance as $\sqrt{N}\Upsilon^n(\hat{\eta}_N - \bar{\eta}^{n(N)})$, where Υ^n is a deterministic matrix. From Galrinho et al. (2018), it follows that $x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ has the same asymptotic covariance and distribution as

$$\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})W_n(\hat{\theta}_N^{\mathrm{LS}})T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \bar{\eta}^{n(N)}). \quad (D.3)$$

The next step is to show that $W_n(\hat{\theta}_N^{\mathrm{LS}})$ can be replaced by $\bar{W}_n(\theta_{\mathrm{o}})$ in (D.3) without affecting the asymptotic distribution and covariance. However, this step does not follow directly from Galrinho et al. (2018), as inverses of the matrices that compose W_n cannot be taken individually here, because T_n is not square. In the following, we do this for a non-square T_n matrix.

First, re-write (D.3) as

$$\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})\bar{W}_n(\theta_{\mathrm{o}})T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \bar{\eta}^{n(N)})$$
$$+\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})[W_n(\hat{\theta}_N^{\mathrm{LS}})-\bar{W}_n(\theta_{\mathrm{o}})]T_n(\theta_{\mathrm{o}})(\hat{\eta}_N-\bar{\eta}^{n(N)}). \quad (D.4)$$

Then, it follows from Lemma B.4 by Söderström and Stoica (1989) that (D.3) and the first term in (D.4) have the same asymptotic distribution and covariance if the second term in (D.4) tends to 0 as $N \to \infty$ w.p.1. We proceed to show this. Consider

$$||\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})[W_n(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n(\theta_{\mathrm{o}})]T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \bar{\eta}^{n(N)})||$$
$$\leq \sqrt{N}||Q_n'(\eta_{\mathrm{o}}^{n(N)})||\ ||W_n(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n(\theta_{\mathrm{o}})||$$
$$\cdot ||T_n(\theta_{\mathrm{o}})||\ ||\hat{\eta}_N - \bar{\eta}^{n(N)}||$$
$$\leq C\sqrt{N}||W_n(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n(\theta_{\mathrm{o}})||\ ||\hat{\eta}_N - \bar{\eta}^{n(N)}||, \quad (D.5)$$

where the last inequality follows from (C.4) and (C.15). Writing

$$W_n(\hat{\theta}_N^{\mathrm{LS}})-\bar{W}_n(\theta_{\mathrm{o}}) = \bar{W}_n(\theta_{\mathrm{o}})[\bar{W}_n^{-1}(\theta_{\mathrm{o}})-W_n^{-1}(\hat{\theta}_N^{\mathrm{LS}})]W_n(\hat{\theta}_N^{\mathrm{LS}}),$$

and because (C.13) and (C.8) guarantee that $\bar{W}_n(\theta_{\mathrm{o}})$ and $W_n(\hat{\theta}_N^{\mathrm{LS}})$ are bounded (in the latter, for sufficiently large N), it follows from (D.5) that

$$||\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})[W_n(\hat{\theta}_N^{\mathrm{LS}})-\bar{W}_n(\theta_{\mathrm{o}})]T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \bar{\eta}^{n(N)})||$$
$$\leq C\sqrt{N}||W_n^{-1}(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n^{-1}(\theta_{\mathrm{o}})||\ ||\hat{\eta}_N - \bar{\eta}^{n(N)}||.$$

Now, the term $||W_n^{-1}(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n^{-1}(\theta_{\mathrm{o}})||$ corresponds to (C.9); so, using (Galrinho et al., 2018)

$$||T_n(\hat{\theta}_N^{\mathrm{LS}}) - T_n(\theta_{\mathrm{o}})|| = \mathcal{O}\left(\sqrt{n^2(N)\frac{\log N}{N}}(1 + d(N))\right), \quad (D.6)$$

(C.4) and (C.10), the first two terms on the right-hand side of (C.9) decay with (D.6). For the third term, we first write

$$||[R_N^n]^{-1} - [\bar{R}^n]^{-1}|| \leq ||\bar{R}_N^n||\ ||\bar{R}^n - R_N^n||\ ||[R_N^n]^{-1}||. \quad (D.7)$$

Ljung and Wahlberg (1992) show that

$$||\bar{R}^n - R_N^n|| = \mathcal{O}\left(2\sqrt{n^2(N)\frac{\log N}{N}} + C\frac{n^2(N)}{N}\right); \quad (D.8)$$

then, using also (D.7), (C.10), and (C.4), we have that the third term on the right-hand side of (C.9) decays according to (D.8). Thus, we have that

$$||W_n(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n(\theta_{\mathrm{o}})|| = \mathcal{O}\left(2\sqrt{n^2(N)\frac{\log N}{N}}\right), \quad (D.9)$$

considering only the slowest-decaying term. Then, from (D.5), (D.9), and (Ljung and Wahlberg, 1992)

$$||\hat{\eta}_N - \bar{\eta}^{n(N)}|| = \mathcal{O}\left(\sqrt{\frac{n(N)\log N}{N}}[1 + d(N)]\right),$$

it follows that

$$C\sqrt{N}||W_n(\hat{\theta}_N^{\mathrm{LS}}) - \bar{W}_n(\theta_{\mathrm{o}})||\ ||\hat{\eta}_N - \bar{\eta}^{n(N)}|| \to 0$$
$$\text{as } N \to \infty, \text{ w.p.1.}$$

Finally, using (D.5) and (D.4), it follows from Lemma B.4 by Söderström and Stoica (1989) that

$$\sqrt{N}Q_n'(\eta_{\mathrm{o}}^{n(N)})\bar{W}_n(\theta_{\mathrm{o}})T_n(\theta_{\mathrm{o}})(\hat{\eta}_N - \bar{\eta}^{n(N)}) \quad (D.10)$$

and (D.3)—and, in turn, $x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$—have the same asymptotic distribution and covariance. Thus, we will analyze (D.10) instead of $x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$.

Applying Theorem 7.3 by Ljung and Wahlberg (1992) to (D.10) with $\Upsilon^n = Q_n'(\eta_{\mathrm{o}}^{n(N)})\bar{W}_n(\theta_{\mathrm{o}})T_n(\theta_{\mathrm{o}})$—and recalling that it has the same asymptotic distribution and covariance as $x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$—we have that

$x(\theta_{\mathrm{o}}; \hat{\eta}_N, \hat{\theta}_N^{\mathrm{LS}})$ is distributed according to (D.1) with

$$
\begin{aligned}
X &= \lim_{n \to \infty} Q_n'(\eta_{\mathrm{o}}^{n(N)}) \bar{W}_n(\theta_{\mathrm{o}}) T_n(\theta_{\mathrm{o}}) \sigma_{\mathrm{o}}^2 \\
&\quad \cdot \bar{R}_N^n T_n'(\theta_{\mathrm{o}}) \bar{W}_n(\theta_{\mathrm{o}}) Q_n(\eta_{\mathrm{o}}^{n(N)}) \\
&= \sigma_{\mathrm{o}}^2 \lim_{n \to \infty} Q_n'(\eta_{\mathrm{o}}^{n(N)}) \bar{W}_n(\theta_{\mathrm{o}}) Q_n(\eta_{\mathrm{o}}^{n(N)}) = \sigma_{\mathrm{o}}^2 M,
\end{aligned}
$$

(D.11)

where the last equality follows from (25). Then, replacing (D.11) in (D.2), we obtain

$$
\sqrt{N}(\hat{\theta}_N^{\mathrm{WLS}} - \theta_{\mathrm{o}}) \sim As\mathcal{N}\left(0, M^{-1}\right). \qquad \square
$$

References

Aguero, J. C. and Goodwin, G. C. (2007). Choosing between open-and closed-loop experiments in linear system identification. *IEEE Transactions on Automatic Control*, 52(8):1475–1480.

Bombois, X., Gevers, M., Hildebrand, R., and Solari, G. (2011). Optimal experiment design for open and closed-loop system identification. *Communications in Information and Systems*, 11(3):197–224.

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

Dufour, J. M. and Jouini, T. (2014). Asymptotic distributions for quasi-efficient estimators in echelon VARMA models. *Computational Statistics & Data Analysis*, 73:69–86.

Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, pages 233–244.

Evans, A. G. and Fischl, R. (1973). Optimal least squares time-domain synthesis of recursive digital filters. *IEEE Transactions on Audio and Electroacoustics*, 21(1):61–65.

Everitt, N., Galrinho, M., and Hjalmarsson, H. (2017). Open-loop asymptotically efficient model reduction with the Steiglitz-McBride method. *under review at Automatica (arXiv:1610.08534)*.

Forssell, U. and Ljung, L. (1999). Closed-loop identification revisited. *Automatica*, 35:1215–1241.

Galrinho, M., Rojas, C. R., and Hjalmarsson, H. (2014). A weighted least-squares method for parameter estimation in structured models. In *53rd IEEE Conference on Decision and Control*, pages 3322–3327, Los Angeles, USA.

Galrinho, M., Rojas, C. R., and Hjalmarsson, H. (2018). Parametric identification using weighted null-space fitting. *provisionaally accepted at IEEE Transactions on Automatic Control* (arXiv 1708.03946).

Gilson, M., Garnier, H., Young, P., and van den Hof, P. (2009). Refined instrumental variable methods for closed-loop system identification. In *15th IFAC Symp. on System Identification*, Saint-Malo, France.

Gilson, M. and van den Hof, P. (2005). Instrumental variable methods for closed-loop system identification. *Automatica*, 41(2):241–249.

Hannan, E. and Kavalieris, L. (1984). Multivariate linear time series models. *Advances in Applied Probability*, 16(3):492–561.

Hannan, E. J. and Kavalieris, L. (1983). Linear estimation of ARMA processes. *Automatica*, 19(4):447–448.

Hjalmarsson, H. and Mårtensson, J. (2011). A geometric approach to variance analysis in system identification. *IEEE Trans. on Automatic Control*, 56(5):983–997.

Jansson, M. (2003). Subspace identification and ARX modeling. In *IFAC Symposium on System Identification*, pages 1585–1590, Rotterdam, The Netherlands.

Lemmerling, P., Vanhamme, L., van Huffel, S., and de Moor, B. (2001). IQML-like algorithms for solving structured total least squares problems: a unified view. *Signal Processing*, 81:1935–1945.

Ljung, L. (1999). *System Identification. Theory for the User.* Prentice-Hall.

Ljung, L. and Wahlberg, B. (1992). Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Advances in Applied Probabilities*, 24:412–440.

Mayne, D. and Firoozan, F. (1982). Linear identification of ARMA processes. *Automatica*, 18(4):461–466.

Partington, J. R. (1989). *An Introduction to Hankel Operators.* London Mathematical Society Student Texts. Cambridge University Press.

Pintelon, R., Schoukens, J., Vandersteen, G., and Barb, K. (2010a). Estimation of nonparametric noise and FRF models for multivariable systems—Part I: theory. *Mechanical Systems and Signal Processing*, 24(3):573–595.

Pintelon, R., Schoukens, J., Vandersteen, G., and Barb, K. (2010b). Estimation of nonparametric noise and FRF models for multivariable systems—Part II: extensions, applications. *Mech. Systems and Signal Processing*, 24(3):596–616.

Qin, S. and Ljung, L. (2003). Closed-loop subspace identification with innovation estimation. In *13th IFAC Symposium on System Identification*, pages 861–866, Rotterdam, The Netherlands.

Reinsel, G., Basu, S., and Yap, S. (1992). Maximum likelihood estimators in the multivariate autoregressive moving-average model from a generalized least squares viewpoint. *Journal of Time Series Analysis*, 13(2):133–145.

Schoukens, J., Rolain, Y., Vandersteen, G., and Pintelon, R. (2011). User friendly Box-Jenkins identification using nonparametric noise models. In *50th IEEE Conference on Decision and Control*, Orlando, Florida, USA.

Schoukens, J., Vandersteen, G., Barbé, K., and Pintelon, R. (2009). Nonparametric preprocessing in system identification: a powerful tool. In *Control Conference (ECC), 2009 European*, pages 1–14. IEEE.

Shaw, A. K. (1994). Optimal identification of discrete-time systems from impulse response data. *IEEE Trans. on Signal Processing*, 42(1):113–120.

Shaw, A. K., Misra, P., and Kumaresan, R. (1994). Identification of a class of multivariable systems from impulse response data: Theory and computational algorithm. *Circuits, Systems and Signal Processing*, 13(6):759–782.

Söderström, T. and Stoica, P. (1983). *Instrumental Variable Methods for System Identificaiton.* Springer.

Söderström, T. and Stoica, P. (1989). *System Identification.* Prentice Hall.

van Overschee, P. and de Moor, B. (1994). N4SID: Subspace algorithms for identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93.

Verhaegen, M. (1993). Application of a subspace model identification technique to identify LTI systems operating in closed-loop. *Automatica*, 29(4):1027–1040.

Zhu, Y. (2001). *Multivariable System Identification for Process Control.* Pergamon.

Zhu, Y. and Hjalmarsson, H. (2016). The Box-Jenkins Steiglitz-McBride algorithm. *Automatica*, 65:170–182.