2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Bias considerations when identifying systems from noisy input-output data — Extensions to general model structures

Torsten Söderström[1] and Umberto Soverini[2]

***Abstract*— Standard identification methods give biased parameter estimates when recorded signals are corrupted by noise on both input and output sides. In previous papers it has been shown that the bias is significant in case the system is almost non-identifiable. This situation is investigated here for some general model structures.**

***Index Terms*— System identification, Errors-in-variables, Bias**

I. INTRODUCTION

In an errors-in-variables situation as described by (2)-(4) below, special action has to be applied due to the presence of input noise $\tilde{u}(t)$, see [2]. All standard identification methods, see for example [1], [7] yield biased (rather, non-consistent) estimates when the measured input signal contains additional noise. The bias can be considerable in case the system is almost not identifiable.

The focus in this paper is to examine the obtained bias when the presence of the input noise is neglected. A preliminary study of the size of the bias was given in [5], where it was assumed that the model structure is an output error model with white output noise, and the prediction error method (PEM) is used for identification. Subsequently, in [4], [6] the analysis of the bias was studied in the case an instrumental variable method is used for the identification.

The parameter bias $\mathbf{b}$ can, as for many other estimation problems, be written as a sum of two terms,

$$\mathbf{b} = \mathbf{b}_{\mathrm{s}} + \mathbf{b}_{\mathrm{r}} \tag{1}$$

where $\mathbf{b}_{\mathrm{s}}$ is a systematic error and $\mathbf{b}_{\mathrm{r}}$ is a random component. The term $\mathbf{b}_{\mathrm{s}}$ will persist also when the number of data points, N, grows without bound. In fact, it is natural to take $\mathbf{b}_{\mathrm{s}}$ as $\lim_{N\to\infty} \mathbf{b}$. The random error appear as soon as N is finite, and describes the dependence on the specific realization of the data. The study in [5] as well as here concerns the systematic error $\mathbf{b}_{\mathrm{s}}$. The random part of the bias is for large N of magnitude $O(1/\sqrt{N})$.

Should the bias term be analyzed by numerical simulations, with finite N, one then also have to deal with the effects of the random bias term $\mathbf{b}_{\mathrm{r}}$. Moreover, the obtained result will be specific for the chosen underlying system and it would not be possible to draw any general conclusions that are valid also for other systems.

[1] Torsten Söderström is with Division of Systems and Control, Department of Information Technology, Uppsala University, P O Box 337, SE-751 05 Uppsala, Sweden torsten.soderstrom@it.uu.se

[2] Umberto Soverini is with Dept of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy, umberto.soverini@unibo.it

In [5] we studied the bias for dynamic models with white output noise. Two specific features were highlighted. It was shown that the influence of the noise on the input signal is of order $O(\lambda_u^2)$, where λ_u^2 is the input noise variance. Another studied aspect is the influence of almost pole-zero cancellations. If the smallest pole-zero separation is δ, then the bias was shown to be $O(1/\delta)$ for small δ.

This paper considers again the study of [5], and extends the results in different ways. First, the case of an output error model with colored output noise is considered. Second, we consider the case of an ARMAX model structure in this regard, and make comparisons to the output error model. Further material about this second aspect can be found in [3].

The paper is organized as follows. The next section contains the general background, and Section III reviews the analysis for the output error model structure from [5]. The output error model structure with colored output noise is treated in Section IV, while Section V contains a discussion for the use of ARMAX model structures. In Section VI the theoretical results are illustrated by some numerical examples. Finally, some conclusions are provided in Section VII.

II. PROBLEM FORMULATION

This section summarizes the general problem setup, considered here as well as in [5], [6].

Assume that the system under consideration is linear and single input-single output. Measurements of both input and output are assumed to be noisy:

$$\begin{aligned} y(t) &= G_0(q)u_0(t) + H_0(q)e(t) \ , && (2)\\ u(t) &= u_0(t) + \tilde{u}(t) \ , && (3)\\ u_0(t) &= F(q)v(t) \ . && (4) \end{aligned}$$

Here $u_0(t)$ denotes the noise-free input signal, while $u(t)$ is the noise-corrupted input and $y(t)$ is the noise-corrupted output. Note that as (2)-(4) refers to an errors-in-variables situation, cf [2], the user cannot design nor influence $u_0(t)$. Further, the transfer functions $G_0(q)$, $H_0(q)$ and $F(q)$ are all assumed to be rational functions of the shift operator q. To simplify expressions in the following the argument q will often be dropped.

The input noise $\tilde{u}(t)$ is assumed to be white with variance λ_u^2. Further, $e(t)$ is assumed to be white noise with variance λ_e^2, and $v(t)$ is assumed to be white noise with variance λ_v^2. The output noise is therefore an ARMA process and it is white only in the special case $H_0(q) = 1$. Note that the output noise $H_0(q)e(t)$ consists of both process noise affecting the system as well as measurement noise. The equation (4) means that the noise-free input $u_0(t)$ is an ARMA process. As F (and its order) is arbitrary, (4) is a fairly general description of a stationary process. The variances λ_u^2, λ_e^2 and λ_v^2 are all assumed to be unknown. It is also assumed that the signals $e(t)$, $v(t)$ and $\tilde{u}(t)$ are independent. This means in particular that open loop operation is assumed. In case feedback would be present, $u_0(t)$ would include (through the feedback) also a term that depends on the output, and thus on $e(t)$. See [2], [8] for details.

Next the model description will be specified. Assume that a model of the form

$$y(t) = G(q)u(t) + H(q)\varepsilon(t) \tag{5}$$

is to be fitted to the recorded input-output data. Here $G(q) = G(q, \theta)$ and $H(q) = H(q, \theta)$ are parameterized with a vector θ. The dependence on θ is mostly not spelled out in what follows.

Assume that the parameterization is such that there is a unique value θ_* that makes

$$G(q, \theta_*) \equiv G_0(q), \quad H(q, \theta_*) \equiv H_0(q) \ . \tag{6}$$

This is a form of identifiability assumption.

Let the estimate (in the asymptotic case when the number of data points $N \to \infty$) be denoted by $\hat{\theta}$. The bias of the estimate is then

$$\tilde{\theta} = \hat{\theta} - \theta_* \ . \tag{7}$$

III. Review of bias due to almost non-identifiability

The results in this section are taken from [5].

Assume that identification is made using the prediction error method (PEM) applied to the data. In the case of no input noise present it is well-known that PEM gives consistent and statistically efficient parameter estimates, [1], [7].

Use of the PEM means that the parameter estimate can be written as

$$\begin{aligned} \hat{\theta} &= \arg\min_{\theta} V(\theta) \ , & (8) \\ V(\theta) &= \frac{1}{2}\mathsf{E}\left\{\varepsilon^2(t,\theta)\right\} \ . & (9) \end{aligned}$$

The expectation in (9), and in what follows, is with respect to all noise sources: $e(t), \tilde{u}(t), v(t)$. In (9) the prediction error $\varepsilon(t, \theta)$ can be found directly from (5), leading to

$$\varepsilon(t) = H(q)^{-1}\left[y(t) - G(q)u(t)\right] \ . \tag{10}$$

An approximate way to express the bias $\tilde{\theta}$ is as follows. Let $\hat{\theta}$ denote the minimum point of $V(\theta)$, and assume that the bias $\tilde{\theta}$ is small. Then using a linearization

$$0 = V'_\theta(\hat{\theta}) \approx V'_\theta(\theta_*) + V''_{\theta\theta}(\theta_*)(\hat{\theta} - \theta_*) \ , \tag{11}$$

leads to

$$\tilde{\theta} \approx -\left[V''_{\theta\theta}(\theta_*)\right]^{-1} V'_\theta(\theta_*) \ . \tag{12}$$

It was shown in [5] that (12) is indeed often a good approximation of the bias $\tilde{\theta}$. Further, $\tilde{\theta}$ will be large when the inverse $[V''_{\theta\theta}(\theta_*)]^{-1}$ is large, which occurs when the Hessian $V''_{\theta\theta}(\theta_*)$ is almost singular. This happens when the system is (almost) not identifiable. Such a situation can happen in two different ways:

- (Almost) overparameterization. This will show up in that some polynomials of the model have (almost) a common factor.
- The noise-free input u_0 is (almost) not persistently exciting of enough order.

For most model structures, the loss function (9) is not convex, [7]. It is assumed here that the global minimum is obtained, so the linearization in (11) is around the true value θ_*.

In the following we will assume that the noise-free input is persistently exciting and we will examine only the first aspect. Next we recall from [5] some more explicit results for the case of an output error model structure with white output noise. This case is characterized by the following equations

$$\begin{aligned} y(t) &= y_0(t) + \tilde{y}(t), \quad E\{\tilde{y}^2(t)\} = \lambda_y^2 \ , & (13) \\ u(t) &= u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2 \ , & (14) \\ Ay_0(t) &= Bu_0(t) \ , & (15) \\ A &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \ , & (16) \\ B &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \ . & (17) \end{aligned}$$

The equation (15) refers to the model to be fitted. The true data ('the system') is assumed to also fulfill (15), but the polynomials are then denoted A_0, B_0. Compared to (2) it here holds that $H_0 = 1$, i.e. the output noise is assumed to be white.

At first, it is necessary to recall the following result. Consider two generic polynomials

$$\begin{aligned} A &= a_0 z^{n_a} + a_1 z^{n_a - 1} + \ldots + a_{n_a} \ , & (18) \\ B &= b_0 z^{n_b} + b_1 z^{n_b - 1} + \ldots + b_{n_b} \ . & (19) \end{aligned}$$

Then the associated Sylvester matrix is the square matrix of dimension $(n_a + n_b) \times (n_a + n_b)$ given by

$$\mathcal{S}(A, B) = \begin{pmatrix} b_0 & b_1 & \ldots & b_{n_b} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & b_0 & b_1 & \ldots & b_{n_b} \\ a_0 & a_1 & \ldots & a_{n_a} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & a_0 & a_1 & \ldots & a_{n_a} \end{pmatrix} \ . \tag{20}$$

The properties of Sylvester matrices have been investigated in many sources. Some basic properties are, for example, reviewed in [7].

Starting from (12), in [5] the following result was proved, that gives an approximation of the expected bias:

$$\beta_1 = -\left(V''_{\theta\theta}(\theta_*)\right)^{-1} V'_\theta(\theta_*) = \mathcal{S}^{-T}(-A_0, B_0) P_{\varphi_u}^{-1} r_0 \ . \tag{21}$$

where

$$
\begin{aligned}
P_{\varphi_u} &= \mathsf{E}\left\{\varphi_u(t)\varphi_u^T(t)\right\} && (22)\\
\varphi_u(t) &= \frac{1}{A^2}\begin{pmatrix} u(t-1)\\ \vdots \\ u(t-n_a-n_b)\end{pmatrix} && (23)\\
r_0 &= \mathsf{E}\{\frac{B_0}{A_0}\tilde{u}(t)\varphi_{\tilde{u}}(t)\}\ . && (24)
\end{aligned}
$$

We remark that the 'input filter' F effects the bias through the term P_{φ_u}.

A somewhat cruder approximation of the bias, labeled β_2 in this paper, is obtained by using only the noise-free part of the input in (22). Specifically, then substitute P_{φ_u} in (21) by $P_{\varphi_{u_0}}$, where $\varphi_{u_0}(t)$ denotes the noise-free part of $\varphi_u(t)$.

When the system has almost a pole-zero cancellation, the matrix inverse $\mathcal{S}^{-T}(-A_0, B_0)$ will have large elements. To be specific, let the system have poles $p_i, i = 1, \ldots, n_a$ and zeros $z_j, j = 1, \ldots, n_b$. Then set

$$\delta = \min_{i,j} |p_i - z_j|\ , \tag{25}$$

which is a measure of the pole-zero separation. It was shown in [5] that for small values of δ the determinant of the Sylvester matrix is proportional to δ. The inverse of the Sylvester matrix will therefore generally have elements of the order $O(1/\delta)$.

A related study appears in [6]. There the focus was to investigate some strategies in order to reduce the large bias in the system parameter estimates, in presence of small pole-zero separation. In particular, two possible solutions were proposed. The first one makes use of a reduced model structure, the second employs a full errors-in-variables model.

IV. A GENERAL LINEAR MODEL STRUCTURE

Consider in this section the general linear model structure (5), where G and H are assumed to be rational functions with independent parameters. One may thus regard this case as an output error structure with colored noise.

Specifically, assume

$$
\begin{aligned}
G(q) &= \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}}\ , && (26)\\
H(q) &= \frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d}}\ . && (27)
\end{aligned}
$$

Then the prediction error becomes

$$
\begin{aligned}
\varepsilon(t,\theta) &= \frac{D}{C}\left[y(t) - \frac{B}{A}u(t)\right] && (28)\\
&= \frac{D}{C}\left[\frac{B_0}{A_0}u_0(t) + \frac{C_0}{D_0}e(t) - \frac{B}{A}\left(u_0(t) + \tilde{u}(t)\right)\right]\ . && (29)
\end{aligned}
$$

Its gradient fulfills

$$
\begin{aligned}
\frac{\partial\varepsilon}{\partial a_i}(t) &= \frac{DBq^{-i}}{CA^2}u(t)\ , && (30)\\
\frac{\partial\varepsilon}{\partial b_i}(t) &= -\frac{Dq^{-i}}{CA}u(t)\ , && (31)\\
\frac{\partial\varepsilon}{\partial c_i}(t) &= -\frac{Dq^{-i}}{C^2}\left(y(t) - \frac{B}{A}u(t)\right)\ , && (32)\\
\frac{\partial\varepsilon}{\partial d_i}(t) &= \frac{q^{-i}}{C}\left(y(t) - \frac{B}{A}u(t)\right)\ . && (33)
\end{aligned}
$$

This leads to

$$\varepsilon'_\theta(t) = \begin{pmatrix} \mathcal{S}(-A,B)\varphi_1(t) \\ \\ \mathcal{S}(C,-D)\varphi_2(t)\end{pmatrix}\ , \tag{34}$$

where

$$
\begin{aligned}
\varphi_1(t) &= \frac{D}{A^2C}\begin{pmatrix} q^{-1}\\ \vdots \\ q^{-n_a-n_b}\end{pmatrix}u(t)\ , && (35)\\
\varphi_2(t) &= \frac{1}{C^2}\begin{pmatrix} q^{-1}\\ \vdots \\ q^{-n_c-n_d}\end{pmatrix}\left(y(t) - \frac{B}{A}u(t)\right)\ . && (36)
\end{aligned}
$$

When evaluating $V'_\theta(\theta_*)$ and $V''_{\theta\theta}(\theta_*)$ one then gets

$$V'_\theta(\theta_*) = \begin{pmatrix} \mathcal{S}(-A_0,B_0)r_1 \\ \mathcal{S}(C_0,-D_0)r_2\end{pmatrix}\ , \tag{37}$$

$$r_1 = -\mathsf{E}\left\{\frac{B_0}{A_0}\tilde{u}(t)\frac{D_0}{A_0^2C_0}\begin{pmatrix}\tilde{u}(t-1)\\ \vdots \\ \tilde{u}(t-n_a-n_b)\end{pmatrix}\right\}\ , \tag{38}$$

$$r_2 = \mathsf{E}\left\{\frac{C_0}{D_0}e(t)\frac{1}{C_0D_0}\begin{pmatrix}e(t-1)\\ \vdots \\ e(t-n_c-n_d)\end{pmatrix}\right\}\ , \tag{39}$$

$$
\begin{aligned}
V''_{\theta\theta}(\theta_*) &= \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22}\end{pmatrix}\ , && (40)\\
V_{11} &= \mathcal{S}(-A_0,B_0)P_{\varphi_1}\mathcal{S}^T(-A_0,B_0)\ , && (41)\\
V_{22} &= \mathcal{S}(C_0,-D_0)P_{\varphi_2}\mathcal{S}^T(C_0,-D_0)\ . && (42)
\end{aligned}
$$

Similar to (21) the approximated bias β_1 can now be written as

$$\beta_1 = \begin{pmatrix} \mathcal{S}^{-T}(-A_0,B_0)P_{\varphi_1}^{-1}r_1 \\ \mathcal{S}^{-T}(C_0,-D_0)P_{\varphi_2}^{-1}r_2\end{pmatrix} \tag{43}$$

The somewhat cruder approximation β_2 of the bias is obtained by using only the noise-free part of the input when forming P_{φ_1} and P_{φ_2}.

Some observations

- As $V''_{\theta\theta}(\theta_*)$ is block diagonal, the estimates of A and B are (asymptotically) uncorrelated with the estimates of C and D.

- If A_0 and B_0 have almost a pole-zero cancellation, then the estimates of A and B are quite uncertain, just as in the output error case treated in [5].
- Similarly, if C_0 and D_0 show almost a pole-zero cancellation, then the estimates of C and D are quite uncertain.

V. The ARMAX model structure

Consider now an ARMAX model. This means that the parameterization is such that G and H in (5) have the same denominator.

This case is characterized by the following equations

$$\begin{aligned} y(t) &= y_0(t) + C/Ae(t), \quad E\{e^2(t)\} = \lambda_e^2 \ , && (44)\\ u(t) &= u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2 \ , && (45)\\ y_0(t) &= B/Au_0(t) \ , && (46)\\ A &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \ , && (47)\\ B &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \ , && (48)\\ C &= 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c} \ . && (49) \end{aligned}$$

Assuming that the unperturbed input signal $u_0(t)$ is persistently exciting, the model is non-identifiable precisely when **all** *the three polynomials* A, B, C *have a common factor*. This corresponds to the case of over-parameterization.

Example. Assume that the true data corresponds to the transfer functions

$$G = \frac{B_1}{A_1}, \quad H = \frac{C_1}{D_1} \ . \qquad (50)$$

This leads quickly to the following ARMAX model polynomials

$$A = A_1 D_1, \quad B = B_1 D_1, \quad C = C_1 A_1 \ . \qquad (51)$$

In case the two transfer functions G and H have some joint poles, this means that the polynomials A_1 and D_1 are not coprime. In fact, the characteristic polynomial formed from these joint poles will be a joint factor of A, B, C. ■

The asymptotic error criterion can still be written as in (9), but now the error (in fact, the one-step ahead prediction error) should instead of (10) be written as

$$\begin{aligned} \varepsilon(t,\theta) &= \frac{A}{C} y(t) - \frac{B}{C} u(t) \\ &= \frac{AB_0 - A_0 B}{A_0 C} u_0(t) + \frac{A}{C}\frac{C_0}{A_0} e(t) - \frac{B}{C}\tilde{u}(t) \ . \end{aligned} \qquad (52)$$

Needless to say, (52) leads to various changes in the expression for the gradient of $\varepsilon(t)$ with respect to the parameters.

In this case the parameter vector θ is given by

$$\theta = \begin{pmatrix} a_1 & \ldots & a_{n_a} & b_1 & \ldots & b_{n_b} & c_1 & \ldots & c_{n_c} \end{pmatrix}^T \ . \qquad (53)$$

The derivatives of the polynomials with respect to θ can be written as

$$\begin{aligned} A'_\theta &= \begin{pmatrix} q^{-1} & \ldots & q^{-n_a} & O_{1\times(n_b+n_c)} \end{pmatrix} \ , && (54)\\ B'_\theta &= \begin{pmatrix} O_{1\times n_a} & q^{-1} & \ldots & q^{-n_b} & O_{1\times n_c} \end{pmatrix} \ , && (55)\\ C'_\theta &= \begin{pmatrix} O_{1\times(n_a+n_b)} & q^{-1} & \ldots & q^{-n_c} \end{pmatrix} \ . && (56) \end{aligned}$$

When evaluating the gradient $\varepsilon'_\theta(t,\theta_*)$ one gets in an intermediate step

$$\begin{aligned} GH_\theta - G_\theta H &= \frac{B_0}{A_0}\left(-\frac{C_0 A_\theta}{A_0^2} + \frac{C_\theta}{A_0}\right) \\ &\quad -\frac{C_0}{A_0}\left(-\frac{B_0 A_\theta}{A_0^2} + \frac{B_\theta}{A_0}\right) \\ &= -\frac{C_0}{A_0^2} B_\theta + \frac{B_0}{A_0^2} C_\theta \ . \end{aligned} \qquad (57)$$

Using this result in (10) leads to

$$\begin{aligned} \varepsilon(t,\theta_*) &= e(t) - \frac{A_0}{C_0}\frac{B_0}{A_0}\tilde{u}(t) = e(t) - \frac{B_0}{C_0}\tilde{u}(t), && (58)\\ \varepsilon'_\theta(t,\theta_*) &= -\frac{-B_0 A_\theta + A_0 B_\theta}{A_0 C_0} u_0(t) \\ &\quad + \frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t) \\ &\quad - \frac{-C_0 A_\theta + A_0 C_\theta}{A_0 C_0} e(t) \ . && (59) \end{aligned}$$

The gradient of V in (12) becomes

$$V'_\theta(\theta_*) = -\mathsf{E}\left\{\left[\frac{B_0}{C_0}\tilde{u}(t)\right]\left[\frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t)\right]\right\} \ , \qquad (60)$$

while the Hessian will be

$$\begin{aligned} V''_{\theta\theta}(\theta_*) &= \operatorname{cov}\left[\frac{-B_0 A_\theta + A_0 B_\theta}{A_0 C_0} u_0(t)\right] \\ &\quad + \operatorname{cov}\left[\frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t)\right] \\ &\quad + \operatorname{cov}\left[\frac{-C_0 A_\theta + A_0 C_\theta}{A_0 C_0} e(t)\right] \ . \end{aligned} \qquad (61)$$

A further simplification is possible using Sylvester matrices. For example, consider the expression

$$\phi_1(t) = \frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t) \ . \qquad (62)$$

It follows from (55) and (56) that $\phi_1(t)$ is an $1\times(n_a+n_b+n_c)$ vector, with the first n_a elements being zero. The vector composed of the elements $n_a+1,\ldots,n_a+n_b+n_c$ can be written as

$$\begin{aligned} &\begin{pmatrix} -C_0 q^{-1} & \ldots & B_0 q^{-1} & \ldots & B_0 q^{-n_c} \end{pmatrix} \frac{1}{C_0^2}\tilde{u}(t) \\ &= \frac{1}{C_0^2}\begin{pmatrix} q^{-1}\tilde{u}(t) & \ldots & q^{-n_b-n_c}\tilde{u}(t) \end{pmatrix} \mathcal{S}^T(-C_0, B_0) \ . \end{aligned} \qquad (63)$$

Recall that the approximate bias term β_1 was defined as the right hand side of (12). The Hessian $V''_{\theta\theta}(\theta_*)$ consists of various covariance elements of the filtered input $u(t)$. The cruder approximation β_2 is obtained by substituting $u(t)$ by the noise-free part $u_0(t)$ in all these covariance elements.

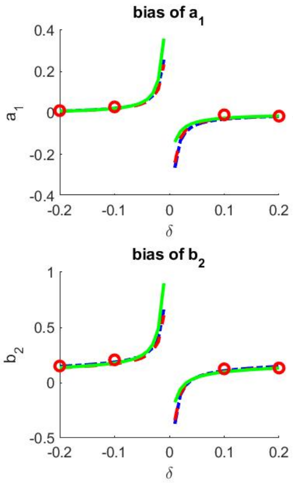

Fig. 1. Parameter biases versus δ, for the output error model. The true biases (β_t) are shown with solid, green lines. The approximate biases (β_1), see (21), are shown with dashed, red lines. The cruder approximate biases (β_2) are shown with dash-dotted, blue lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the input noise variance was $\lambda_u^2 = 0.1$.

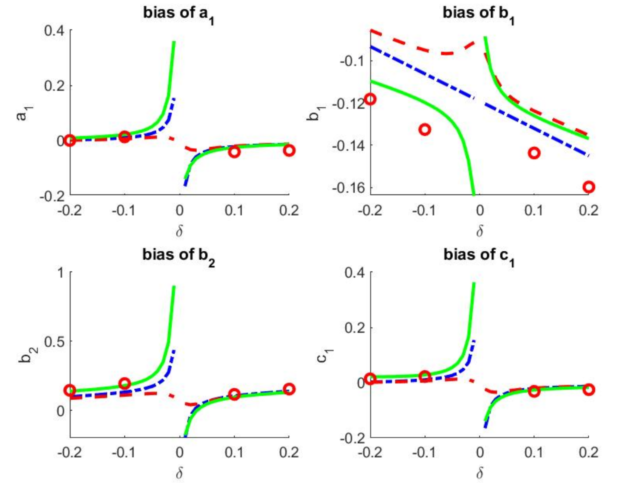

Fig. 2. Parameter biases versus δ, for the ARMAX model. The true biases (β_t) are shown with solid, green lines. The approximate biases (β_1) are shown with dashed, red lines. The cruder approximate biases (β_2) are shown with dash-dotted, blue lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the input noise variance was $\lambda_u^2 = 0.1$.

VI. Numerical examples

As a background, and for comparison purposes, we first repeat an example from [5].

A. Use of an output error model

To illustrate the above results in more detail consider a simple numerical example with $n_a = 1, n_b = 2$ and where $u_0(t)$ is an AR(1) process,

$$u_0(t) = Fv(t), \quad F = \left(1 - 0.9q^{-1}\right)^{-1}, \quad \mathsf{E}\{v^2(t)\} = 1 \ . \tag{64}$$

The other parameters in the numerical example are

$$a_1 = -0.8, \quad \lambda_y^2 = 10, \quad b_1 = 2 \ . \tag{65}$$

In the numerical study the input noise variance λ_u^2 was varied. So was also the coefficient $b_2 = 2(-0.8 - \delta)$. Note that the value $\delta = 0$ corresponds to A_0 and B_0 having a common zero, and where identifiability is lost.

In the numerical study the approximate bias expressions β_1 and β_2 were computed. They are compared to the 'true' bias β_t, which was computed by minimizing the loss function (9). The results were also compared numerically to some Monte Carlo simulations, where the output error identification method was applied to a number of realizations. 100 input-output realizations, each of length 1000 were used.

In Figure 1 the parameter biases versus the parameter δ are displayed.

B. Use of an ARMAX model

Consider the same example and data as examined for the output error case in Section VI-A. Note that the output error model corresponds to the ARMAX model *with the constraint* $C = A$. Expressed differently, for the example studied here, it holds for *the true data* $A = 1 - 0.8q^{-1} = C$, while

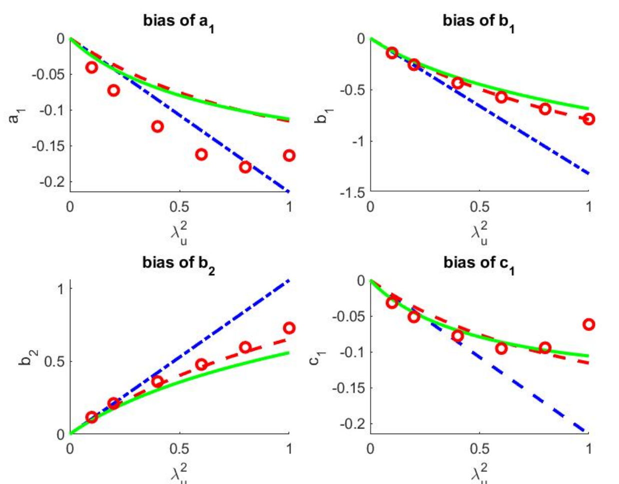

Fig. 3. Parameter biases versus λ_u^2, for the ARMAX model. The true biases (β_t) are shown with solid, green lines. The approximate biases (β_1) are shown with dashed, red lines. The cruder approximate biases (β_2) are shown with dash-dotted, blue lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the parameter δ is $\delta = 0.1$.

$B = 2(1 - (0.8 + \delta)q^{-1})$. We estimate though the dynamics with an ARMAX model (thus not exploiting that $A = C$ in the estimation).

The numerical investigations are displayed in Figures 2 and 3.

Some observations

- The OE and ARMAX models have mostly similar qualitative properties.
- The ARMAX models seem to be more robust than the OE models, in the sense that no problems with false

local minima were observed.

- The Monte Carlo simulations give mostly similar results to those predicted by theory.

C. Further considerations for an ARMAX model

The parameter bias when using a PEM for an ARMAX model was further considered in the technical report [3]. When a prediction error method is used with either an output error model or a model with independent parameters for the transfer functions G and H, then a small pole-zero separation δ leads to a parameter bias of order $O(1/\delta)$. The examples presented in [3] shows that this behavior *does not extend* to the ARMAX case. These examples are summarized here.

Consider a first order system with $n_a = 1,\ n_b = 2,\ n_c = 1$. The bias for four different cases will be presented.

Case 1. Let B and C have a joint zero. The parameters are chosen as

$$a = -0.8 + \delta, \quad b_1 = 1, \quad b_2 = -0.8, \ c = -0.8\ . \tag{66}$$

Let the noise-free input be white noise of zero mean and variance σ^2. The variances are chosen as

$$\sigma^2 = 1, \quad \lambda_u^2 = 1, \quad \lambda_e^2 = 1\ . \tag{67}$$

The obtained results are:

- The biases of a and c are zero.
- The bias of b_1 is 0.5 and that of b_2 is -0.4, independent of the value of δ.

Case 2. Consider the same system and parameters as in Case 1, with the modification that the noise-free input is no longer white noise, but a first order regression

$$u_0(t) + f u_0(t-1) = v(t), \quad \mathsf{E}\left\{v^2(t)\right\} = \sigma^2 = 1, \quad f = -0.9\ . \tag{68}$$

The obtained results differ drastically from Case 1:

- The bias of a, b_2 and c behave as $O(1/\delta)$.
- The bias of b_1 varies very slowly with δ.

Case 3. Next modify the original example slightly in another way. Force A and B to have a joint zero. The noise-free input is still assumed to be white noise. The parameters for this example are thus

$$a = -0.8, \quad b_1 = 1, \quad b_2 = -0.8, \quad c = -0.8 + \delta\ . \tag{69}$$

Let the noise-free input be white of zero mean and variance $\sigma^2 = 1$.

The obtained results are:

- The biases of a and c behave as $O(\delta)$.
- The biases of b_1 and b_2 vary very slowly with δ.

Case 4. Modify the original example slightly in still another way. Force A and C to have a joint zero. The noise-free input is still assumed to be white noise. The parameters for this example are thus

$$a = -0.8, \quad b_1 = 1, \quad b_2 = -0.8 + \delta, \quad c = -0.8\ . \tag{70}$$

Let the noise-free input be white of zero mean and variance $\sigma^2 = 1$.

The obtained results are:

- The bias terms $|a|$ and $|c|$ are both $O(\delta)$.
- The bias term $|b_1|$ does not vary with δ. It is equal to 0.5.
- The bias term $|b_2|$ varies very slowly with δ. For small values of δ it is equal to -0.4.

A general conclusion from the four cases in this subsection is that

> *For an ARMAX model the behavior of the bias as a function of the pole-zero separation δ is quite different in the 4 cases, despite the fact that the examples themselves differ only by slightly modifying the system parameters.*

VII. Conclusions

When standard identification methods are applied to noise-corrupted input-output data, biased parameter estimates occur due to the presence of input noise. It has been assumed that a standard prediction error method is applied. When the system is close to be not identifiable due to an almost pole-zero cancellation, the bias will be large. It was shown that the bias is $O(1/\delta)$ where δ is the pole-zero separation. This result applies for the pole-zero separation of the system transfer function as well as for the noise-shaping filter. Further, a comparison between using the output error model structure or an ARMAX model was performed. On the contrary to the general linear model case in Section IV, *there is no general result* in the ARMAX case that the bias is always $O(1/\delta)$ for a small pole-zero separation δ.

References

[1] L. Ljung. *System Identification - Theory for the User, 2nd edition*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[2] T Söderström. *Errors-in-Variables Methods in System Identification*. Springer-Verlag, London, UK, 2018.

[3] T. Söderström and U. Soverini. Analyzing the parameter bias when an ARMAX model is fitted to noise-corrupted data. Technical Report 2022-009, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2022. Available as http://www.it.uu.se/research/publications/ reports/2022-009.

[4] T. Söderström and U. Soverini. Analyzing the parameter bias when an instrumental variable method is used with noise-corrupted data. Technical Report 2022-008, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2022. Available as http://www.it.uu.se/research/publications/ reports/2022-008.

[5] T. Söderström and U. Soverini. When are errors-in-variables aspects important to consider in system identification? In *2022 European Control Conference (ECC)*, pages 315–320, London, UK, July 12-15 2022.

[6] T. Söderström and U. Soverini. Aspects of errors-in-variables identification: Some ways to mitigate a large bias. In *22nd IFAC World Congress*, Yokohama, Japan, July 9-14 2023.

[7] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, Hemel Hempstead, UK, 1989.

[8] T. Söderström, L. Wang, R. Pintelon, and J. Schoukens. Can errors-in-variables systems be identified from closed-loop experiments? *Automatica*, 49(2):681–684, February 2013.