Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 4019–4024

Aspects on errors-in-variables identification: Some ways to mitigate a large bias

Torsten Söderström * **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*

Abstract: Standard identification methods give biased parameter estimates when the recorded signals are corrupted by noise on both input and output sides. When the system is close to be non-identifiable, the bias can be large. The paper discusses the possibilities and potential benefits when using either a reduced model structure or a full errors-in-variables model. The case of using an instrumental variable estimator is also treated.

Keywords: System Identification, Errors-in-variables, Bias

1. INTRODUCTION

The presence of input noise causes biased (non-consistent) parameter estimates when system identification is applied, Söderström and Stoica (1989), Söderström (2007). When the purpose of identification is to gain physical insight rather than to design a model-based controller, large parameter biases should be avoided, Söderström (2018). In Söderström and Soverini (2022b) it was shown that the bias is significant when the system is almost not identifiable.

This paper discusses three options for how one can cope with such cases.

A first possibility is to use an instrumental variable (IV) estimator. The bias will then often be reduced as compared to the use of a prediction error method. A second option is to use a reduced order model structure to circumvent the identifiability issue. The estimates will still be biased though. Another option is to apply a full errors-in-variables (EIV) model structure, Söderström (2018). As will be seen, in this case the bias is strongly reduced.

The paper is organized as follows. The next section gives a general background on the setup and the modeling. Section 3 reviews the bias issues for the case of an almost non-identifiable system, based on Söderström and Soverini (2022b). In that paper the evaluation of the bias was carried out by using the prediction error method. In Section 4 this analysis is extended, and the bias is evaluated when an instrumental variable method is used for identification of the system. Use of a reduced model structure is considered in Section 5, while the use of a full EIV model is treated in Section 6. Some conclusions are given in Section 7.

2. BACKGROUND AND MODELLING

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.

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 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 (3) means that the noise-free input $u_0(t)$ is an ARMA process. The variances λ_u^2, λ_e^2 and λ_v^2 are all assumed to be unknown. These noise assumptions are fairly general in an EIV setting, Söderström (2018).

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.

10.1016/j.ifacol.2023.10.1383

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, and $\varepsilon(t)$ denotes the prediction error. 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.

3. REVIEW OF BIAS DUE TO ALMOST NON-IDENTIFIABILITY

The results in this section are taken from Söderström and Soverini (2022b).

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, Ljung (1987), Söderström and Stoica (1989).

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

$$\hat{\theta} = \arg\min_{\theta} V(\theta) \ , \qquad V(\theta) = \frac{1}{2} E\left\{\varepsilon^2(t,\theta)\right\} \ . \tag{7}$$

In (7) the prediction error $\varepsilon(t,\theta)$ can be found directly from (4), leading to

$$\varepsilon(t,\theta) = H(q)^{-1}\left[y(t) - G(q)u(t)\right] \ . \tag{8}$$

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

$$0 = V'_\theta(\hat{\theta}) \approx V'_\theta(\theta_*) + V''_{\theta\theta}(\theta_*)(\hat{\theta} - \theta_*) \ , \tag{9}$$

leads to

$$\tilde{\theta} \approx -\left[V''_{\theta\theta}(\theta_*)\right]^{-1} V'_\theta(\theta_*) \ . \tag{10}$$

It was shown in Söderström and Soverini (2021), Söderström and Soverini (2022b) that (10) 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, meaning that identifiability is lost if the system parameters are slightly changed. Such a situation can occur 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.

An explicit analysis of (10) for the case of an output error model is given in Söderström and Soverini (2022b). In particular it was shown that when the system is almost overparameterized, the bias can be large.

To be specific, let the system have poles $p_i, i = 1,\dots,n_a$ and zeros $z_j, j = 1,\dots,n_b$. Then set

$$\delta = \min_{i,j} |p_i - z_j| \ , \tag{11}$$

which is a measure of the pole-zero separation. It was shown in Söderström and Soverini (2022b) that for small values of δ (which we refer to as almost non-identifiability), the parameter bias is $O(1/\delta)$.

4. PARAMETER BIAS WHEN USING AN INSTRUMENTAL VARIABLE ESTIMATOR

An alternative to use PEM is to apply an instrumental variable (IV) method, Söderström and Stoica (1989). The effect on the parameter bias for such cases is analyzed in some detail in the report Söderström and Soverini (2022a). Its main findings are summarized here.

The model structure considered is

$$A(q^{-1})y(t) = B(q^{-1})u(t) + \varepsilon(t) \ , \tag{12}$$

and the parameter vector to be estimated is

$$\theta = (\ a_1\ \dots\ a_{n_a}\ b_1\ \dots\ b_{n_b}\ )^T \ . \tag{13}$$

Let θ_0 denote the true value of the parameter vector. Introduce also the following notations for the regressor vector:

$$\begin{aligned}\varphi(t) &= (\ -y(t-1)\ \dots\ -y(t-n_a) \\ &\qquad u(t-1)\ \dots\ u(t-n_b)\ )^T \qquad (14)\\ &= \varphi_0^T(t) + \tilde{\varphi}^T(t) \ , \qquad (15)\\ \varphi_0(t) &= (\ -y_0(t-1)\ \dots\ -y_0(t-n_a) \\ &\qquad u_0(t-1)\ \dots\ u_0(t-n_b)\ )^T \ . \qquad (16)\end{aligned}$$

The model (12) can be written as

$$y(t) = \varphi^T(t)\theta + \varepsilon(t) \ . \tag{17}$$

The general form for the IV estimate is then

$$\hat{\theta} = \left(\hat{R}_{z\varphi}^T \hat{R}_{z\varphi}\right)^{-1} \hat{R}_{z\varphi}^T \hat{r}_{zy} \ , \tag{18}$$

where

$$\hat{R}_{z\varphi} = \frac{1}{N}\sum_{t=1}^{N} z(t)\varphi^T(t), \quad \hat{r}_{zy} = \frac{1}{N}\sum_{t=1}^{N} z(t)y(t) \ . \tag{19}$$

Several assumptions have to be applied in the analysis:

- The unperturbed input signal $u_0(t)$ is persistenly exciting, at least of order $n_a + n_b$. This assumption is needed to guarantee identifiability.

- The polynomials A and B are coprime. This assumption is also needed to guarantee identifiability.
- The three signals $u_0(t)$, $\tilde{u}(t)$ and $\tilde{y}(t)$ are uncorrelated. This is a convenient and mild assumption.
- The elements of the instrumental variable vector $z(t)$ is uncorrelated with the output noise $\tilde{y}(t)$. A natural way to achieve this is to let the elements of $z(t)$ consists of delayed and/or filtered values of the input.
- The asymptotic case with an infinite amount of data is considered, that is $N \to \infty$. Then due to ergodicity the sample estimates can be substituted by expectations, such as $\hat{R}_{z\varphi} \to R_{z\varphi} = E\left\{z(t)\varphi^T(t)\right\}$, etc.

The parameter bias $\tilde{\theta} = \hat{\theta} - \theta_0$ is found to be

$$\begin{aligned}\tilde{\theta} &= R_{z\varphi}^{\dagger} r_{zy} - \theta_0 = R_{z\varphi}^{\dagger}\left[r_{zy} - R_{z\varphi}\theta_0\right] \\ &= R_{z\varphi}^{\dagger} E\left\{z(t)\left[y(t) - \varphi^T(t)\theta_0\right]\right\} \\ &= -R_{z\varphi}^{\dagger} E\left\{z(t)\left[\tilde{\varphi}^T(t)\theta_0\right]\right\} = -R_{z\varphi}^{\dagger} r_{z\tilde{\varphi}\theta_0}\ , \end{aligned} \quad (20)$$

where $R_{z\varphi}^{\dagger}$ is the pseudoinverse of $R_{z\varphi}$.

One can conclude so far that the factor $r_{z\tilde{\varphi}\theta_0}$ is proportional to λ_u^2, and thus that the bias for small values of the input noise variance also is proportional to λ_u^2. Recall that both $z(t)$ and $\tilde{\varphi}(t)$ depends linearly on the input noise $\tilde{u}$. This can be expressed as $\tilde{\theta} = O(\lambda_u^2)$, which is a result of the same sort as when a prediction error method is used, cf. Söderström and Soverini (2021), Söderström and Soverini (2022b).

Apparently the size of $R_{z\varphi}^{\dagger}$ is indeed of importance for the size of the bias $\tilde{\theta}$, cf. (20). It holds

$$R_{z\varphi} = R_{z\varphi_0} + R_{z\tilde{\varphi}}\ . \quad (21)$$

A detailed analysis in Söderström and Soverini (2022a) leads to $\sigma_{\min}\left(R_{z\varphi_0}\right) = O(\delta)$, where $\sigma_{\min}$ denotes the smallest singular value.

Indeed, the term $R_{z\tilde{\varphi}}$ is proportional to λ_u^2. When λ_u^2 is small, the bias $\tilde{\theta}$ may therefore be approximated by $\tilde{\theta}_{\text{app}}$ as

$$\tilde{\theta} = \underbrace{R_{z\varphi_0}^{\dagger} r_{z\tilde{\varphi}\theta_0}}_{\tilde{\theta}_{\text{app}}} + O(\lambda_u^4) \quad (22)$$

The bias $\tilde{\theta}$ and its approximation $\tilde{\theta}_{\text{app}}$ do only partly behave in similar ways. The analysis in Söderström and Soverini (2022a) has established the following:

- For small values of the input noise variance λ_u^2 it holds

$$\tilde{\theta} = O(\lambda_u^2), \quad \tilde{\theta}_{\text{app}} = O(\lambda_u^2) \quad (23)$$

$$\tilde{\theta} = \tilde{\theta}_{\text{app}} + O(\lambda_u^4) \quad (24)$$

- For small values of a pole-zero separation δ it holds

$$\tilde{\theta}_{\text{app}} = O(1/\delta) \quad (25)$$

$$\tilde{\theta} = O(1) \text{ in general} \quad (26)$$

Due to (26) the bias term of an IV estimate is often smaller than that of a prediction error method (PEM) used with an output error model. PEM as well as the approximate bias $\tilde{\theta}_{\text{app}}$ will both be of order $O(1/\delta)$.

5. SMALL POLE-ZERO SEPARATION: USE OF A REDUCED MODEL STRUCTURE

As described above and illustrated in Söderström and Soverini (2022b), when the pole-zero separation δ is small it 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 now examined, by making reference to the PEM estimator recalled in Section 3.

As models of different orders are to be compared a measure of fit has to be selected with care. (Models with different model structures have parameter vectors of different dimensions, and it is generally not meaningful to compare individual parameters.)

To that aim consider the relative error in the transfer function G, taken as

$$\delta G = \frac{\int |\frac{B(e^{i\omega})}{A(e^{i\omega})} - \frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 \mathrm{d}\omega}{\int |\frac{B_0(e^{i\omega})}{A_0(e^{i\omega})}|^2 \mathrm{d}\omega}\ . \quad (27)$$

Numerical example

Consider a simple numerical example with $n_a = 1, n_b = 2$ and let $u_0(t)$ be an AR(1) process,

$$u_0(t) = Fv(t), \quad F = \left(1 - 0.9q^{-1}\right)^{-1}, \quad E\{v^2(t)\} = 1\ . \quad (28)$$

The other parameters in the numerical example are

$$a_1 = -0.8, \quad \lambda_y^2 = 10, \quad b_1 = 2\ . \quad (29)$$

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. This example was also considered in Söderström and Soverini (2022b).

Figure 1 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 (i.e. the filter F in (3) into account.

Figure 2 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.

Finally, Figure 3 shows how the criterion δG varies with the pole-zero separation δ for some different values of the input noise variance λ_u^2.

Some observations

- From Figure 1, 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.

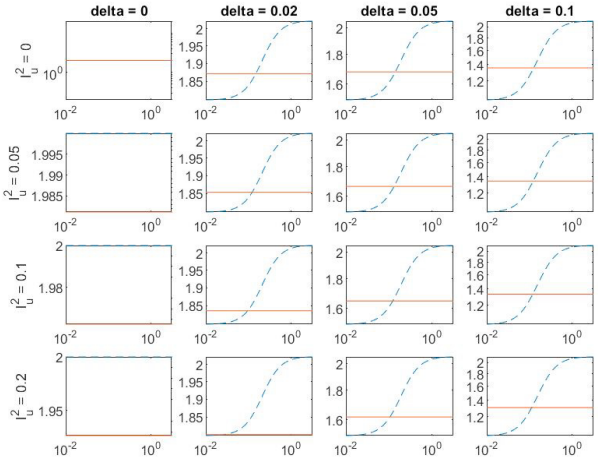

Fig. 1. Frequency plots of the true system (dash-dotted lines) and the approximate models (solid lines) for some different values of λ_u^2 and δ.

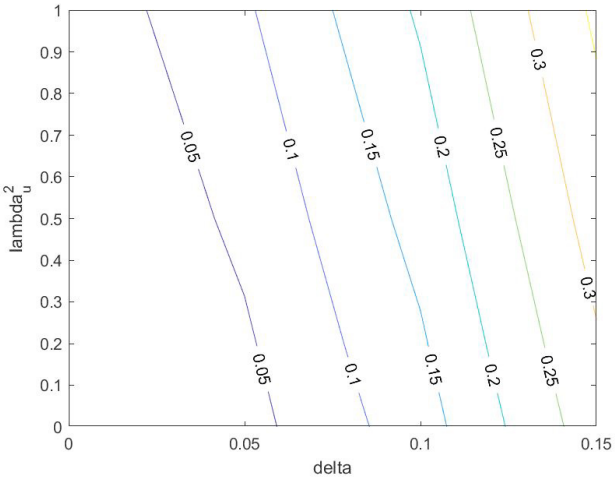

Fig. 2. Contour plots of the criterion δG.

- Figure 3 shows that when the pole-zero separation δ is small, a reduced order model gives a better fit than a full order model. This behaviour is more pronounced when λ_u^2 is not small.

6. SMALL POLE-ZERO SEPARATION: USE OF AN EIV MODEL STRUCTURE

Assume that there is a pole-zero distance δ that is small. It was shown in Söderström and Soverini (2022b) that, when neglecting that there is input noise in the measurement and applying a standard PEM 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. Concerning the covariance matrix a crude approximation

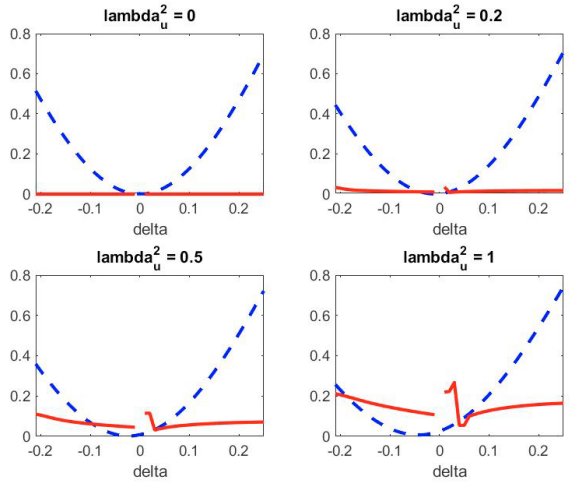

Fig. 3. The criterion δG as a function of δ for some different values of λ_u^2. The dashed lines refer to the deviation between the true system and the reduced order model. The solid lines refer to the deviation between the true system and the full order model. The irregular behaviour of the solid lines for small positive values of δ is due to the existence of local minimum points, cf the discussion in Söderström and Soverini (2021).

would be that it is proportional to $[V_{\theta\theta}''(\theta_*)]^{-1}$ just as in the case of no input noise.

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, using a simple example. The same example as the one in Section 5 is used.

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 based on the theory developed in the cited references, and the influence of the pole-zero separation δ is examined in particular.

The following methods are considered:

(1) Generalized IV estimates (GIVE), see Söderström (2011). This class of identification methods include the Frisch scheme, Guidorzi et al. (2008), Frisch (1934), as well as the bias-eliminating method (BELS), Zheng (1998). An overview of how these seemingly different method can be viewed as alternative algorithms for solving the same set of nonlinear equations, such as (30) below, appear in Söderström et al. (2014a).

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

$$\frac{1}{N}\sum_{t=1}^{N} z(t)y(t) = \frac{1}{N}\sum_{t=1}^{N} z(t)\varphi^T(t)\theta \ , \qquad (30)$$

$$\varphi^T(t) = (\ -y(t-1)\ u(t-1)\ u(t-2)\ )\ , \quad (31)$$

$$z(t) = (\ y(t)\ \dots\ y(t-1-s_y)$$
$$u(t-1)\ \dots\ u(t-2-s_u)\ )^T\ , \quad (32)$$

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
$$s_y = 1, \quad s_u = 0\ . \quad (33)$$
- GIVE2, which corresponds to an overdetermined system of equations. The selected choice corresponds to
$$s_y = 3, \quad s_u = 2\ . \quad (34)$$

(2) Covariance matching method (CM), see Söderström et al. (2009), Söderström et al. (2014b).

The GIVE estimates are obtained as a weighted nonlinear least squares solution to a system of equations, see (30), 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

$$r_y(\tau), \tau = 0, \dots, p_y, \quad r_u(\tau), \tau = 0, \dots, p_u,$$
$$r_{yu}(\tau), \tau = p_1, \dots, p_2\ . \quad (35)$$

CM is applied for two cases:

- CM1, which corresponds to the same covariance elements as GIVE1. This choice corresponds to
$$p_y = 2, p_u = 1, p_1 = -1, p_2 = 2\ . \quad (36)$$
- CM2, which corresponds to the same covariance elements as GIVE2. This choice corresponds to
$$p_y = 4, p_u = 3, p_1 = -3, p_2 = 4\ . \quad (37)$$

(3) ML, the maximum likelihood estimate, Ljung (1987) 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 Söderström (2006).

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 theoretical 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 4 and 5.

Some observations

- The ML estimates are much more accurate than those of the other methods. This is in line with general findings in Söderström (2018). 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.

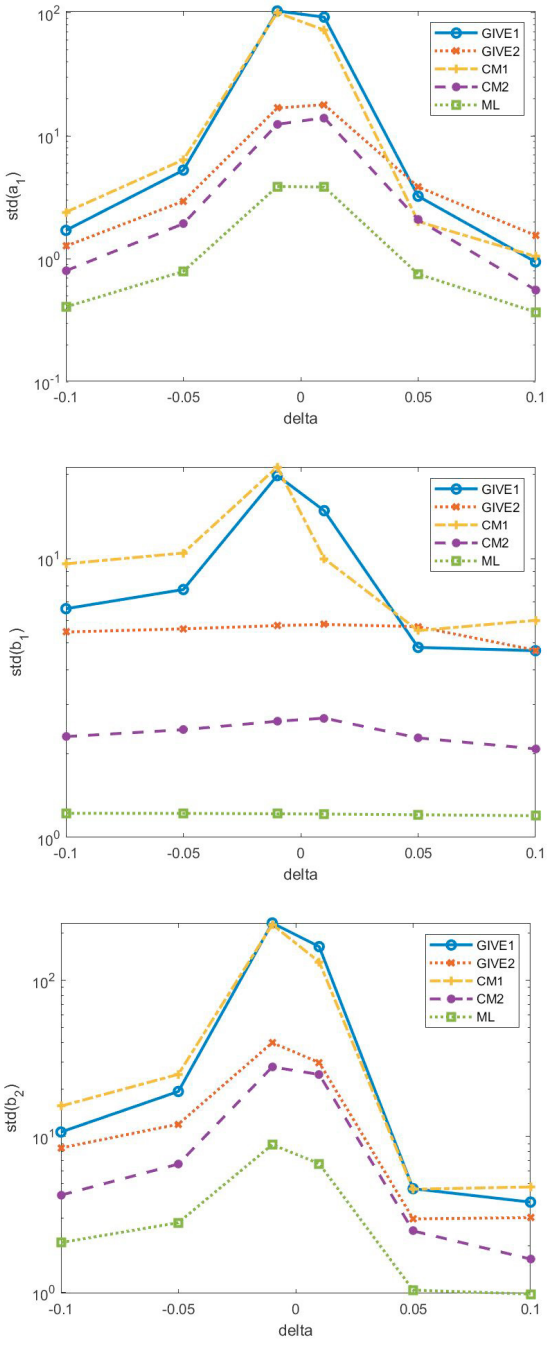

Fig. 4. Standard deviations of the estimated a_1, b_1 and b_2.

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

$$\text{std}(\hat{a}_1) = O(1/\delta), \quad \text{std}(\hat{b}_1) = O(1), \quad \text{std}(\hat{\delta}) = O(1)\ . \quad (38)$$

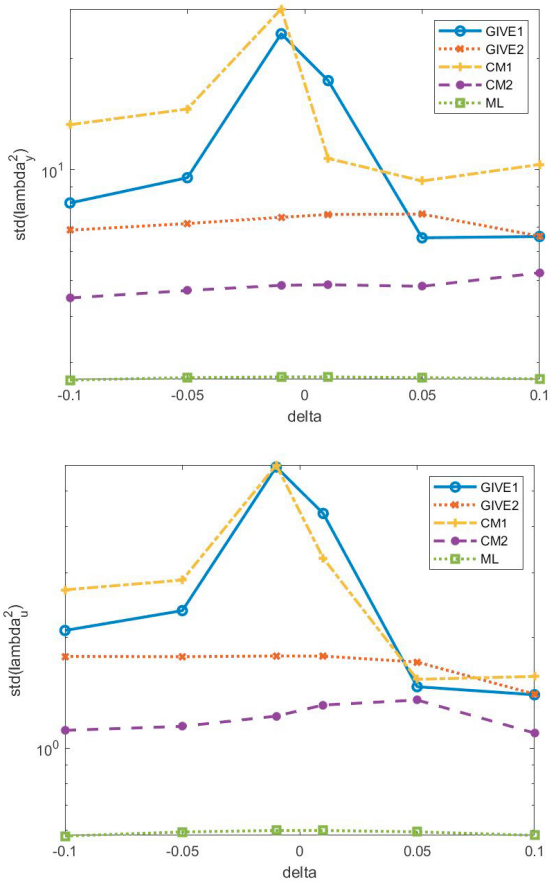

Fig. 5. Standard deviations of the estimated λ_y^2 and λ_u^2.

A much more general study of the accuracy of estimated poles and zeros can be found in Mårtensson and Hjalmarsson (2009).

7. CONCLUSIONS

When a standard system identification method that does not model the input noise is used, biases appear in the parameter estimates. Such biases are more pronounced when the system dynamics has a small pole-zero separation, and the system is then close to being non-identifiable.

When the pole-zero separation is small, the system transfer function can be better estimated by using a reduced order model. Another type of modelling error appears, but the sensitivity due to almost non-identifiability is lost. In total, the modelling error is decreased.

When a full EIV model is used, the modelling error is in many cases smaller than otherwise, also when there is a small pole-zero separation.

REFERENCES

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

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

Ljung, L. (1987). *System Identification - Theory for the User*. Prentice Hall, Englewood Cliffs, NJ, USA.

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

Söderström, T. (2006). 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.

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

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

Söderström, T. (2018). *Errors-in-Variables Methods in System Identification*. Springer-Verlag, London, UK.

Söderström, T., Diversi, R., and Soverini, U. (2014a). 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. Available as http://www.it.uu.se/research/publications/reports/2014-008.

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

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

Söderström, T. and Soverini, U. (2021). When are errors-in-variables aspects particularly important to consider in system identification? Technical Report 2021-006, Department of Information Technology, Uppsala University, Uppsala, Sweden. Available as http://www.it.uu.se/research/publications/reports/2021-006.

Söderström, T. and Soverini, U. (2022a). 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. Available as http://www.it.uu.se/research/publications/reports/2022-008.

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

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall International, Hemel Hempstead, UK.

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