When Are Errors-in-Variables Aspects Important to Consider in System Identification?

When Are Errors-in-Variables Aspects Important to Consider in System Identification?

Torsten Söderström[1] and Umberto Soverini[2]

***Abstract*— When recorded signals are corrupted by noise on both input and output sides, standard identification methods give biased parameter estimates, due to the presence of input noise. This paper discusses in what situations such a bias is large and, consequently, when errors-in-variables identification methods should preferably be used.**

***Index Terms*— system identification, errors-in-variables, bias, parameter estimation, output error model**

I. Introduction

All standard identification methods, see for example [10], [19] yield biased (rather, non-consistent) estimates when the measured input signal contains additional noise. In case also the input data are affected by noise, the estimation of the parameters for linear dynamic systems is recognized as a more difficult problem. Representations where errors or measurement noises are present on both inputs and outputs are usually called 'errors-in-variables' (EIV) models. Such models play an important role when the purpose is determination of the physical laws that describe the process, rather than the prediction of its future behavior.

A thorough description of various methods that give consistent identification of dynamic systems in an EIV setting is given in the book [17]. See also the survey paper [15]. Aspects from a user perspective on how to cope with EIV problems are discussed in [18]. For some further overviews, see [5] and [16].

There are many papers and publications where EIV methods are used for various types of applications. As examples, one can mention identifying transmissibility functions in a mechanical mass-spring system, [21], data-driven controller design algorithms, [20], [1], electromagnetic mineral exploration, [6], roll dynamics of a ship, [7], structural health monitoring used to check the status of (large) mechanical constructions and buildings, [4], [14]. Additional applications are listed in e.g. [17].

The focus in this paper is somewhat different. Assume that there is in fact some measurement noise on the recorded input signal, but that this is neglected when applying a standard system identification technique. The problem under discussion is to find out in what situations the obtained bias will be significant, or even large. For example, one may want to know how the size of the bias is influenced by the system dynamics, and by the character of the true input. The paper contains a preliminary study of this issue, with particular attention to the case of an output error model structure. More general model structures will be investigated in forthcoming papers. It will be shown that a critical situation where the standard identification methods yield highly biased estimates arises when the system is almost not identifiable due to the presence of a small pole-zero separation. These considerations will be supported by simple numerical examples.

The paper is organized as follows. The next section describes the modelling and general problem formulation. The main ideas behind this paper are presented in Section III. More specific and explicit analysis is then given in Section IV when the models structure is an output error model. Section V contains some concluding discussion.

II. Modelling and problem formulations

This section starts off by giving assumptions on the recorded data. This is formulated as a description of the unknown system to be identified. Next a general model description is postulated for identification purposes, and the problem formulation is given.

Assume that the system (the mathematical description of the unknown dynamics to be identified) is linear and single input-single output. Measurements of both input and output are assumed to be noise-corrupted. In mathematical form, these assumptions are expressed as

$$y(t) = G_0(q)u_0(t) + H_0(q)e(t)\ , \tag{1}$$

$$u(t) = u_0(t) + \tilde{u}(t)\ , \tag{2}$$

$$u_0(t) = F(q)v(t)\ . \tag{3}$$

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. 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 specific input model (3) is not used in the general analysis, but in the numerical example.

The input noise $\tilde{u}(t)$ is assumed to be white with unknown 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 (3) means that the noise-free input $u_0(t)$ is an ARMA process.

[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

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.

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

$$y(t) = G(q)u(t) + H(q)\varepsilon(t) \tag{4}$$

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.

In the study assume that the parameterization is such that there is a unique value θ_* that makes

$$G(q,\theta_*) \equiv G_0(q), \quad H(q,\theta_*) \equiv H_0(q) \ . \tag{5}$$

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

$$\tilde{\theta} = \hat{\theta} - \theta_* \ . \tag{6}$$

For consistency it is required that the bias is zero. The focus here is on the asymptotic case, and one may think of the bias in this sense as a systematic error that does not disappear when the number of data points tends to infinity.

The problem to be discussed is the following:

Problem. What factors influence the size of the bias $\tilde{\theta}$? This problem involves several aspects. For example, how is the size of $\tilde{\theta}$ is influenced by the filters G_0, H_0, F?

III. Main ideas

From now on 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, [9], [19]. The PEM method has been chosen since it gives highly accurate estimates in general. However, other identification method could be used, for example the instrumental variable method.

The PEM means that the parameter estimate can be written as (assuming for simplicity that the number of data is infinite)

$$\hat{\theta} = \arg\min_{\theta} V(\theta) \ , \tag{7}$$

$$V(\theta) = \frac{1}{2} E\left\{\varepsilon^2(t,\theta)\right\} \ . \tag{8}$$

where E denotes expectation. In (8) the prediction error $\varepsilon(t,\theta)$ can be found directly from (4), leading to

$$\varepsilon(t) = H(q)^{-1}\left[y(t) - G(q)u(t)\right] \ . \tag{9}$$

To the best of our knowledge, there is no general analytical and explicit way to express the bias $\tilde{\theta}$. An approximate way is as follows. The validity of this approximation will be tested in the next section. Let $\hat{\theta}$ denote the minimum point of $V(\theta)$, and assume that the bias $\tilde{\theta}$ is small. Then try a linearization

$$0 = V_\theta'(\hat{\theta}) \approx V_\theta'(\theta_*) + V_{\theta\theta}''(\theta_*)(\hat{\theta} - \theta_*) \ , \tag{10}$$

leading to

$$\tilde{\theta} \approx -\left[V_{\theta\theta}''(\theta_*)\right]^{-1} V_\theta'(\theta_*) \ . \tag{11}$$

Next develop explicit expressions for the terms in the right hand side of (11). To this aim, use the notations

$$\varepsilon_\theta = \frac{\partial \varepsilon(t,\theta)}{\partial \theta} \ , \tag{12}$$

$$G_\theta = \frac{\partial G(q,\theta)}{\partial \theta} \ , \tag{13}$$

$$H_\theta = \frac{\partial H(q,\theta)}{\partial \theta} \ . \tag{14}$$

and let, similarly, $\varepsilon_{\theta\theta}$ denote the derivative of ε_θ. Direct differentiation gives

$$V_\theta' = E\left\{\varepsilon(t,\theta)\varepsilon_\theta(t,\theta)\right\} \ , \tag{15}$$

$$V_{\theta\theta}'' = E\left\{\varepsilon(t,\theta)\varepsilon_{\theta\theta}(t,\theta)\right\} + E\left\{\varepsilon_\theta(t,\theta)\varepsilon_\theta^T(t,\theta)\right\} \tag{16}$$

and also

$$\varepsilon_\theta(t,\theta) = -\frac{H_\theta}{H^2}\left[y(t) - G(q)u(t)\right] - \frac{G_\theta}{H}u(t) \ . \tag{17}$$

The expressions so far hold for a general value of the parameter vector θ. Next specialize to the value $\theta = \theta_*$ and use the data description in (1)-(3). This leads to

$$\begin{aligned}\varepsilon(t,\theta_*) &= H_0^{-1}\left[G_0u_0(t) + H_0e(t) - G_0\left(u_0(t) + \tilde{u}(t)\right)\right] \\ &= e(t) - H_0^{-1}G_0\tilde{u}(t) \ ,\end{aligned} \tag{18}$$

$$\begin{aligned}\varepsilon_\theta(t,\theta_*) &= -\frac{H_\theta}{H_0^2}\left[G_0u_0(t) + H_0e(t) - G_0u_0(t) - G_0\tilde{u}(t)\right] \\ &\quad -\frac{G_\theta}{H_0}\left[u_0(t) + \tilde{u}(t)\right] \\ &= -\frac{G_\theta}{H_0}u_0(t) + \left(\frac{G_0H_\theta}{H_0^2} - \frac{G_\theta}{H_0}\right)\tilde{u}(t) - \frac{H_\theta}{H_0}e(t).\end{aligned} \tag{19}$$

In (19), G_θ and H_θ are to be evaluated for $\theta = \theta_*$.

Assume that the system operates in open loop, and thus that $v(t)$, $e(t)$ and $\tilde{u}(t)$ are independent. Then the gradient $V_\theta'(\theta_*)$ can be evaluated as follows:

$$V_\theta'(\theta_*) = -E\left\{\left(H_0^{-1}G_0\tilde{u}(t)\right)\left(\frac{G_0H_\theta}{H_0^2} - \frac{G_\theta}{H_0}\right)\tilde{u}(t)\right\} \ . \tag{20}$$

Note that $E\{e(t)\frac{H_\theta}{H_0}e(t)\} = 0$, as the innovation $e(t)$ is white noise, and H_θ contains a delay.

Evaluating the expectation in (20) gives that $V_\theta'(\theta^*)$ is proportional to the noise variance λ_u^2. Thus, from (11) it follows that $\tilde{\theta} \to 0$ as $\lambda_u^2 \to 0$ (the bias will be small, if the input noise level is small). On the other hand, the bias $\tilde{\theta}$ will be large if $V_{\theta\theta}''(\theta^*)$ is almost singular. This in turn 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 (one may think of an ARMAX model or an output error model) have (almost) a common factor. See Section IV for explicit details.
- The noisefree input u_0 is (almost) not persistently exciting of enough order. One can think of extreme

cases such as $u_0(t) \equiv 1$, or even $u_0(t) \equiv 0$. This reason for (almost) loss of identifiability is not examined in further details here. Some relevant papers for studying the problem of identification under poor excitation include [8], [13], [2], [3], [12], [11]. Typically, one can expect all parameter estimates to be (very) uncertain. A low order of excitation means that the input signal is (almost) the sum of a few sinusoids. One can expect that the system transfer function is reasonably well estimated for precisely the frequencies of these sinusoids, and will have large uncertainties otherwise.

To sum up so far, what can be expected is very natural: When the system is almost not identifiable several things happen:

- The Hessian $V''_{\theta\theta}(\theta_*)$ will be almost singular.
- The bias $\tilde{\theta}$ will be large.
- The covariance matrix of the parameter estimates $\hat{\theta}$ will be large.

IV. Detailed analysis

A. Preliminaries

This section gives more explicit forms of the analysis of Section III. Specialization to the output error model structure is given in Section IV-B. First some preliminaries for the so called Sylvester matrices are given in this subsection.

Consider two polynomials

$$\begin{aligned} A &= a_0 z^{n_a} + a_1 z^{n_a-1} + \ldots + a_{n_a} \ , & (21)\\ B &= b_0 z^{n_b} + b_1 z^{n_b-1} + \ldots + b_{n_b} \ . & (22) \end{aligned}$$

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

$$\mathcal{S}(A,B) = \begin{pmatrix} b_0 & b_1 & \ldots & b_{n_b} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & b_0 & b_1 & \ldots & b_{n_b} \\ a_0 & a_1 & \ldots & a_{n_a} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & a_0 & a_1 & \ldots & a_{n_a} \end{pmatrix} . \quad (23)$$

The properties of Sylvester matrices have been investigated in many sources. Some basic properties are, for example, considered in [19]. Among the properties are the following:

- If A and B are coprime (that is, have no common zero), then the Sylvester matrix $\mathcal{S}(A,B)$ is nonsingular.
- If A and B have precisely $k > 0$ common zeros, then the Sylvester matrix $\mathcal{S}(A,B)$ singular. Further, its null space has dimension k. The null space of $\mathcal{S}^T(A,B)$ can be characterized, see [19].
- Assume that A has zeros in $\alpha_j, j = 1, \ldots, n_a$ and that B has zeros in $\beta_k, k = 1, \ldots, n_b$. Then one can show that it holds

$$\det\left(\mathcal{S}(A,B)\right) = (-1)^{n_a \times n_b} a_0^{n_b} b_0^{n_a} \prod_{j=1}^{n_a} \prod_{k=1}^{n_b} (\alpha_j - \beta_k) \ . \quad (24)$$

Observe that $(-1)^{n_a \times n_b} = 1$ unless both n_a and n_b are odd numbers. A proof of (24) is given in the appendix.

B. The output error model structure

The case of output error (OE) model structure is charaterized by the following equations

$$\begin{aligned} y(t) &= y_0(t) + \tilde{y}(t), \quad E\{\tilde{y}^2(t)\} = \lambda_y^2 \ , & (25)\\ u(t) &= u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2 \ , & (26)\\ Ay_0(t) &= Bu_0(t) \ , & (27)\\ A &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \ , & (28)\\ B &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \ . & (29) \end{aligned}$$

The equation (27) refers to the model to be fitted. The true data ('the system') is assumed to also fulfil (27), but the polynomials are then denoted $A_0,\ B_0$. Compared to (1) it here holds that $H_0 = 1$, i.e. the output noise is assumed to be white.

The asymptotic error criterion for OE is

$$V = \frac{1}{2} E\{\varepsilon^2(t)\} \ , \quad (30)$$

where the output error $\varepsilon(t)$ is

$$\varepsilon(t) = y(t) - \frac{B}{A}u(t) = \left(\frac{B_0}{A_0} - \frac{B}{A}\right) u_0(t) + \tilde{y}(t) - \frac{B}{A}\tilde{u}(t) \ . \quad (31)$$

Its gradient is easily found to be

$$\begin{aligned} \varepsilon_\theta(t) &= \begin{pmatrix} \frac{B}{A^2} q^{-1} u(t) & \ldots & \frac{B}{A^2} q^{-n_a} u(t) \\ -\frac{1}{A} q^{-1} u(t) & \ldots & -\frac{1}{A} q^{-n_b} u(t) \end{pmatrix}^T \\ &= \mathcal{S}(-A,B) \frac{1}{A^2} \begin{pmatrix} u(t-1) \\ \vdots \\ u(t-n_a-n_b) \end{pmatrix} \\ &\triangleq \mathcal{S}(-A,B)\varphi_u(t) \ . & (32) \end{aligned}$$

One can now write the Hessian $V''_{\theta\theta}$ and the gradient V'_θ as

$$\begin{aligned} V''_{\theta\theta}(\theta_*) &= \mathcal{S}(-A_0,B_0) P_{\varphi_u} \mathcal{S}^T(-A_0,B_0) \\ &= \mathcal{S}(-A_0,B_0) P_{\varphi_{u_0}} \mathcal{S}^T(-A_0,B_0) \\ &\quad + \mathcal{S}(-A_0,B_0) P_{\varphi_{\tilde{u}}} \mathcal{S}^T(-A_0,B_0) \ , & (33)\\ V'_\theta(\theta_*) &= E\{\varepsilon(t)\varepsilon'(t,\theta_*)\} \\ &= -\mathcal{S}(-A_0,B_0) E\{\frac{B_0}{A_0}\tilde{u}(t)\varphi_{\tilde{u}}(t)\} \\ &\triangleq -\mathcal{S}(-A_0,B_0) r_0 \ , & (34) \end{aligned}$$

where P_{φ_u} denotes the covariance matrix of $\varphi_u(t)$.

An approximation of the expected bias then becomes, according to (11),

$$\beta_1 = -\left(V''_{\theta\theta}(\theta_*)\right)^{-1} V'_\theta(\theta_*) = \mathcal{S}^{-T}(-A_0,B_0) P_{\varphi_u}^{-1} r_0 \ . \quad (35)$$

A somewhat cruder approximation is obtained by neglecting the influence of the input noise in the matrix $P_{\varphi_u}^{-1}$, which would lead to

$$\beta_2 = \mathcal{S}^{-T}(-A_0,B_0) P_{\varphi_{u_0}}^{-1} r_0 \ . \quad (36)$$

One can now see that

- β_2 increases linearly with λ_u^2.
- $\beta_1 \approx \beta_2$ for small values of λ_u^2.

The true value of the (asymptotic) bias can be found by numerically minimizing the loss function (30).

Both the more exact expression β_1 and the cruder approximation β_2 include the matrix inverse $\mathcal{S}^{-T}(-A_0, B_0)$. When the system has almost a pole-zero cancellation, this matrix inverse 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

$$\delta = \min_{i,j} |p_i - z_j| \tag{37}$$

which is a measure of the pole-zero separation. It now follows from (24) that for small values 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 $O(1/\delta)$. Also the bias expressions (35) and (36) will be of the order $O(1/\delta)$.

Numerical example

To explore 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,

$$u_0(t) = Fv(t), \quad F = \left(1 - 0.9q^{-1}\right)^{-1}, \quad E\{v^2(t)\} = 1 \ . \tag{38}$$

The other parameters in the numerical example are

$$a_1 = -0.8, \quad \lambda_y^2 = 10, \quad b_1 = 2 \ . \tag{39}$$

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 identifiability is then lost.

One can object that results obtained with a first order simulated model have no general validity. However, higher order numerical examples would lead to a more complicated analysis, without introducing any novelty with respect to the considerations reported after (37), which are true for a general model order.

In the numerical study the approximate bias expressions β_1, see (35), and β_2, see (36) were computed. They are compared to the 'true' bias β_t, computed by minimizing the loss function (30). The results were also compared numerically to some Monte Carlo simulations, where the output error identification method was applied to a number of realizations.

In the numerical examples, 100 realizations, each of length 1000 input-output pairs were used.

In Figure 1 the parameter biases versus the parameter δ are displayed. In Figure 2 the parameter biases versus the input noise variance λ_u^2 are displayed.

The approximate bias expressions β_1 and β_2 are reasonably good, except for the bias of b_1 for small values of δ, see Figure 1.

As a further examination of how the bias errors are influenced by the pole-zero separation δ, and by the input

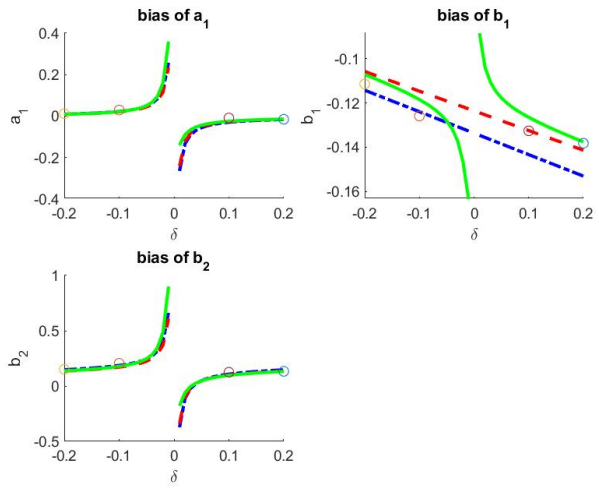

Fig. 1. Parameter biases versus δ. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the input noise variance was $\lambda_u^2 = 0.1$.

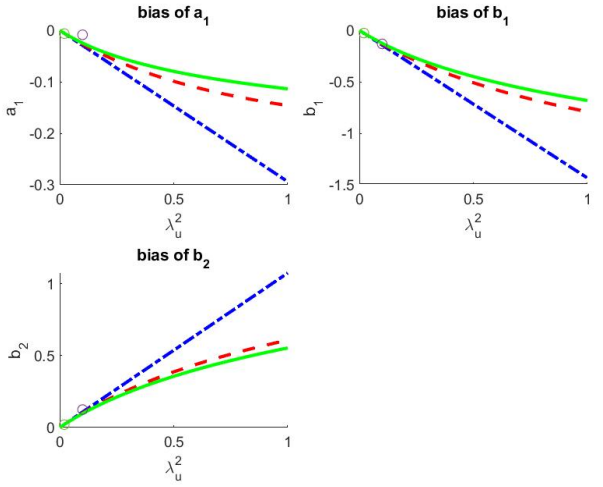

Fig. 2. Parameter biases versus λ_u^2. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the parameter δ is $\delta = 0.1$.

noise variance λ_u^2 let both these two quantities vary, and examine the contour levels of the biases.

In addition explore a scalar measure of the total parameter bias. To that aim consider the relative error in the transfer function G, taken as

$$\delta G = \frac{\int |\frac{B(e^{i\omega})}{A(e^{i\omega})} - \frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 d\omega}{\int |\frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 d\omega} \ . \tag{40}$$

The study of how the biases of a_1, b_1 and b_2 as well as δG depend on δ and λ_u^2 is displayed in Figures 3 and 4, for positive and negative values of δ, respectively.

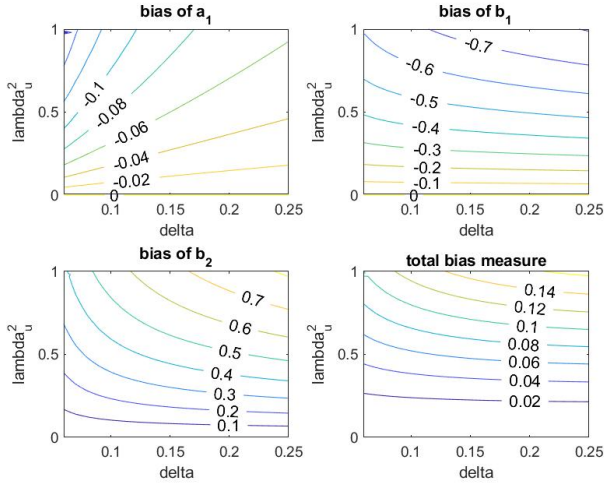

Fig. 3. Contour levels for parameter biases and the total bias measure δG as functions of δ and λ_u^2, $\delta > 0$.

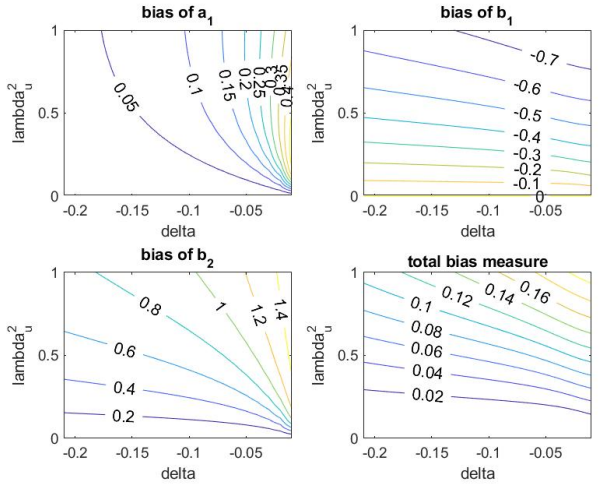

Fig. 4. Contour levels for parameter biases and the total bias measure δG as functions of δ and λ_u^2, $\delta < 0$.

The contour plots looked quite distorted for small positive values of δ. This is the reason why the δ axis starts at $\delta = 0.05$ in Figure 3. It seems reasonable to expect that this is connected to false local minima.

A further examination of false local minima was undertaken in the following way. The asymptotic loss function (with the constant λ_y^2 subtracted) can be written as

$$\begin{aligned} V(a,b_1,b_2) &= E\{[y(t) - \frac{B}{A}u(t)]^2\} - \lambda_y^2 \\ &= E\{[\left(\frac{B_0}{A_0} - \frac{B}{A}\right)u_0(t) - \frac{B}{A}\tilde{u}(t)]^2\} \ . \end{aligned} \tag{41}$$

As the loss function V is quadratic in b_1, b_2, this minimization can be done analytically and one can consider

$$W(a) = \min_{b_1,b_2} V(a,b_1,b_2) \ . \tag{42}$$

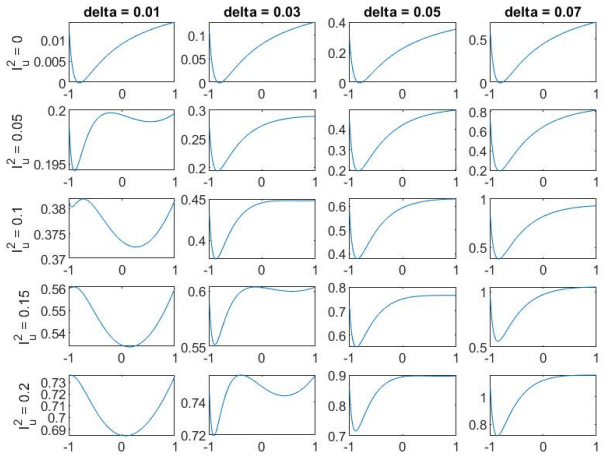

Fig. 5. The graph of $W(a)$ for various values of the parameters δ and λ_u^2.

Apparently, a local minimum of $V(a,b_1,b_2)$ has a corresponding local minimum of $W(a)$. The behaviour of the concentrated loss function $W(a)$ for some various values of δ and λ_u^2 are given in Figure 5.

It is clear from Figure 5 that there are 'false' local minima of $W(a)$, and hence also of $V(a,b_1,b_2)$ for small positive values of δ. Note that the true value of the parameter a is -0.8.

1) Some observations:

- For the given example the experimental and numerical results coincide well with what could be expected from theory.
- The approximate values for the biases coincide well with the 'theoretical values', especially for small values of λ_u^2.
- Also for moderate values of λ_u^2, the approximate bias values (β_1 and β_2) give fairly good values of the biases.
- It is shown in Figure 1 that the biases grows without bound when $\delta \to 0$. This is fairly natural. The limiting case $\delta = 0$, corresponds to a non-identifiable system, as then there is a pole-zero cancellation in the transfer function B/A.
- In the Monte Carlo simulations, when λ_u^2 was chosen higher (in Figure 2), the output error estimates often failed, and in quite a number of the realizations the parameter estimates were indeed (very) far from the true values. This may be due to convergence to a 'false' local minimum.
- The cases displayed in Figure 1 show that indeed when the system becomes unidentifiable (in this case due to pole-zero cancellations), then the parameter biases will become large.
- Some MC simulations were also tried for the case when $\delta = 0.02$, but it was then hard to get similarities to the theoretical parameter biases.

V. Concluding discussion

When standard identification methods are applied to input-output data that are noise-corrupted, biased parameter estimates occur due to the presence of input noise. This paper has addressed what factors that influence the size of this bias. It has been assumed that a regular prediction error method is applied. When the input noise variance λ_u^2 is small, the bias will be small and of order $O(\lambda_u^2)$. When the system is close to not identifiable due to almost pole-zero cancellation, the bias will be large. It was shown that the bias is of order $O(1/\delta)$ where δ is the pole-zero separation.

In the presence of almost pole-zero cancellation in the system structure, two problematic issues occur. First, it is a well-known fact that the parameter estimates become uncertain (have large standard deviations), independently of the used identification method. Second, in this paper it was shown that, for this type of almost non-identifiable systems, the presence of input noise also leads to a large bias. Note that in this paper, we did not start with considering almost non-identifiable systems, and then investigating what can happen in an EIV situation. Rather, the starting point was to examine in what situations neglecting the presence of input noise can lead to a large/significant bias.

References

[1] V. Cerone, D. Regruto, and M. Abuabiah. Direct data-driven control design through set-membership errors-in-variables identification techniques. In *Proc American Control Conference (ACC)*, Seattle, WA, USA, May 24-26 2017.

[2] K. Colin, X. Bombois, L. Bako, and F. Morelli. Data informativity for the open-loop identification of MIMO systems in the prediction error framework. *Automatica*, 117:109000, 2020.

[3] M. R. Gevers, A S Bazanella, X Bombois, and L Miškovic. Identification and the information matrix: How to get just sufficiently rich? *IEEE Transactios on Automatic Control*, 54(12):272–285, 2009.

[4] R. Guidorzi and R. Diversi. Structural health monitoring application of errors-in-variables identification. In *Proc 21st IEEE Mediterranean Conference on Control and Automation (MED)*, Platanias, Greece, June 25-28 2013.

[5] R. Guidorzi, R. Diversi, and U. Soverini. The Frisch scheme in algebraic and dynamic identification problems. *Kybernetika*, 44(5):585–616, 2008.

[6] K. Lau, J. H. Braslavsky, J. C. Agüero, and G. C. Goodwin. An errors-in-variables method for non-stationary data with application to mineral exploration. *Automatica*, 45:2971–2976, 2009.

[7] J. Linder. *Indirect system identification for unknown input problems with applications to ships*. PhD thesis, Linköping University, Sweden, 2017.

[8] L. Ljung. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions on Automatic Control*, AC-30(9):834–844, 1985.

[9] L. Ljung. *System Identification - Theory for the User*. Prentice Hall, Englewood Cliffs, NJ, USA, 1987.

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

[11] J. Mårtensson, N. Everitt, and H. Hjalmarsson. Covariance analysis in SISO linear systems identification. *Automatica*, 77:82–92, 2017.

[12] J. Mårtensson and H. Hjalmarsson. How to make bias and variance errors insensitve to system and model complexity in identification. *IEEE Transactions in Automatic Control*, 56(1):100–112, January 2011.

[13] B. M. Ninness and H. Hjalmarsson. Variance error quantifications that are exact for finite model order. *IEEE Transactions on Automatic Control*, 49(8):1275–1291, 2004.

[14] A. Rougée, M. Basseville, A. Benveniste, and G. V. Moustakides. Optimum robust detection of changes in the AR part of a multivariable ARMA process. *IEEE Transactions on Automatic Control*, 32:1116–1120, 1987.

[15] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, June 2007. Survey paper.

[16] T. Söderström. System identification for the errors-in-variables problem. *Transactions of the Institute of Measurement and Control*, 34(7):780–792, October 2012.

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

[18] T. Söderström. A user perspective on errors-in-variables methods in system identification. *Control Engineering Practice*, 89:56–69, 2019. A feature paper for the series of Frontiers in Control Engineering Practice.

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

[20] K. van Heusden, A. Karimi, and T. Sï¿½derstrï¿½m. On identification methods for direct data-driven controller tuning. *International Journal of Adaptive Control and Signal Processing*, 25(5):448–465, May 2011.

[21] E. Zhang, R. Pintelon, and J. Schoukens. Errors-in-variables identification of dynamic systems excited by arbitrary non-white input. *Automatica*, 49(12):3032–3041, October 2013.

Appendix

Proof of (24)

First note that for any given zero α_j. the coefficients $a_k, k = 0, \ldots, n_a$ are affine functions of α_j. This means just that they have one constant term and one linear term. For example, it holds

$$a_1 = -a_0 \sum_k \alpha_k, \quad a_{n_a} = (-1)^{n_a} a_0 \alpha_j \prod_{k \neq j}^{n_a} \alpha_k \ .$$

Then consider the determinant as a function of α_j. Due to the above observation it follows that it must be a polynomial in α_j, and the order is n_a. This is true for all $j = 1, \ldots, n_a$.

Further, as the determinant is zero as soon as A and B have any joint zero, it now follows that the determinant can be written as

$$\det(\mathcal{S}(A,B)) = C \prod_{j=1}^{n_a} \prod_{k=1}^{n_b} (\alpha_j - \beta_k) \ ,$$

where C is a constant, that remains to be determined.

To find C, consider the product of highest powers of the zeros $\alpha_j, j = 1, \ldots n_a$ among all terms summing up to the determinant. The highest powers will be obtained precisely when considering the main diagonal. This leads to

$$\begin{aligned}
& C \prod_{j=1}^{n_a} \prod_{k=1}^{n_b} \alpha_j = b_0^{n_a} a_{n_a}^{n_b} \\
& \Rightarrow \\
& C \left(\prod_{j=1}^{n_a} \alpha_j \right)^{n_b} = b_0^{n_a} \left(a_0 (-1)^{n_a} \prod_{k=1}^{n_a} \alpha_k \right)^{n_b} \\
& = b_0^{n_a} a_0^{n_b} (-1)^{n_a \times n_b} \left(\prod_{k=1}^{n_a} \alpha_k \right)^{n_b} \\
& \Rightarrow \\
& C = (-1)^{n_a \times n_b} \ ,
\end{aligned}$$

which finally proves (24).