WHEN ARE ERRORS-IN-VARIABLES ASPECTS PARTICULARLY IMPORTANT TO CONSIDER IN SYSTEM IDENTIFICATION?

Torsten Söderström * **and Umberto Soverini** +

*Division of Systems and Control,
Department of Information Technology, Uppsala University
P O Box 337, SE-751 05 Uppsala, Sweden
Email: torsten.soderstrom@it.uu.se

+Dept of Electrical, Electronic and Information Engineering
University of Bologna
Bologna, Italy
Email: umberto.soverini@unibo.it

Keywords: system identification, errors-in-variables, bias, parameter estimation, output error model, ARMAX model

Abstract

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

1 Introduction

All standard identification methods, see for example [21], [37] 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 [32]. See also the survey paper [29]. Aspects from a user perspective on how to cope with EIV problems are discussed in [33]. For some further overviews, see [10] and [31].

For static systems, EIV representations are closely related to other well-known topics such as *latent variables* models and *factor models*, [6], [40], [27].

When the data are noisy data, the user will need to apply some assumptions. This is needed to distinguish between on one hand the effects on the recorded output of measurement noise, and on the other hand the effects of the underlying noise-free dynamics. The assumptions (*prejudices*) which lie behind the identification procedure have been thoroughly analyzed in [12], [13] with particular attention to the so called Frisch scheme, [5].

There are many papers and publications where errors-in-variables methods are used for various types of applications. As examples, one can mention identifying transmissibility functions in a mechanical mass-spring system, [41], data-driven controller design algorithms, [39], [14], [38], [3], electromagnetic mineral exploration, [17], [15], [16], roll dynamics of a ship, [18], structural health monitoring used to check the status of (large) mechanical constructions and buildings, [9], [11], [26], [1]. Additional applications are listed in e.g. [32].

The focus in this report 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. Phrased differently, when is it important to apply EIV techniques? 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. In the report 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. In this case, it is shown that EIV identification techniques lead to more accurate results. All these considerations will be supported by simple numerical examples.

The report is organized as follows. The next section describes the modelling and general problem formulation. The main ideas behind this paper are presented in Section 3. More specific and explicit analysis for some typical model structures are then given in Section 4. Section 5 describes the performance of EIV based techniques for the situation where standard identification methods give significantly biased parameter estimates. Section 6 contains some concluding discussion.

2 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

$$\begin{aligned} y(t) &= G_0(q)u_0(t) + H_0(q)e(t)\,, & (1)\\ u(t) &= u_0(t) + \tilde{u}(t)\,, & (2)\\ u_0(t) &= F(q)v(t)\,. & (3) \end{aligned}$$

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. 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 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.

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

$$y(t) = G(q)u(t) + H(q)\varepsilon(t) \qquad (4)$$

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

$$G(q,\theta_*) \equiv G_0(q), \;\; H(q,\theta_*) \equiv H_0(q)\,. \qquad (5)$$

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

$$\tilde{\theta} = \hat{\theta} - \theta_*\,. \qquad (6)$$

For consistency it is required that the bias is zero. For finite N the situation is more subtle, and it is then hard to have precise expressions for the bias. The focus here is therefore 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 problems to be discussed are the following:

- **Problem 1**. 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?
- **Problem 2**. Another aspect concerns the fact that the bias is a vector. Are there pertinent ways to express the size of the bias? Can one use a scalar measure? Is it possible to weight the importance of different components?
- **Problem 3**. For cases when the bias is large (or significant), how does the EIV estimates behave? Assuming that the presence of input noise is taken into account, what are the properties of the parameter estimates?

Problems 1 and 2 are treated in Section 3, while Section 4 contains a more explicit and detailed analysis of Problem 1. Section 5 is devoted to a treatment of Problem 3.

3 Main ideas

3.1 Treatment of Problem 1

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, [20], [37].

The PEM means that the parameter estimate can be written as

$$\begin{aligned} \hat{\theta} &= \arg\min_\theta V(\theta)\,, & (7)\\ V(\theta) &= \frac{1}{2}E\left\{\varepsilon^2(t,\theta)\right\}\,. & (8) \end{aligned}$$

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

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

There is no general analytical 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

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

leading to

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

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

$$\begin{aligned} \varepsilon_\theta &= \frac{\partial \varepsilon(t,\theta)}{\partial \theta}\,, & (12)\\ G_\theta &= \frac{\partial G(q,\theta)}{\partial \theta}\,, & (13)\\ H_\theta &= \frac{\partial H(q,\theta)}{\partial \theta}\,. & (14) \end{aligned}$$

Direct differentiation gives

$$\begin{aligned} V_\theta' &= E\left\{\varepsilon(t,\theta)\varepsilon_\theta(t,\theta)\right\}\,, & (15)\\ V_{\theta\theta}'' &= E\left\{\varepsilon(t,\theta)\varepsilon_{\theta\theta}(t,\theta)\right\} + E\left\{\varepsilon_\theta(t,\theta)\varepsilon_\theta^T(t,\theta)\right\} & (16) \end{aligned}$$

and also

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

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

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

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

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:

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

Recall now (11), which describes the bias error as

$$\tilde{\theta} = Pr\lambda_u^2\ , \qquad (21)$$

where

$$\begin{aligned}P &= -\left[V_{\theta\theta}''(\theta_*)\right]^{-1}\ , \qquad (22)\\ r &= V_\theta'(\theta_*)/\lambda_u^2\ . \qquad (23)\end{aligned}$$

It holds that $\tilde{\theta} \to 0$ as $\lambda_u^2 \to 0$ (the bias will be small, if the input noise level is small). Further, we can expect r to be 'modest' or 'well bounded' in size. See equation (20), that shows how r is directly proportional to the input noise variance λ_u^2. Therefore, for the bias $\tilde{\theta}$ to be large, it is needed that P is large. This occurs when $V_{\theta\theta}''(\theta_*)$ 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 4 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 [19], [25], [4], [7], [24], [22]. 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 estimate $\hat{\theta}$ will be large.

3.2 Treatment of Problem 2

Next consider the possibility to use a scalar measure for expressing the size of the bias vector. The norm $\| \tilde{\theta} \|$ is not a good measure of the size of the bias. Indeed, the different components of the parameter vector should not and cannot be compared directly. Indeed they may represent different physical quantities, and how to compare 'apples' and 'pears' is then nontrivial. Some sort of weighted norm is therefore a better alternative, such as

$$W = \tilde{\theta}^T Q \tilde{\theta}\ . \qquad (24)$$

where Q is a nonnegative weighting matrix. Further, a measure like the one in (24) should have a physical meaning, which implies that the elements of the weighting matrix Q have different physical units.

Example 1. One way to express the size of the bias vector is to compare the true transfer function G with the one associated with the parameter estimate $\hat{\theta}$. As a measure of the difference between the true and the estimated transfer function, one may use the L^2-norm, and assume that the bias vector is small. Then using linearization,

$$\begin{aligned}W &= \int |\hat{G}(e^{i\omega}) - G(e^{i\omega})|^2 d\omega \approx \int |G_\theta(e^{i\omega})\tilde{\theta}|^2 d\omega \\ &= \tilde{\theta}^T \underbrace{\int G_\theta^T(e^{-i\omega})G_\theta(e^{i\omega})d\omega}_{Q}\,\tilde{\theta}\ , \qquad (25)\end{aligned}$$

and (25) can be taken as a definition of a weighting matrix Q. In general Q in (25) will be positive semidefinite (and singular), unless the transfer function G explicitly depends on *all* the components of the parameter vector θ.

Example 2. Another way of introducing a weighting matrix Q is to consider predictability. How much is the prediction error variance degraded when an estimated model is used, instead of a true system description? Again by linearization, assuming a small bias, one can obtain

$$V = E\left\{\varepsilon^2(t,\hat{\theta})\right\} \approx E\left\{\left[\varepsilon(t,\theta_*) + \varepsilon_\theta(t,\theta_*)\tilde{\theta}\right]^2\right\}$$

$$= E\left\{e^2(t)\right\} + \tilde{\theta}^T \underbrace{E\left\{\varepsilon_\theta^T(t,\theta_*)\varepsilon_\theta(t,\theta_*)\right\}}_{Q}\tilde{\theta}\,. \quad (26)$$

4 Detailed analysis of Problem 1

4.1 Preliminaries

This section gives more explicit forms of the analysis of Section 3.1. Specialization to the output error model structure is given in Section 4.2, and the case of the ARMAX model structure is described in Section 4.3. Section 4.4 generalizes the output error model structure to a general linear model with correlated output noise. Section 4.5 is devoted to the case of almost unidentifiable systems, and what happens when they are identified using reduced order models.

First some preliminaries for the so called Sylvester matrices are given in this subsection.

Consider two polynomials

$$\begin{aligned} A &= a_0 z^{n_a} + a_1 z^{n_a-1} + \ldots + a_{n_a}\,, & (27)\\ B &= b_0 z^{n_b} + b_1 z^{n_b-1} + \ldots + b_{n_b}\,. & (28)\end{aligned}$$

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

$$\mathcal{S}(A,B) = \begin{pmatrix} b_0 & b_1 & \ldots & b_{n_b} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & b_0 & b_1 & \ldots & b_{n_b} \\ a_0 & a_1 & \ldots & a_{n_a} & & 0 \\ 0 & \ddots & & & \ddots & \\ & & a_0 & a_1 & \ldots & a_{n_a} \end{pmatrix}. \quad (29)$$

The properties of Sylvester matrices have been investigated in many sources. Some basic properties are, for example, considered in [37]. 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 [37].
- 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

$$\det\left(\mathcal{S}(A,B)\right) = (-1)^{n_a \times n_b} a_0^{n_b} b_0^{n_a} \prod_{j=1}^{n_a}\prod_{k=1}^{n_b}(\alpha_j - \beta_k)\,. \quad (30)$$

Observe that $(-1)^{n_a \times n_b} = 1$ unless both n_a and n_b are odd numbers. A proof of (30) is given in the appendix.

4.2 The output error model structure

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

$$\begin{aligned} y(t) &= y_0(t) + \tilde{y}(t), \quad E\{\tilde{y}^2(t)\} = \lambda_y^2\,, & (31)\\ u(t) &= u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2\,, & (32)\\ Ay_0(t) &= Bu_0(t)\,, & (33)\\ A &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}\,, & (34)\\ B &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}\,. & (35)\end{aligned}$$

The equation (33) refers to the model to be fitted. The true data ('the system') is assumed to also fulfil (33), 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

$$V = \frac{1}{2}E\{\varepsilon^2(t)\}\,, \quad (36)$$

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

$$\varepsilon(t) = y(t) - \frac{B}{A}u(t) = \left(\frac{B_0}{A_0} - \frac{B}{A}\right)u_0(t) + \tilde{y}(t) - \frac{B}{A}\tilde{u}(t)\,. \quad (37)$$

Its gradient is easily found to be

$$\begin{aligned} \varepsilon'_\theta(t) &= \left(\begin{array}{cccc} \frac{B}{A^2}q^{-1}u(t) & \ldots & \frac{B}{A^2}q^{-n_a}u(t) \\ -\frac{1}{A}q^{-1}u(t) & \ldots & -\frac{1}{A}q^{-n_b}u(t) \end{array}\right)^T \\ &= \mathcal{S}(-A,B)\frac{1}{A^2}\begin{pmatrix} u(t-1) \\ \vdots \\ u(t-n_a-n_b) \end{pmatrix} \\ &\triangleq \mathcal{S}(-A,B)\varphi_u(t)\,. & (38)\end{aligned}$$

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

$$\begin{aligned} V''_{\theta\theta}(\theta_*) &= \mathcal{S}(-A_0,B_0)P_{\varphi_u}\mathcal{S}^T(-A_0,B_0) \\ &= \mathcal{S}(-A_0,B_0)P_{\varphi_{u_0}}\mathcal{S}^T(-A_0,B_0) \\ &\quad +\mathcal{S}(-A_0,B_0)P_{\varphi_{\tilde{u}}}\mathcal{S}^T(-A_0,B_0)\,, & (39)\\ V'_\theta(\theta_*) &= E\{\varepsilon(t)\varepsilon'(t,\theta_*)\} \\ &= -\mathcal{S}(-A_0,B_0)E\{\frac{B_0}{A_0}\tilde{u}(t)\varphi_{\tilde{u}}(t)\} \\ &\triangleq -\mathcal{S}(-A_0,B_0)r_0\,. & (40)\end{aligned}$$

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

$$\beta_1 = -\left(V''_{\theta\theta}(\theta_*)\right)^{-1}V'_\theta(\theta_*) = \mathcal{S}^{-T}(-A_0,B_0)P_{\varphi_u}^{-1}r_0\,. \quad (41)$$

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

$$\beta_2 = \mathcal{S}^{-T}(-A_0,B_0)P_{\varphi_{u_0}}^{-1}r_0\,. \quad (42)$$

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 (36).

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

$$\delta = \min_{i,j} |p_i - z_j| \tag{43}$$

which is a measure of the pole-zero separation. It now follows from (30) 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 (41) and (42) 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,

$$u_0(t) = Fv(t), \quad F = \left(1 - 0.9q^{-1}\right)^{-1}, \quad E\{v^2(t)\} = 1\,. \tag{44}$$

The other parameters in the numerical example are

$$a_1 = -0.8, \quad \lambda_y^2 = 10, \quad b_1 = 2\,. \tag{45}$$

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.

In the numerical study the approximate bias expressions β_1, see (41), and β_2, see (42) were computed. They are compared to the 'true' bias β_t, computed by minimizing the loss function (36). 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.

As a further examination of how the bias errors are influenced by the pole-zero separation δ, and by the input noise variance λ_u^2 let both these two quantities vary, and examine the contour levels of the biases.

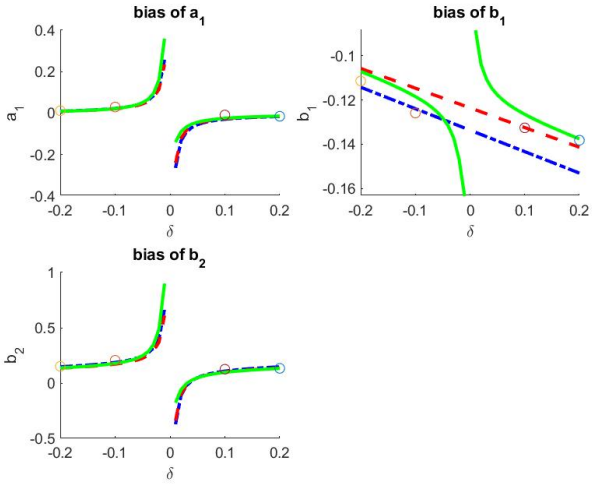

Figure 1: Parameter biases versus δ. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100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

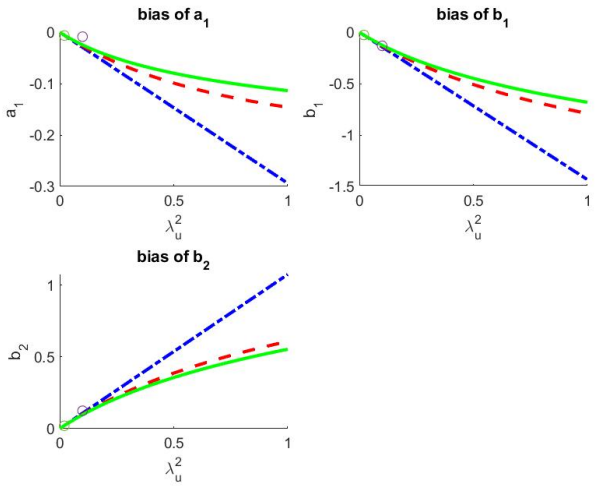

Figure 2: Parameter biases versus λ_u^2. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from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

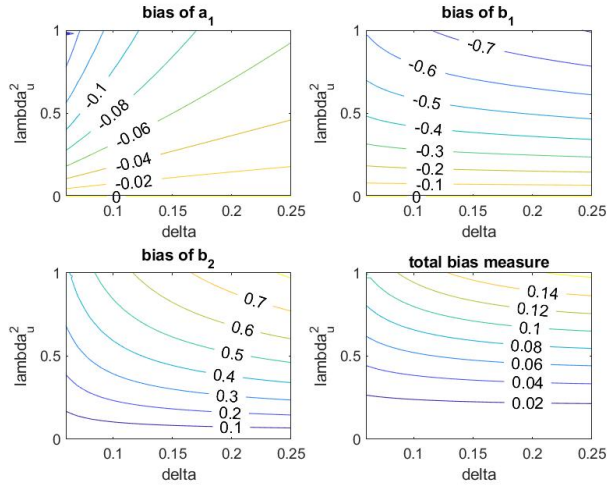

Figure 3: Contour levels for parameter biases and the total bias measure δG as functions of δ and λ_u^2, $\delta > 0$.

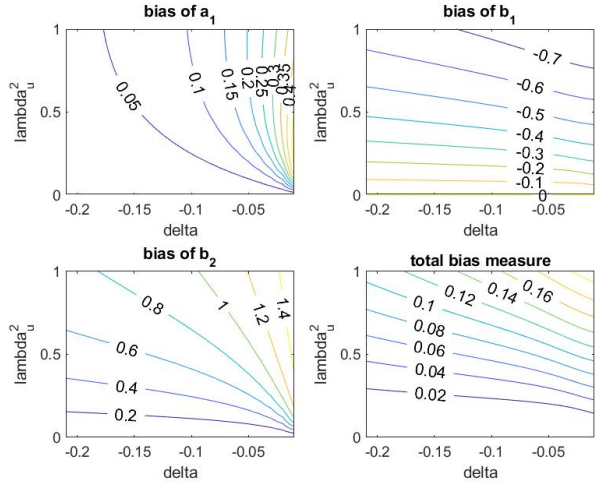

Figure 4: Contour levels for parameter biases and the total bias measure δG as functions of δ and λ_u^2, $\delta < 0$.

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

$$\delta G = \frac{\int |\frac{B(e^{i\omega})}{A(e^{i\omega})} - \frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 d\omega}{\int |\frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 d\omega} \ . \tag{46}$$

This is indeed a scaled version of the criterion W in (25).

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.

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.

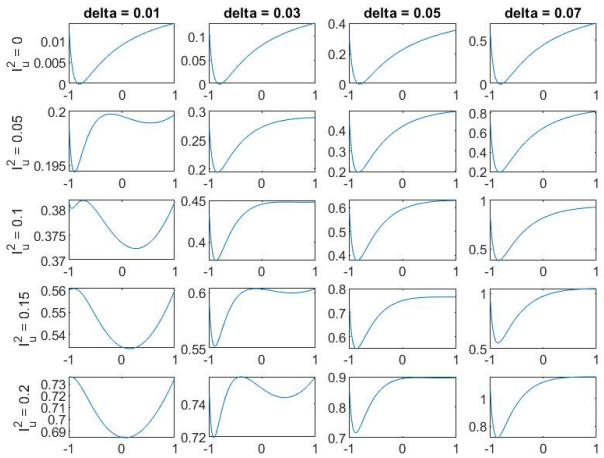

Figure 5: The graph of $W(a)$ for various values of the parameters δ and λ_u^2.

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

$$\begin{aligned} V(a, b_1, b_2) &= E\{[y(t) - \frac{B}{A}u(t)]^2\} - \lambda_y^2 \\ &= E\{[\left(\frac{B_0}{A_0} - \frac{B}{A}\right) u_0(t) - \frac{B}{A}\tilde{u}(t)]^2\} \ . \end{aligned} \tag{47}$$

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

$$W(a) = \min_{b_1, b_2} V(a, b_1, b_2) \ . \tag{48}$$

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.

4.2.1 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 approximations 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.

4.3 The ARMAX model structure

Consider now an ARMAX model. This means that G and H in (4) are given the same denominator.

This case is characterized by the following equations

$$\begin{aligned} y(t) &= y_0(t) + C/A\tilde{y}(t), \quad E\{\tilde{y}^2(t)\} = \lambda_y^2 , & (49)\\ u(t) &= u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2 , & (50)\\ Ay_0(t) &= Bu_0(t) , & (51)\\ A &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} , & (52)\\ B &= b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} , & (53)\\ C &= 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c} . & (54) \end{aligned}$$

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

Assume that the true data corresponds to the transfer functions

$$G = \frac{B_1}{A_1}, \quad H = \frac{C_1}{D_1} . \tag{55}$$

This leads quickly to the following ARMAX model polynomials

$$A = A_1 D_1, \quad B = B_1 D_1, \quad C = C_1 A_1 . \tag{56}$$

In case the two transfer functions G and H has 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 (36), but now the error (in fact, the one-step ahead prediction error) should instead of (37) be written as

$$\begin{aligned} \varepsilon(t,\theta) &= \frac{A}{C} y(t) - \frac{B}{C} u(t) \\ &= \frac{AB_0 - A_0 B}{A_0 C} u_0(t) + \frac{A}{C}\frac{C_0}{A_0}\tilde{y}(t) - \frac{B}{C}\tilde{u}(t) . \end{aligned} \tag{57}$$

Needless to say, (57) 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

$$\theta = \begin{pmatrix} a_1 & \ldots & a_{n_a} & b_1 & \ldots & b_{n_b} & c_1 & \ldots & c_{n_c} \end{pmatrix}^T . \tag{58}$$

The derivatives of the polynomials can be written as

$$\begin{aligned} A_\theta &= \begin{pmatrix} q^{-1} & \ldots & q^{-n_a} & O_{1\times(n_b+nb_c)} \end{pmatrix} , & (59)\\ B_\theta &= \begin{pmatrix} O_{1\times n_a} & q^{-1} & \ldots & q^{-n_b} & O_{1\times n_c} \end{pmatrix} , & (60)\\ C_\theta &= \begin{pmatrix} O_{1\times(n_a+n_b)} & q^{-1} & \ldots & q^{-n_c} \end{pmatrix} . & (61) \end{aligned}$$

When evaluating an expression in (19) one gets

$$\begin{aligned} GH_\theta - G_\theta H &= \frac{B_0}{A_0}\left(-\frac{C_0 A_\theta}{A_0^2} + \frac{C_\theta}{A_0}\right) \\ &\quad -\frac{C_0}{A_0}\left(-\frac{B_0 A_\theta}{A_0^2} + \frac{B_\theta}{A_0}\right) \\ &= -\frac{C_0}{A_0^2} B_\theta + \frac{B_0}{A_0^2} C_\theta . \end{aligned} \tag{62}$$

Using the above expressions in (18), (19) leads to

$$\varepsilon(t,\theta_*) = e(t) - \frac{A_0}{C_0}\frac{B_0}{A_0}\tilde{u}(t) = e(t) - \frac{B_0}{C_0}\tilde{u}(t) , \tag{63}$$

$$\begin{aligned} \varepsilon_\theta(t,\theta_*) &= -\frac{-B_0 A_\theta + A_0 B_\theta}{A_0 C_0} u_0(t) \\ &\quad + \frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t) \\ &\quad - \frac{-C_0 A_\theta + A_0 C_\theta}{A_0 C_0} e(t) . \end{aligned} \tag{64}$$

The gradient of V in (11) becomes

$$V_\theta'(\theta_*) = -E\left\{\left[\frac{B_0}{C_0}\tilde{u}(t)\right]\left[\frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t)\right]\right\} , \tag{65}$$

while the Hessian will be

$$\begin{aligned} V_{\theta\theta}''(\theta_*) &= \operatorname{cov}\left[\frac{-B_0 A_\theta + A_0 B_\theta}{A_0 C_0} u_0(t)\right] \\ &\quad + \operatorname{cov}\left[\frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t)\right] \\ &\quad + \operatorname{cov}\left[\frac{-C_0 A_\theta + A_0 C_\theta}{A_0 C_0} e(t)\right] . \end{aligned} \tag{66}$$

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

$$\phi_1(t) = \frac{-C_0 B_\theta + B_0 C_\theta}{C_0^2}\tilde{u}(t) . \tag{67}$$

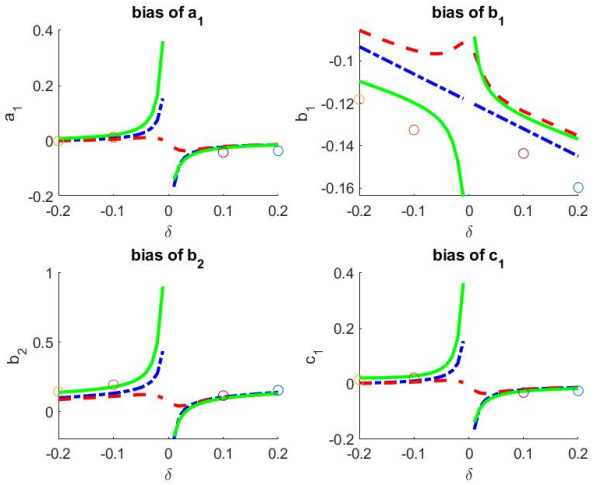

Figure 6: Parameter biases versus δ. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the input noise variance was $\lambda_u^2 = 0.1$.

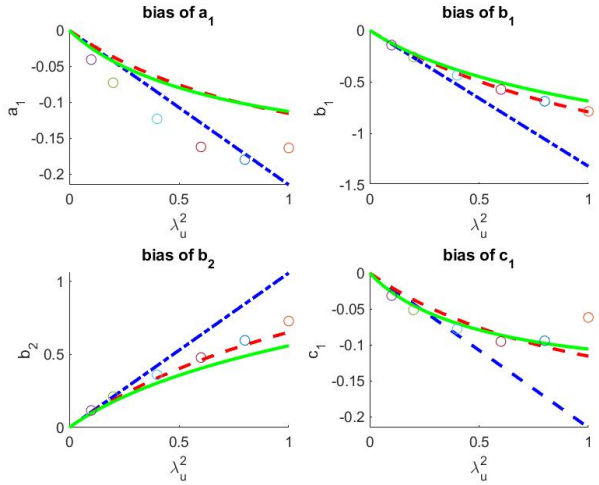

Figure 7: Parameter biases versus λ_u^2. The true biases (β_t) are shown with solid lines. The approximate biases (β_1) are shown with dashed lines. The cruder approximate biases (β_2) are shown with dashdotted lines. The circles show the empirical biases obtained by the Monte Carlo simulations from 100 realizations of length 1000. The value of the parameter δ is $\delta = 0.1$.

It follows from (60) and (61) 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

$$\begin{pmatrix} -C_0 q^{-1} & \ldots & -C_0 q^{-n_b} & B_0 q^{-1} & \ldots & B_0 q^{-n_c} \end{pmatrix} \frac{1}{C_0^2} \tilde{u}(t)$$
$$= \frac{1}{C_0^2} \begin{pmatrix} q^{-1}\tilde{u}(t) & \ldots & q^{-n_b - n_c}\tilde{u}(t) \end{pmatrix} \mathcal{S}^T(-C_0, B_0) \,. \tag{68}$$

Numerical example

As a numerical illustration one can now consider the same example and data as examined for the output error case in Section 4.2. 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 $B = 2(1 - (0.8 + \delta)q^{-1})$.

The numerical investigations are displayed in Figures 6 and 7.

4.3.1 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 (but not always) similar results to those predicted by theory.

4.4 A general linear model structure

Consider in this subsection the general linear model structure (4), 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 correlated noise.

Specifically, assume

$$G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}} \,, \tag{69}$$
$$H(q) = \frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d}} \,. \tag{70}$$

Then the prediction error becomes

$$\varepsilon(t,\theta) = \frac{D}{C}\left[y(t) - \frac{B}{A}u(t)\right] \tag{71}$$
$$= \frac{D}{C}\left[\frac{B_0}{A_0}u_0(t) + \frac{C_0}{D_0}e(t) - \frac{B}{A}\left(u_0(t) + \tilde{u}(t)\right)\right] \,. \tag{72}$$

Its gradient fulfils

$$\frac{\partial \varepsilon}{\partial a_i}(t) = \frac{DBq^{-i}}{CA^2}u(t) \,, \tag{73}$$
$$\frac{\partial \varepsilon}{\partial b_i}(t) = -\frac{Dq^{-i}}{CA}u(t) \,, \tag{74}$$
$$\frac{\partial \varepsilon}{\partial c_i}(t) = -\frac{Dq^{-i}}{C^2}\left(y(t) - \frac{B}{A}u(t)\right) \,, \tag{75}$$
$$\frac{\partial \varepsilon}{\partial d_i}(t) = \frac{q^{-i}}{C}\left(y(t) - \frac{B}{A}u(t)\right) \,. \tag{76}$$

This leads to

$$\varepsilon'_\theta(t) = \begin{pmatrix} \mathcal{S}(-A,B)\varphi_1(t) \\ \\ \mathcal{S}(C,-D)\varphi_2(t) \end{pmatrix} , \tag{77}$$

where

$$\varphi_1(t) = \frac{D}{A^2C}\begin{pmatrix} q^{-1} \\ \vdots \\ q^{-n_a-n_b} \end{pmatrix} u(t) , \tag{78}$$

$$\varphi_2(t) = \frac{1}{C^2}\begin{pmatrix} q^{-1} \\ \vdots \\ q^{-n_c-n_d} \end{pmatrix} \left(y(t) - \frac{B}{A}u(t) \right) . \tag{79}$$

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

$$V'_\theta(\theta_*) = \begin{pmatrix} \mathcal{S}(-A_0,B_0)r_0 \\ \mathcal{S}(C_0,-D_0)r_1 \end{pmatrix} , \tag{80}$$

$$r_0 = -E\left\{ \frac{B_0}{A_0}\tilde{u}(t)\frac{D_0}{A_0^2C_0}\begin{pmatrix} \tilde{u}(t-1) \\ \vdots \\ \tilde{u}(t-n_a-n_b) \end{pmatrix} \right\} , \tag{81}$$

$$r_1 = E\left\{ e(t)\frac{1}{C_0D_0}\begin{pmatrix} e(t-1) \\ \vdots \\ e(t-n_c-n_d) \end{pmatrix} \right\} , \tag{82}$$

$$V''_{\theta\theta}(\theta_*) = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix} , \tag{83}$$

$$V_{11} = \mathcal{S}(-A_0,B_0)P_{\varphi_1}\mathcal{S}^T(-A_0,B_0) , \tag{84}$$

$$V_{22} = \mathcal{S}(C_0,-D_0)P_{\varphi_2}\mathcal{S}^T(C_0,-D_0) . \tag{85}$$

4.4.1 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.
- Similarly, If C_0 and D_0 have almost a pole-zero cancellation, then the estimates of C and D are quite uncertain.

4.5 Using a reduced model structure

It was found above that when the pole-zero separation δ is small is can be very hard to identify the system using the postulated model structure. For identification, it may be an alternative to instead using a reduced model where a pole-zero cancellation is enforced. This idea is examined in this subsection.

As models of different orders are to be compared it is then meaningful to use the criterion (46) as a measure of fit.

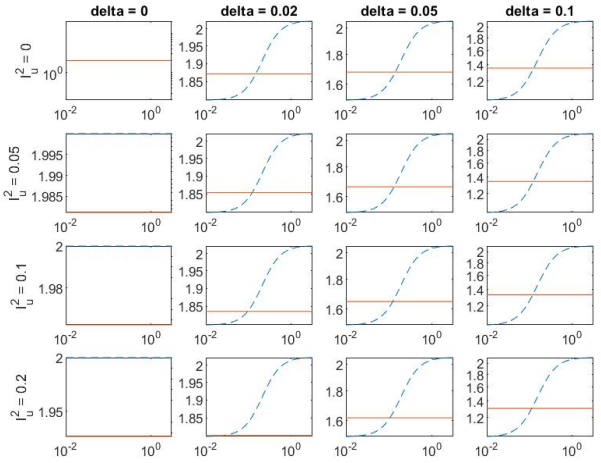

Figure 8: Frequency plots of the true system (dash-dotted lines) and the approximate models (solid lines) for some different values of λ_u^2 and δ.

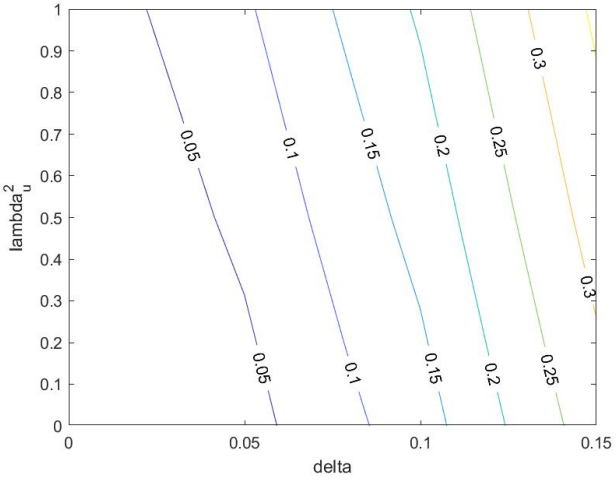

Figure 9: Contour plots of the criterion δG.

Numerical example

The same numerical example as in (44), (45) is considered.

Figure 8 shows the frequency plots of the true system (dash-dotted lines) and the approximate models (solid lines) for some different values of λ_u^2 and δ. As the true system is of first order, the reduced model will be a constant in this case. Its value is determined as the minimizer of the (asymptotic) output error criterion, taking the character of the noise-free input into account.

Figure 9 shows contour plots of the criterion δG which expresses the deviation of the estimated transfer function G of the reduced order model from the true system transfer function G_0.

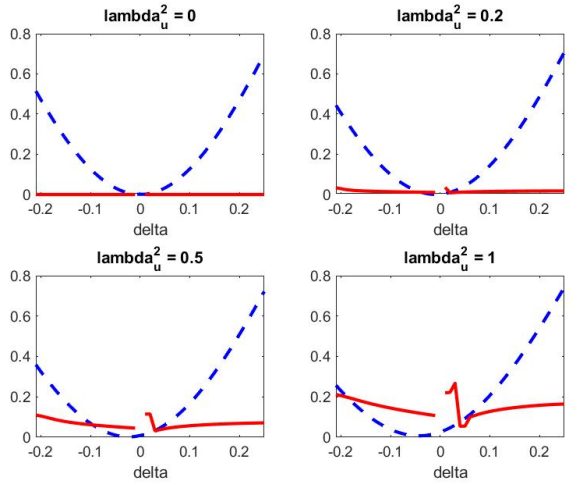

Figure 10: The criterion δG as a function of δ for some different values of λ_u^2. The dashed lines refer to the deviation between the true system and the reduced order model. The solid lines refer to the deviation between the true system and the full order model. The irregular behaviour of the solid lines for small positive values of δ is due to the existence of local minimum points, cf the discussion in Section 4.2.

Finally, Figure 10 shows how the criterion δG varies with the pole-zero separation δ for some different values of the input noise variance λ_u^2.

4.5.1 Some observations

- The difference between the true system transfer function and that of the reduced order model is increasing when the noise variance λ_u^2 is increasing. It is also increasing when the pole-zero separation δ is increasing.
- When the pole-zero separation δ is small, a reduced order model gives a better fit than a full order model. This phenomenon happens to be more pronounced in the example for negative values of δ.

5 Treatment of Problem 3: How accurate are EIV estimates for small pole-zero separations?

5.1 Introduction

So far it was shown that when a system is almost not identifiable due to an almost pole-zero cancellation, then there will be a considerable bias in the parameter estimates when EIV aspects are not taken into account. Concerning the covariance matrix a crude approximation would be that it is proportional to $V_{\theta\theta}''(\theta_*)$ just as in the case of no input noise.

More specifically, assume that there is a pole-zero distance δ that is small. It was shown that, when neglecting that there is input noise in the measurement and applying a standard identification method:

- There will be a bias that is $O(1/\delta)$.
- The estimates will have a standard deviation that is $O(1/\delta)$.

What happens if an identification method taking the EIV aspects into account is applied? When $\delta \to 0$, identifiability is lost. On the other hand, for small δ there should not be any systematic estimation errors. One would therefore expect:

- There will be no asymptotic bias.
- The estimates will have a standard deviation that is $O(1/\delta)$.

The purpose of this section is to examine and illustrate the standard deviation of the parameter estimates, for small values of the pole-zero separation, for a simple example.

5.2 The example

Consider the following case, a first order output error model structure. This is the same example as studied earlier in the report.

$$y(t) = y_0(t) + \tilde{y}(t), \quad E\{\tilde{y}^2(t)\} = \lambda_y^2 , \tag{86}$$

$$u(t) = u_0(t) + \tilde{u}(t), \quad E\{\tilde{u}^2(t)\} = \lambda_u^2 , \tag{87}$$

$$Ay_0(t) = Bu_0(t) , \tag{88}$$

$$A = 1 + a_1 q^{-1} , \tag{89}$$

$$B = b_1 q^{-1} + b_2 q^{-2}, \quad b_2 = b_1(a - \delta) , \tag{90}$$

$$u_0(t) = Fv(t), \quad F = \left(1 - 0.9q^{-1}\right)^{-1}, \quad E\{v^2(t)\} = 1 , \tag{91}$$

$$a_1 = -0.8, \quad \lambda_y^2 = 10, \quad b_1 = 2 . \tag{92}$$

It is assumed that both the input noise $\tilde{u}(t)$ and the output noise $\tilde{y}(t)$ are white.

A number of different identification methods are considered. They all apply to EIV models, and their asymptotic distribution of the parameter estimates are known. The asymptotic normalized standard deviations of the estimates are computed numerically, and the influence of the pole-zero separation δ is examined in particular.

The following methods are considered:

1. Generalized IV estimates (GIVE), see [30]. This class of identification methods include the Frisch scheme, see [2], [8], [5], as well as the bias-eliminating method (BELS), see [43], [42]. An overview of how these seemingly different method can be viewed as alternative algorithms for solving the same set of nonlinear equations, such as (93) below, appear in [34].

In GIVE, the following, possibly overdetermined, system of equations is considered:

$$\frac{1}{N}\sum_{t=1}^{N} z(t)y(t) = \frac{1}{N}\sum_{t=1}^{N} z(t)\varphi^T(t)\theta \ , \tag{93}$$

$$\varphi^T(t) = \begin{pmatrix} -y(t-1) & u(t-1) & u(t-2) \end{pmatrix} , \tag{94}$$

$$z(t) = \begin{pmatrix} y(t) & \dots & y(t-1-s_y) \\ u(t-1) & \dots & u(t-2-s_u) \end{pmatrix}^T , \tag{95}$$

with $\theta, \lambda_y^2, \lambda_u^2$ as unknowns.

GIVE is applied here for two cases:

- GIVE1, with a minimal number of equations ($s_u + s_y = 1$). This choice is taken as

$$s_y = 1, \ \ s_u = 0 \ . \tag{96}$$

- GIVE2, which corresponds to an overdetermined system of equations. The selected choice corresponds to

$$s_y = 3, \ \ s_u = 2 \ . \tag{97}$$

2. Covariance matching method (CM), see [36], [35].

The GIVE estimates are obtained as a weighted nonlinear least squares solution to a system of equations, see (93), formed by covariance elements $r_y(\tau), \ r_u(\tau), \ r_{yu}(\tau)$, for a number of τ values. The aim of the CM method is to make a more efficient use of the information in these sample covariance elements.

In the general CM case (as implemented) the used covariance elements are

$$\begin{aligned} &r_y(\tau), \tau = 0, \dots, p_y, \quad r_u(\tau), \tau = 0, \dots, p_u, \\ &r_{yu}(\tau), \tau = p_1, \dots, p_2 \ . \end{aligned} \tag{98}$$

CM is applied for two cases:

- CM1, which corresponds to the same covariance elements as GIVE1. This choice corresponds to

$$p_y = 2, p_u = 1, p_1 = -1, p_2 = 2 \ . \tag{99}$$

- CM2, which corresponds to the same covariance elements as GIVE2. This choice corresponds to

$$p_y = 4, p_u = 3, p_1 = -3, p_2 = 4 \ . \tag{100}$$

3. ML, the maximum likelihood estimate, [20] using the full EIV model structure. In this case also the input model $u_0(t) = Fv(t)$ is estimated from the data. Assuming $v(t), e(t), \tilde{u}(t)$ are independent white and Gaussian noise, the asymptotic covariance matrix is given by the Cramér-Rao lower bound. A way to compute it is presented in [28].

The asymptotic covariance matrices of all the parameter estimates are computed numerically for a number of δ values. The matrices are normalized, such that for a large N the result should be divided by N. The standard deviations are computed for the estimates of a_1, b_1, b_2, λ_y^2 and λ_u^2.

The results are plotted in Figures 11 and 12.

5.2.1 Some observations

- The ML estimates are much more accurate than those of the other methods. Further, for ML only the estimates of a_1 and b_2 are sensitive to the pole-zero gap δ.

- Similarly, for GIVE2 and CM2, only the estimates of a_1 and b_2 are sensitive to the pole-zero gap δ.

- It is fairly natural that the estimates of a_1 and b_2 are sensitive to the pole-zero gap δ, as they are linked to the pole and zero positions. The parameter b_1 on the other hand is primarily linked to the static gain of the system.

- Both GIVE1 and CM1 are quite sensitive (show large standard deviations) for small values of δ.

- GIVE2 shows great improvement over GIVE1. Similarly, CM2 shows great improvement over CM1.

It was thus found that when there is a small pole-zero separation, the parameter estimates become uncertain, which is fairly natural. In fact, for the model structure of this subsection one can show by straightforward, but lengthy, calculations that for small δ

$$\mathrm{std}(\hat{a}_1) = O(1/\delta), \ \ \mathrm{std}(\hat{b}_1) = O(1), \ \ \mathrm{std}(\hat{\delta}) = O(1) \ . \tag{101}$$

A much more general of the accuracy of estimated poles and zeros can be found in [23].

6 Conclusions

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 $O(\lambda_u^2)$. 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 $O(1/\delta)$ where δ is the pole-zero separation. Finally, when there is a small pole-zero separation and some input noise, a reduced order model will typically give a better fit than using the 'true' model order.

If instead a model structure incorporating the presence of input noise is applied, the estimates are consistent, but some of the estimates show a standard deviation $O(1/\delta)$ for a small pole-zero separation.

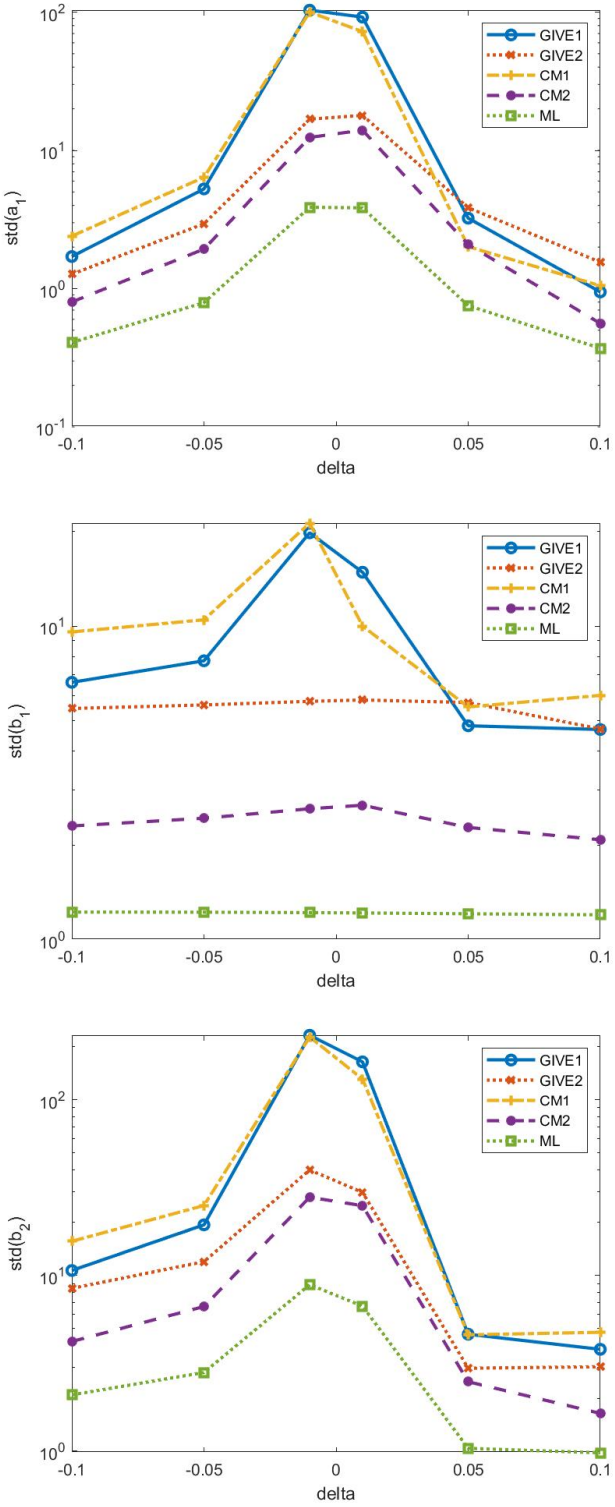

Figure 11: Standard deviations of the estimated a_1, b_1 and b_2.

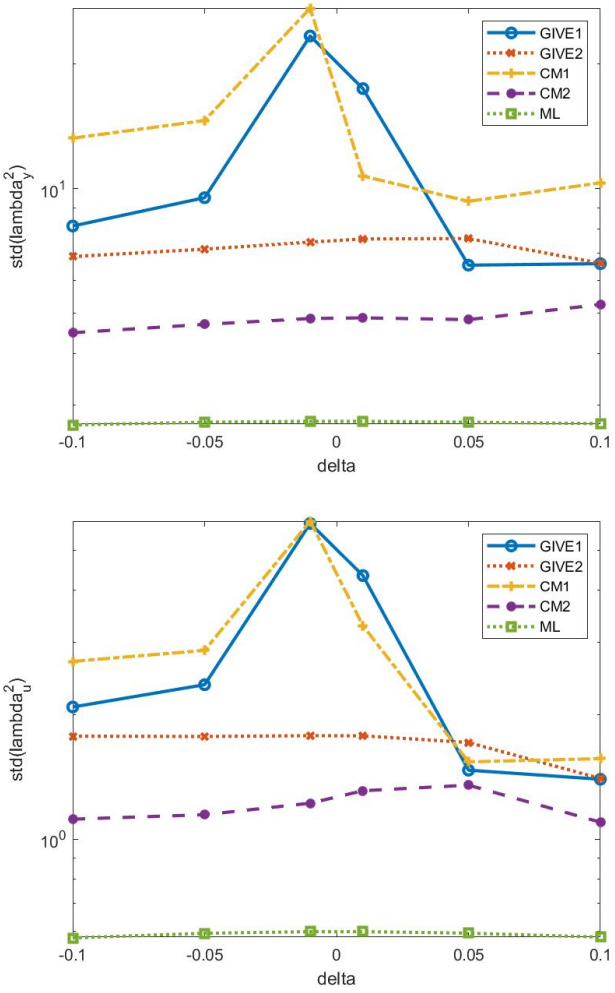

Figure 12: Standard deviations of the estimated λ_y^2 and λ_u^2.

References

[1] M. Basseville, A. Benveniste, G. V. Moustakides, and A. Rougée. Optimal sensor location for detecting changes in dynamical behavior. *IEEE Transactions on Automatic Control*, 32(12):1067–1075, December 1987.

[2] S. Beghelli, R.P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.

[3] 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.

[4] 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.

[5] R. Frisch. Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway, 1934.

[6] W. A. Fuller. *Measurement Error Models*. Wiley, New York, NY, 1987.

[7] 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.

[8] R. Guidorzi. Identification of multivariable processes in the Frisch scheme context. In *Proc 12th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, St Louis, USA, 1996.

[9] 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.

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

[11] R. Guidorzi, R. Diversi, L. Vincenzi, and V. Simioli. AR + noise versus AR and ARMA models in SHM-oriented identification. In *Proc 23rd IEEE Mediterranean Conference on Control and Automation (MED)*, Torremolinos, Spain, June 16-19 2015.

[12] R. E. Kalman. Identification from real data. In M. Hazewinkel and A. H. G. Rinnoy Kan, editors, *Current Developments in the Interface: Economics, Econometrics, Mathematics*, Dordrecht, The Netherlands, 1982. D. Reidel.

[13] R. E. Kalman. System identification from noisy data. In A. R. Bednarek and L. Cesari, editors, *Dynamical Systems II*, New York, N.Y., 1982. Academic Press.

[14] A. Karimi, K. van Heusden, and D. Bonvin. Non-iterative data-driven controller tuning using the correlation approach. In *Proc European Control Conference*, Kos, Greece, July 2-5 2007.

[15] K. Lau, J. H. Braslavsky, J. C. Agüero, and G. C. Goodwin. Application of non-stationary EIV methods to transient electromagnetic mineral exploration. In *Proc 17th IFAC World Congress*, Seoul, Korea, July 6-11 2008.

[16] 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.

[17] K. Lau, J. H. Braslavsky, and G. C. Goodwin. Errors-in-variables problems in transient electromagnetic mineral exploration. In *Proc 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, December 12-14 2007.

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

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

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

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

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

[23] J. Mårtensson and H. Hjalmarsson. Variance-error quantification of identified poles and zeros. *Automatica*, 45(11):2512–2525, November 2009.

[24] 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.

[25] 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.

[26] 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.

[27] W. Scherrer and M. Deistler. A structure theory for linear dynamic errors-in-variables models. *SIAM Journal on Control and Optimization*, 36(6):2148–2175, November 1998.

[28] T. Söderström. On computing the Cramér-Rao bound and covariance matrices for PEM estimates in linear state space models. In *Proc 14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

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

[30] T. Söderström. A generalized instrumental variable estimation method for errors-in-variables identification problems. *Automatica*, 47(8):1656–1666, August 2011.

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

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

[33] 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.

[34] T. Söderström, R. Diversi, and U. Soverini. A generalized instrumental variable framework for EIV identification methods in the presence of mutually correlated noises. Technical Report 2014-008, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2014. Available as http://www.it.uu.se/research/publications/reports/2014-008.

[35] T. Söderström, D. Kreiberg, and M. Mossberg. Extended accuracy analysis of a covariance matching approach for identifying errors-in-variables systems. *Automatica*, 50(10):2597–2605, October 2014.

[36] T. Söderström, M. Mossberg, and M. Hong. A covariance matching approach for identifying errors-in-variables systems. *Automatica*, 45(9):2018–2031, September 2009.

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

[38] K. van Heusden, A. Karimi, and D. Bonvin. Data-driven model reference control with asymptotically guaranteed stability. *Journal of Adaptive Control and Signal Processing*, 25(4):331–351, 2011.

[39] 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.

[40] J. H. van Schuppen. Stochastic realization problems. In N. Nijmeijer and J. M. Schumacher, editors, *Three Decades of Mathematical System Theory*, Berlin, 1989. Springer-Verlag.

[41] 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.

[42] W. X. Zheng. Transfer function estimation form noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, 12:365–380, 1998.

[43] W. X. Zheng and C. B. Feng. Unbiased parameter estimation of linear systems in presence of input and output noise. *International Journal of Adaptive Control and Signal Processing*, 3:231–251, 1989.

A Proof of (30)

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 (30).