A counterexample to the uniqueness of the asymptotic estimate in ARMAX model identification via the correlation approach

S. Garatti*

Abstract

This paper deals with the identifiability of an ARMAX system when the correlation approach is adopted. In general, identifiability depends on both the parametrization of the model class and on the informativeness of the data. Here, we focus on the latter aspect and, therefore, a full-order model class is considered. The main goal is to provide a counterexample to the uniqueness of the asymptotic estimate when a persistently exciting input is adopted. This shows the somehow counterintuitive fact that the identifiability of ARMAX systems within the correlation approach is related to the "color" of the input.

1 Introduction

Consider the problem of identifying a model for the ARMAX system

$$S^o: \quad A^o(z^{-1})y(t) = B^o(z^{-1})u(t) + C^o(z^{-1})w(t), \quad w(t) \sim \mathrm{WN}(0, \lambda_w^2),$$

*Dipartimento di Elettronica, Informazione e Biongegneria - Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy. E-mail: simone.garatti@polimi.it

based on a data record $\{u(1), y(1), \ldots, u(N), y(N)\}$ collected from S^o. Here, $\mathrm{WN}(0, \lambda_w^2)$ denotes a white noise with zero mean and variance λ_w^2. It is also assumed that $w(t)$ is i.i.d. and that it has bounded moments of order $4+\delta$ for some $\delta > 0$. Moreover, $w(t)$ is uncorrelated with $u(t)$.

Identification is performed within the class of ARMAX(n_a,n_b,n_c) models

$$\mathcal{M}_\vartheta = \left\{ y(t) = \frac{B(z^{-1},\vartheta)}{A(z^{-1},\vartheta)} u(t) + \frac{C(z^{-1},\vartheta)}{A(z^{-1},\vartheta)} \xi(t),\ \xi(t) \sim \mathrm{WN}(0,\lambda^2), \quad \vartheta \in \Theta \right\},$$

where $\vartheta = [a_1 \cdots a_{n_a}\ b_1\ \cdots\ b_{n_b}\ c_1\ \cdots\ c_{n_c}]'$,

$$\begin{aligned} A(z^{-1},\vartheta) &= 1 + a_1 z^{-1} + \ldots + a_{n_a} z^{-n_a}, \\ B(z^{-1},\vartheta) &= b_1 z^{-1} + \ldots + b_{n_b} z^{-n_b}, \\ C(z^{-1},\vartheta) &= 1 + c_1 z^{-1} + \ldots + c_{n_c} z^{-n_c}, \end{aligned}$$

and $\Theta \subseteq \mathbb{R}^n$, $n = n_a + n_b + n_c$, is the set of the parameter vectors such that the roots of $z^{n_a} A(z^{-1},\vartheta)$ and of $z^{n_c} C(z^{-1},\vartheta)$ belong to the interior of the unit circle in the complex domain. We assume that:

1. $S^o \in \mathcal{M}_\vartheta$, that is, $\exists\ \vartheta^o \in \Theta$ such that $A^o(z^{-1}) = A(z^{-1},\vartheta^o)$, $B^o(z^{-1}) = B(z^{-1},\vartheta^o)$, and $C^o(z^{-1}) = C(z^{-1},\vartheta^o)$ (full-order parametrization);
2. there is no common factor to all of $A(z^{-1},\vartheta^o)$, $B(z^{-1},\vartheta^o)$, and $C(z^{-1},\vartheta^o)$.

As for the fitting criterion, consider the so-called correlation approach, which aims at finding a model whose associated prediction error is as white as possible, [10, 13]. Precisely, letting

$$\widehat{y}(t,\vartheta) = \left(1 - \frac{A(z^{-1},\vartheta)}{C(z^{-1},\vartheta)}\right) y(t) + \frac{B(z^{-1},\vartheta)}{C(z^{-1},\vartheta)} u(t)$$

be the optimal 1-step linear predictor for the model corresponding to ϑ, the parameter estimate $\widehat{\vartheta}_N$ is computed as the solution of the following system of equations[1]:

$$\frac{1}{N} \sum_{t=1}^{N} \varepsilon(t,\vartheta) \zeta(t,\vartheta) = 0,$$

[1] If the solution is not unique, it is assumed that a tie-break rule is introduced.

where $\varepsilon(t,\vartheta) = y(t) - \widehat{y}(t,\vartheta)$ is the prediction error and $\zeta(t,\vartheta)$ is a n-long correlation vector constructed based on data up to time $t-1$. For instance, [10, 13], customary choices for ARMAX models are

$$\zeta(t,\vartheta) = [y(t-1) \ \cdots \ y(t-n_a) \ u(t-1) \ \cdots \ u(t-n_b) \ \varepsilon(t-1,\vartheta) \ \cdots \ \varepsilon(t-n_c,\vartheta)]', \tag{1}$$

$$\zeta(t,\vartheta) = [u(t-1) \ \cdots \ u(t-n_a-n_b) \ \varepsilon(t-1,\vartheta) \ \cdots \ \varepsilon(t-n_c,\vartheta)]', \tag{2}$$

$$\zeta(t,\vartheta) = [u(t-1) \ \cdots \ u(t-n_b) \ \varepsilon(t-1,\vartheta) \ \cdots \ \varepsilon(t-n_a-n_c,\vartheta)]'. \tag{3}$$

In the first case, $\widehat{\vartheta}_N$ takes the name of pseudo-linear regression estimate, [10].
Letting Θ^* be the set of solutions to the system of equations

$$\mathbb{E}[\varepsilon(t,\vartheta)\zeta(t,\vartheta)] = 0, \tag{4}$$

it can be proven under mild assumptions that the distance between $\widehat{\vartheta}_N$ and Θ^* tends to zero as $N \to \infty$, [10, 13]. Since ϑ^o belongs to Θ^* as it can be easily verified, if Θ^* is a singleton, then $\widehat{\vartheta}_N \to \vartheta^o$, i.e. the estimated model tends to the true data-generating system. If instead Θ^* contains multiple points, the identifiability of ϑ^o is no longer guaranteed, and this may cause severe problems in the assessment of the quality of the obtained model as shown in [7, 8, 6, 9].
In general, the conditions under which Θ^* is a singleton depend both on the model class parametrization (roughly speaking, Θ^* may not be a singleton because of over-parametrization – see [1, 13, 10, 3, 2] for results along this line) and on the informativeness of the data (if $u(t) = 0$, $\forall t$, then it is clear that $b_1^o \ \cdots \ b_{n_b}^o$ cannot be retrieved by any identification algorithm). In the present setup, however, the informativeness of data is the sole significant aspect, being the model class globally identifiable at ϑ^o according to [10, Definition 4.6].
For PEM (Prediction Error Methods) identification, the problem of the informativeness of the data has been studied in full detail, leading to the concepts of *informative enough data sequence* and *persistent excitation*, [10, Definitions 8.1 and 13.2], which give quite general conditions for Θ^* to be a singleton. Moreover, in [5], methods have been developed to design the input so as to reduce the presence of local minima and improve the convergence of $\widehat{\vartheta}_N$ to ϑ^o.

As for the correlation approach, instead, the problem of the informativeness of data is much more convoluted. If, on the one hand, the case of Instrumental Variable identification, where one disregards the identification of $C(z^{-1}, \vartheta^o)$, has been fully analyzed, leading to many general conditions guaranteeing the identifiability of $A(z^{-1}, \vartheta^o)$ and $B(z^{-1}, \vartheta^o)$, [12, 10], on the other hand, the case of ARMAX models here considered is largely unexplored and, to the best of our knowledge, partial achievements are available only. The following theorem, taken from [4], is one of the few available results.

Theorem 1 *Let*

$$\zeta(t, \vartheta) = [u(t-1) \ \cdots \ u(t-n_a-n_b) \ \varepsilon(t-1, \vartheta) \ \cdots \ \varepsilon(t-n_c, \vartheta)]'.$$

In the present setup, if $u(t) \sim \mathrm{WN}(0, \sigma^2)$ *with* $\sigma^2 > 0$, *then* $\Theta^* = \{\vartheta^o\}$.

When the input is not white, one might expect that the same conclusion of Theorem 1 can be drawn by assuming that the input is persistently exciting[2], in full analogy with what happens in PEM identification. This leads to the following conjecture.

Conjecture 1 *Let*

$$\zeta(t, \vartheta) = [u(t-1) \ \cdots \ u(t-n_a-n_b) \ \varepsilon(t-1, \vartheta) \ \cdots \ \varepsilon(t-n_c, \vartheta)]'.$$

In the present setup, if $u(t)$ *is persistently exciting, then* $\Theta^* = \{\vartheta^o\}$.

In the next Section 2, we prove that Conjecture 1 is false by giving a counterexample where Θ^* is not a singleton even though $u(t)$ is persistently exciting. Moreover, the counterexample shows that Θ^* may not be a singleton also when either the correlation vector (1) or (3) is used. This shows that the identifiability of an ARMAX system within the correlation approach may depend on the "color" of the input, even for standard choices of the correlation vector.

Paper structure. The counterexample is constructed and discussed in Section 2. Some conclusions are eventually drawn in Section 3.

[2] Note that, thanks to the uncorrelation with $w(t)$, the persistent excitation property implies that the collected data sequence is also informative enough, see [10, Theorem 13.1].

2 A counterexample to Conjecture 1

Consider the ARMAX(0,1,3) models with $A(z^{-1}, \vartheta) = 1$, $B(z^{-1}, \vartheta) = bz^{-1}$, $C(z^{-1}, \vartheta) = 1 + c_1 z^{-1} + c_2 z^{-2} + c_3 z^{-3}$, and $\vartheta = [b\ c_1\ c_2\ c_3]'$. The data-generating system is

$$S^o: \quad y(t) = b^o u(t-1) + w(t) + c_1^o w(t-1) + c_2^o w(t-2) + c_3^o w(t-3).$$

where $w(t) \sim \mathrm{WN}(0,1)$. We let $u(t) = e(t) + e(t-1) = (1 + z^{-1})e(t)$, where $e(t) \sim \mathrm{WN}(0,1)$ and $e(t)$ is independent of $w(t)$. Note that $u(t)$ is a persistently exciting input. Finally, we consider

$$\zeta(t, \vartheta) = [u(t-1)\ \varepsilon(t-1, \vartheta)\ \cdots\ \varepsilon(t-n_c, \vartheta)]' \tag{5}$$

as correlation vector. Being $n_a = 0$, (5) simultaneously encompasses the three cases in (1)–(3).

In the current setting, the system of equations (4) writes as

$$\mathbb{E}[\varepsilon(t, \vartheta) u(t-1)] = 0, \tag{6}$$

$$\mathbb{E}[\varepsilon(t, \vartheta) \varepsilon(t-i, \vartheta)] = 0 \quad i = 1, \ldots, 3. \tag{7}$$

The aim of this section is to show that equations (6) and (7) admit for some ϑ^o a solution ϑ^* which is not equal to ϑ^o. The counterexample is built indirectly: the parameter vector ϑ^* is fixed, and a ϑ^o not equal to ϑ^* is sought such that ϑ^* is a solution to (6) and (7) when data are generated by the S^o corresponding to ϑ^o. It follows that, for the found S^o, equations (6) and (7) admit at least two different solutions, namely, $\vartheta = \vartheta^o$ and $\vartheta = \vartheta^*$.

Preliminarily, note that, given a ϑ, the polynomial $C(z^{-1}, \vartheta)$ can be represented in the form

$$C(z^{-1}, \vartheta) = (1 + \alpha z^{-1})(1 + \beta z^{-1})(1 + \gamma z^{-1}), \tag{8}$$

where α, β, and γ may be complex. In the following, we will denote by ρ the vector $[b\ \alpha\ \beta\ \gamma]'$. Note that, given ρ, there is a unique ϑ such that equation (8) holds true, while, given ϑ, vector ρ is uniquely determined up to a permutation of α, β and γ.

Fix a real-valued $\rho^* = [b^*\ \alpha^*\ \beta^*\ \gamma^*]'$ such that $\alpha^* + \beta^* + \gamma^* = 2$, $|\alpha^*| < 1$, $|\beta^*| < 1$, $|\gamma^*| < 1$ and $\alpha^* \neq \beta^* \neq \gamma^*$, and let ϑ^* be the corresponding parameter vector. We aim at finding a ϑ^o not equal to ϑ^* such that (6) and (7) admit this ϑ^* as a solution.

First, consider equation (6). In the current example, the prediction error in correspondence of ϑ^* can be written as

$$\varepsilon(t, \vartheta^*) = \frac{(b^o - b^*)}{C(z^{-1}, \vartheta^*)} u(t-1) + \frac{C(z^{-1}, \vartheta^o)}{C(z^{-1}, \vartheta^*)} w(t),$$

and, owning to the independence of $u(t)$ and $w(t)$, equation (6) writes

$$\mathbb{E}\Big[\frac{(b^o - b^*)}{C(z^{-1}, \vartheta^*)} u(t-1) \cdot u(t-1)\Big] = 0. \tag{9}$$

The transfer function $1/C(z^{-1}, \vartheta^*)$ admits the following series expansion

$$\begin{aligned}
\frac{1}{C(z^{-1}, \vartheta^*)} &= \frac{1}{(1+\alpha^* z^{-1})(1+\beta^* z^{-1})(1+\gamma^* z^{-1})} \\
&= 1 - (\alpha^* + \beta^* + \gamma^*) z^{-1} + \sum_{k=2}^{\infty} h_k z^{-k} \\
&= 1 - 2z^{-1} + \sum_{k=2}^{\infty} h_k z^{-k},
\end{aligned}$$

where h_k's are suitable coefficients depending on α^*, β^*, γ^*. By substituting the latter expression in the left-hand-side of equation (9), we obtain

$$\begin{aligned}
&\mathbb{E}\Big[(b^o - b^*)\Big(u(t-1) - 2u(t-2) + \sum_{k=3}^{\infty} h_{k-1} u(t-k)\Big) \cdot u(t-1)\Big] \\
&= (b^o - b^*)\Big(\mathbb{E}[u(t-1)^2] - 2\mathbb{E}[u(t-2)u(t-1)] + \sum_{k=3}^{\infty} h_{k-1} \mathbb{E}[u(t-k)u(t-1)]\Big),
\end{aligned}$$

which is indeed equal to 0 $\forall\ b^o$, because $\mathbb{E}[u(t-1)^2] = 2$, $\mathbb{E}[u(t-2)u(t-1)] = 1$ and $\sum_{k=3}^{\infty} h_{k-1} \mathbb{E}[u(t-k)u(t-1)] = 0$ (recall that $u(t) = e(t) + e(t-1)$). Hence, the conclusion is that equation (6) always admits ϑ^* as a solution whatever ϑ^o is.

Turn now to the equations in (7). Since $u(t)$ is independent of $w(t)$, they can be rewritten as

$$\begin{aligned}
&\mathbb{E}\Big[\frac{(b^o - b^*)}{C(z^{-1}, \vartheta^*)} u(t-1) \cdot \frac{(b^o - b^*)}{C(z^{-1}, \vartheta^*)} u(t-1-i)\Big] + \\
&\mathbb{E}\Big[\frac{C(z^{-1}, \vartheta^o)}{C(z^{-1}, \vartheta^*)} w(t) \cdot \frac{C(z^{-1}, \vartheta^o)}{C(z^{-1}, \vartheta^*)} w(t-i)\Big] = 0, \quad i = 1, \ldots, 3.
\end{aligned} \tag{10}$$

Taking into account the representation (8), the equations in (10) can be seen as a system of equations in the unknown ρ^o, namely,

$$(b^o - b^*)^2 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} f_1(\alpha^o,\beta^o,\gamma^o) \\ f_2(\alpha^o,\beta^o,\gamma^o) \\ f_3(\alpha^o,\beta^o,\gamma^o) \end{bmatrix}, \tag{11}$$

where we let

$$\begin{aligned} v_i &= \mathbb{E}\Big[\frac{1}{C(z^{-1},\vartheta^*)}u(t-1)\cdot\frac{1}{C(z^{-1},\vartheta^*)}u(t-1-i)\Big] \quad \text{and} \\ f_i(\alpha^o,\beta^o,\gamma^o) &= -\mathbb{E}\Big[\frac{C(z^{-1},\vartheta^o)}{C(z^{-1},\vartheta^*)}w(t)\cdot\frac{C(z^{-1},\vartheta^o)}{C(z^{-1},\vartheta^*)}w(t-i)\Big]. \end{aligned}$$

Clearly, (11) is satisfied by $\rho^o = \rho^*$, which amounts to taking $\vartheta^o = \vartheta^*$.

Suppose first that $[v_1\ v_2\ v_3]' = 0$.

In this case, when $(\alpha^o,\beta^o,\gamma^o) = (\alpha^*,\beta^*,\gamma^*)$, (11) is satisfied no matter of the value of b^o. This means that (11) are satisfied e.g. by taking $\rho^o = [b^o\ \alpha^*\ \beta^*\ \gamma^*]'$ with $b^o = b^*+1$. Clearly, $\rho^o \neq \rho^*$ and it corresponds to a $\vartheta^o \neq \vartheta^*$, because $b^o = b^*+1 \neq b^*$.

Suppose now that $[v_1\ v_2\ v_3]' \neq 0$.

The right hand side of (11) is a mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$ that is continuously differentiable (the coefficients of $C(z^{-1},\vartheta^o)$ "smoothly" depend on α^o, β^o, and γ^o, [10]). It holds that

$$\begin{aligned} \frac{\partial f_1}{\partial \alpha^o}(\alpha^*,\beta^*,\gamma^*) = \mathbb{E}[\frac{z^{-1}}{1+\alpha^* z^{-1}}w(t)\cdot w(t-1)] &= 1 \\ \frac{\partial f_2}{\partial \alpha^o}(\alpha^*,\beta^*,\gamma^*) = \mathbb{E}[\frac{z^{-1}}{1+\alpha^* z^{-1}}w(t)\cdot w(t-2)] &= -\alpha^* \\ \frac{\partial f_3}{\partial \alpha^o}(\alpha^*,\beta^*,\gamma^*) = \mathbb{E}[\frac{z^{-1}}{1+\alpha^* z^{-1}}w(t)\cdot w(t-3)] &= (\alpha^*)^2, \end{aligned}$$

and, similarly,

$$\begin{aligned} \tfrac{\partial f_1}{\partial \beta^o}(\alpha^*,\beta^*,\gamma^*) &= 1 & \tfrac{\partial f_1}{\partial \gamma^o}(\alpha^*,\beta^*,\gamma^*) &= 1 \\ \tfrac{\partial f_2}{\partial \beta^o}(\alpha^*,\beta^*,\gamma^*) &= -\beta^* & \tfrac{\partial f_2}{\partial \gamma^o}(\alpha^*,\beta^*,\gamma^*) &= -\gamma^* \\ \tfrac{\partial f_3}{\partial \beta^o}(\alpha^*,\beta^*,\gamma^*) &= (\beta^*)^2 & \tfrac{\partial f_1}{\partial \gamma^o}(\alpha^*,\beta^*,\gamma^*) &= (\gamma^*)^2. \end{aligned}$$

Hence, the Jacobian of $[f_1(\alpha^o,\beta^o,\gamma^o)\ f_2(\alpha^o,\beta^o,\gamma^o)\ f_3(\alpha^o,\beta^o,\gamma^o)]'$ evaluated in $(\alpha^*,\beta^*,\gamma^*)$ is given as follows:

$$\left[\begin{array}{ccc} \frac{\partial f_1}{\partial \alpha^o} & \frac{\partial f_1}{\partial \beta^o} & \frac{\partial f_1}{\partial \gamma^o} \\ \frac{\partial f_2}{\partial \alpha^o} & \frac{\partial f_2}{\partial \beta^o} & \frac{\partial f_2}{\partial \gamma^o} \\ \frac{\partial f_3}{\partial \alpha^o} & \frac{\partial f_3}{\partial \beta^o} & \frac{\partial f_3}{\partial \gamma^o} \end{array}\right]_{(\alpha^*,\beta^*,\gamma^*)} = \left[\begin{array}{ccc} 1 & 1 & 1 \\ -\alpha^* & -\beta^* & -\gamma^* \\ (\alpha^*)^2 & (\beta^*)^2 & (\gamma^*)^2 \end{array}\right],$$

and is nonsingular because it is a Vandermonde matrix with $\alpha^* \neq \beta^* \neq \gamma^*$. Eventually, we have that

$$\left[\begin{array}{c} f_1(\alpha^*,\beta^*,\gamma^*) \\ f_2(\alpha^*,\beta^*,\gamma^*) \\ f_3(\alpha^*,\beta^*,\gamma^*) \end{array}\right] = \left[\begin{array}{c} \mathbb{E}[w(t)w(t-1)] \\ \mathbb{E}[w(t)w(t-2)] \\ \mathbb{E}[w(t)w(t-3)] \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right].$$

The inverse function theorem – see [11, Theorem 9.24] – can be now invoked to assert that there is an open neighborhood U of $[\alpha^*\ \beta^*\ \gamma^*]'$ and an open neighborhood V of $[0\ 0\ 0]'$ such that the mapping $[f_1(\alpha^o,\beta^o,\gamma^o)\ f_2(\alpha^o,\beta^o,\gamma^o)\ f_3(\alpha^o,\beta^o,\gamma^o)]'$ is one-to-one between U and V. This implies that $\forall\ [v_1\ v_2\ v_3]' \in \mathbb{R}^3$, $\forall\ \epsilon \geq 0$, ϵ small enough, $\exists\ [\alpha^o\ \beta^o\ \gamma^o]'$ such that

$$\left[\begin{array}{c} f_1(\alpha^o,\beta^o,\gamma^o) \\ f_2(\alpha^o,\beta^o,\gamma^o) \\ f_3(\alpha^o,\beta^o,\gamma^o) \end{array}\right] = \epsilon \left[\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array}\right].$$

Hence, by letting $b^o = b^* \pm \sqrt{\epsilon}$, $\epsilon > 0$ small enough, a $\rho^o = [b^o\ \alpha^o\ \beta^o\ \gamma^o]'$ can be found such that the system of equations (11) is satisfied and $\rho^o \neq \rho^*$. This ρ^o corresponds to a ϑ^o that is different from ϑ^* because $b^o = b^* \pm \sqrt{\epsilon} \neq b^*$.

Summarizing, for the given ϑ^*, a $\vartheta^o \neq \vartheta^*$ can be found such that equations (6) and (7) admit ϑ^* as a solution. Since ϑ^o is a solution to (6) and (7) too, this proves that Θ^* is not a singleton when data are generated in correspondence of the found ϑ^o.

Remark 1 *It is perhaps worth noticing that the found ϑ^o is not an isolated singularity. As a matter of fact, the given counterexample shows that for the input $u(t) = e(t) + e(t-1)$, $e(t) \sim \mathrm{WN}(0,1)$, there is a whole set of ϑ^o's for which Θ^* is not a singleton. This set is obtained by considering all possible values of ϵ in the above construction and by letting ρ^* vary in all possible ways under the conditions $\alpha^* + \beta^* + \gamma^* = 2$, $|\alpha^*| < 1$,*

$|\beta^*| < 1$, $|\gamma^*| < 1$, *and* $\alpha^* \neq \beta^* \neq \gamma^*$. *Also, the given argument can be easily extended to the case where* $u(t) = e(t) + ke(t-1)$, $e(t) \sim \mathrm{WN}(0, \lambda_e^2)$.

3 Conclusions

In this paper, we considered the identification of an ARMAX system by means of a full-order model class and showed that identifiability may be not attained despite the use of a persistently exciting input. The counterexample holds true in spite of the choice of the correlation vector among some standard options, including the most common pseudo-linear regression. Though the specificity of the counterexample does not allow one to draw general conclusions about the applicability of the correlation approach, still, we believe, this paper reveals some difficulties of this identification method that perhaps deserve further investigation. To the best of our knowledge, whether a correlation vector exists guaranteeing identifiability under a persistent excitation condition remains an open problem. Similarly, the conditions securing identifiability for the correlation vectors considered in this paper are not clear. The hope is that this paper may foster further research along these directions.

References

[1] K.J. Åstrom and T. Söderström. Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model. *IEEE Transactions on Automatic Control*, 19:769–773, 1974.

[2] A.S. Bazanella, X. Bombois, and M. Gevers. Necessary and sufficient conditions for uniqueness of the minimum in Prediction Error Identification. *Automatica*, 48:1621–1630, 2012.

[3] C.I. Byrnes, P. Enqvist, and A. Lindquist. Identifiability and well-posedness of shaping-filter parameterizations: a global analysis approach. *SIAM Journal on Control and Optimization*, 41:23–59, 2003.

[4] M.C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41:1751–1764, 2005.

[5] D. Eckhard, A.S. Bazanella, C.R. Rojas, and H. Hjalmarsson. Input design as a tool to improve the convergence of PEM. *Automatica*, 49:3282–3291, 2013.

[6] S. Garatti and R.R. Bitmead. On resampling and uncertainty estimation in linear system identification. *Automatica*, 46:785–795, 2010.

[7] S. Garatti, M.C. Campi, and S. Bittanti. Assessing the quality of identified models trough the asymptotic theory - When is the result reliable? *Automatica*, 40:1319–1332, 2004.

[8] S. Garatti, M.C. Campi, and S. Bittanti. The asymptotic model quality assessment for Instrumental Variable identification revisited. *Systems & Control Letters*, 55:494–500, 2006.

[9] S. Garatti, M.C. Campi, and S. Bittanti. Iterative robust control: speeding up improvement through iterations. *Systems & Control Letters*, 59:139–146, 2010.

[10] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, NJ, 1999.

[11] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, NY, 1976.

[12] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Lecture Notes in Control and Information Sciences. Springer-Verlag, New York, NY, 1983.

[13] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall, Englewood Cliffs, NJ, 1989.