IDENTIFICATION OF A TWO-INPUT SYSTEM: VARIANCE ANALYSIS

M. Gevers [*,1] **L. Mišković** [**,2] **D. Bonvin** [**] **A. Karimi** [**]

* *Center for Systems Engineering and Applied Mechanics*
(CESAME)
Université Catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium
gevers@csam.ucl.ac.be
** *Laboratoire d'Automatique*
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{ljubisa.miskovic,dominique.bonvin,alireza.karimi}@epfl.ch

Abstract: This paper examines the identification of a single-output two-input system. Motivated by an experiment design problem (should one excite the two inputs simultaneously or separately), we examine the effect of the (second) input signal on the variance of the various polynomial coefficients in the case of FIR, ARX, ARMAX, OE and BJ models. A somewhat surprising result is to show that the addition of a second input in an ARMAX model reduces the variance of all polynomials estimates. *Copyright*©*2005 IFAC*

Keywords: Identification, variance analysis, input design, two-input systems

1. INTRODUCTION

Our initial motivation for this piece of research was as follows. Suppose we are to identify a system with one output and two inputs, and we are given a limited amount of experiment time to do so. Is it then better to excite one input at a time, each over half the experiment time say, or should one excite both inputs simultaneously, assuming of course that both inputs are independent?

In the literature, it is argued that in industry the guideline for identification practice is still single-variable based thinking, e.g (Zhu, 1998). However, several authors propose to excite all inputs simultaneously (Zhu and Backx, 1993; Dayal and MacGregor, 1997). Still,

there is no pertinent analysis of the variance to support these statements.

To obtain a reliable answer to the question stated in the first paragraph, we undertook to perform an analysis of the variance of the estimated polynomial coefficients, assuming in all cases that the correct model structure is used. The expressions for the distribution of parameter estimates that are asymptotic in the number of data is a classical result (Ljung, 1999). The theoretical conditions for the validity of this asymptotic theory have recently been treated in (Garatti *et al.*, 2004).

It appeared to us that the answer to the above question might well be different for different model structures such as FIR, ARX, ARMAX, OE, or BJ. The conclusion of our analysis is that, whatever the model structure, the simultaneous application of two independent input signals during the whole experimentation time always leads to a smaller covariance matrix of the estimated polynomial coefficients than the consecutive

application of the two input signals, each during part of the experimentation time.

This result in itself is probably not too surprising, even though a number of system identification experts would probably be inclined to apply the excitation signals during separate experimentation periods in the case of a two-input system. However, our covariance analysis led to some results that run contrary to commonly held beliefs in the system identification community. Let us just foreshadow here two of these findings.

- It is a commonly held view that, in open loop identification, the variance of the noise model is unaffected by the input signal. This view finds its origins in the asymptotic (both in model order and number of data) variance formula for the estimated noise model, $Var(\hat{H}_N(e^{j\omega})) \simeq \frac{n}{N} \frac{\Phi_v(e^{j\omega})}{|H_0(e^{j\omega})|^2}$ in which the input signal plays no role whatsoever (Ljung, 1999). Recently, exact variance expressions that are not asymptotic in model order have been derived in (Xie and Ljung, 2001) and (Ninness and Hjalmarsson, 2004). Ninness and Hjalmarsson have considered the BJ and OE models and found that, in these cases, the variance of the noise model is independent of the input signal. However, no result is available so far for the ARX and ARMAX model, which are treated in this paper. We show in particular that, for the ARMAX model $A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})e(t)$ (with z^{-1} being the delay operator), the input signal $u(t)$ does in fact contribute to a reduction of the variance of $\hat{C}_N$, as well as that of $\hat{A}_N$, of course.
- It follows from the first result that the addition of a second input signal, say $u_2(t)$, always reduces the variance of the estimated noise model $\frac{\hat{C}_N(e^{j\omega})}{\hat{A}_N(e^{j\omega})}$ of an ARMAX model. In addition, we show that, for ARX and ARMAX models, the addition of $u_2(t)$ also reduces the variance of the estimated coefficients of the polynomial $B_1(z^{-1})$ related to the other input.

Of course, our results do not contradict any established result in the identification literature. It so happens that, to the best of our knowledge, very little attention has been paid to the variance analysis of the parameters in recent years, most probably because of the widespread use of and reliance on the asymptotic (in model order) variance formulas for the transfer function estimates quoted above. These results, which are based on an assumption that the model order goes to infinity, have contributed to the widely held but incorrect view that the input signal has no bearing on the precision of the noise model in open loop identification.

Our results do not only provide a clear answer to the question of whether or not one should excite both inputs of a two-input system simultaneously. They also suggest that, when that is at all possible, it pays to add additional input energy into the system for the sake of identifying model parameters whenever the input-output and noise models have common parameters. For example, if the physics of the problem allow one to add a second input signal somewhere in a single-input single-output system, so as to convert the system $Ay = Bu + Ce$ into the system $Ay = B_1u_1 + B_2u_2 + Ce$ say, then the addition of the signal u_2 improves the precision of all other estimated polynomials A, B_1 and C. Such thought would certainly not have occurred to the authors prior to this piece of research.

The paper is organized as follows. Preliminary material is given in Section 2. Section 3 treats the linear regression FIR and ARX models. The ARMAX model is considered in Section 4 where it is shown that the addition of a second input reduces the variance of all estimated coefficients. Section 5 extends the conclusion of the variance analysis to other model structures and brings some final remarks.

2. PRELIMINARIES ON PREDICTION ERROR IDENTIFICATION

We consider the identification of an unknown linear time-invariant "true" system with one output and possibly two inputs (hereafter denoted a DISO system, for double input single output):

$$\mathcal{S} : y(t) = G_1(z)u_1(t) + G_2(z)u_2(t) + H(z)e(t), \quad (1)$$

where $G_1(z)$ and $G_2(z)$ are strictly causal rational transfer functions, y is the measured output, u_1 and u_2 are two input signals, and where the noise, assumed to be quasistationary, is modelled as the output of a causal and causally invertible noise model $H(z)$ driven by a white noise input e with variance σ_e^2. A special case of (1) is a single input single output (SISO) system, when $u_2 = 0$. We consider the identification of a model for $\mathcal{S}$ using a parametrized model set:

$$\mathcal{M} = \{G_1(z,\theta), G_2(z,\theta), H(z,\theta), \theta \in D_\theta \subset \mathbf{R}^k\} \quad (2)$$

where $G_1(z,\theta)$, $G_2(z,\theta)$ and $H(z,\theta)$ are rational transfer functions, $\theta \in \mathbf{R}^k$ is a parameter vector, and D_θ is a subset of admissible values for θ. The analysis of this paper will focus entirely on variance errors; thus we assume from now on that $\mathcal{S} \in \mathcal{M}$. To every θ corresponds a one-step ahead predictor $\hat{y}(t|t-1,\theta)$, and hence a one-step ahead prediction error:

$$\varepsilon(t,\theta) \triangleq y(t) - \hat{y}(t|t-1,\theta) \quad (3)$$
$$= H^{-1}(z,\theta)[y(t) - G_1(z,\theta)u_1(t)$$
$$- G_2(z,\theta)u_2(t)]$$

The least squares PE estimate $\hat{\theta}_N$ based on N input-output data is defined as

$$\hat{\theta}_N = \arg \min_{\theta \in D_\theta} V_N(\theta), \quad \text{with} \quad (4)$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[\varepsilon(t,\theta)]^2. \qquad (5)$$

The estimate $\hat{\theta}_N$ defines the model $\hat{G}_1 = G_1(z,\hat{\theta}_N)$, $\hat{G}_2 = G_2(z,\hat{\theta}_N)$, $\hat{H} = H(z,\hat{\theta}_N)$. Since $\mathcal{S} \in \mathcal{M}$, it follows that under reasonable conditions $\hat{\theta}_N \xrightarrow{N\to\infty} \theta_0$, where $\theta_0 \in D_\theta$ is such that $G_1(z,\theta_0) = G_1(z)$, $G_2(z,\theta_0) = G_2(z)$, and $H(z,\theta_0) = H(z)$. Thus the transfer function estimates converge to the true transfer functions. The parameter error then converges to a Gaussian random variable (Ljung, 1999):

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_\theta), \quad \text{as } N \to \infty \qquad (6)$$

where

$$P_\theta = \sigma_e^2[E\psi(t,\theta_0)\psi^T(t,\theta_0)]^{-1} \triangleq \sigma_e^2 M^{-1}, \qquad (7)$$

with an obvious definition for the information matrix $M \in \mathbf{R}^{k\times k}$. Here $\psi(t,\theta)$ is the gradient of the prediction error with respect to the parameters (i.e. the sensitivity to these errors):

$$\psi(t,\theta) \triangleq \frac{\partial\varepsilon(t,\theta)}{\partial\theta} = -\frac{\partial\hat{y}(t|t-1,\theta)}{\partial\theta} \qquad (8)$$

It follows from the formulas (8), (3) and (1) that this gradient ψ can be rewritten as follows:

$$\psi(t,\theta) = \frac{-1}{H(z,\theta)}\left[\frac{\partial G_1(z,\theta)}{\partial\theta}u_1(t)+\right.$$
$$\left.\frac{\partial G_2(z,\theta)}{\partial\theta}u_2(t) + \frac{\partial H(z,\theta)}{\partial\theta}\varepsilon(t,\theta)\right].(9)$$

Combining (7) and (9), and using Parseval's theorem, we can then write the following expression for the information matrix M:

$$M = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{d\omega}{|H|^2}\left\{\left(\frac{\partial G_1}{\partial\theta}\right)\left(\frac{\partial G_1}{\partial\theta}\right)^*\Phi_{u_1} \qquad (10)\right.$$
$$\left.+ \left(\frac{\partial G_2}{\partial\theta}\right)\left(\frac{\partial G_2}{\partial\theta}\right)^*\Phi_{u_2} + \left(\frac{\partial H}{\partial\theta}\right)\left(\frac{\partial H}{\partial\theta}\right)^*\sigma_e^2\right\}$$

For a large enough number N of data, the asymptotic covariance formulas (6)-(7) are typically used to compute approximate expressions for the covariance of the parameter vector estimate:

$$cov(\hat{\theta}_N) \approx \frac{1}{N}P_\theta = \frac{\sigma_e^2}{N}M^{-1}. \qquad (11)$$

The expression (10) will be our key tool for the analysis of the effect of one input signal, say u_2, on the covariance of the coefficients of the various model polynomials. By specializing this formula to any of the classically used polynomial model structures (FIR, ARX, OE, BJ, ARMAX), one can examine whether or not the input u_2 has a variance reduction effect on any or all of the estimated coefficients of the polynomials A, B_1 and C in an ARMAX model structure, for example. We mention that (Klein and Mélard, 1994) have computed the gradients appearing in (10) for all commonly used polynomial models.

For model structures that are linear in the parameters, such as FIR or ARX, the one-step ahead predictor can be written as:

$$\hat{y}(t|t-1,\theta) = \varphi^T(t)\theta, \qquad (12)$$

where $\varphi(t) \in \mathbf{R}^k$ is a vector containing known input and output signals only, i.e. independent of θ. In such cases, the matrix $M = E\varphi(t)\varphi^T(t)$ is very easy to compute, and there is no need to resort to the frequency domain integral expression (10).

In the sequel, we use the formula (10), or its simplified version $M = E\varphi(t)\varphi^T(t)$ whenever appropriate, to study the effect of a particular input on the variance of the parameter estimates, for the classical polynomial models mentioned above. Given the space limitation in this conference paper, we present a full analysis only for the linear in the parameter FIR and ARX models, and the ARMAX model. The analysis methods for the other cases can be easily derived from the methods presented here for the FIR, ARX and ARMAX models. We present in section 5 the conclusions of our analysis for all model structures.

3. THE ARX AND FIR MODELS

In this section we consider the special case of ARX (Auto-Regressive with eXogeneous inputs) and FIR (Finite Impulse Response) model structures.

A. ARX models with two inputs

Consider an ARX model structure with two independent inputs:

$$A(z^{-1})y(t) = B_1(z^{-1})u_1(t) + B_2(z^{-1})u_2(t) + e(t) \qquad (13)$$

where $A(z^{-1}) = 1 + \sum_{j=1}^{n}a_j z^{-j}$, $B_1(z^{-1}) = \sum_{j=1}^{n}b_1^j z^{-j}$ and $B_2(z^{-1}) = \sum_{j=1}^{n}b_2^j z^{-j}$ are polynomials in the delay operator z^{-1}. For simplicity, we choose all polynomials in our models to have the same degree. The system can also be rewritten in linear regression form:

$$y(t) = \varphi^T(t)\theta_0 + e(t), \quad \text{with} \qquad (14)$$

$$\theta^T = [a_1 \ldots a_n; b_1^1 \ldots b_1^n; b_2^1 \ldots b_2^n]$$
$$\triangleq [\ \mathbf{a}^T \ ; \ \mathbf{b_1}^T \ ; \ \mathbf{b_2}^T]$$
$$\varphi^T(t) = [-y(t-1)\ldots -y(t-n);$$
$$u_1(t-1)\ldots u_1(t-n); u_2(t-1)\ldots u_2(t-n)]$$
$$\triangleq [Y^T(t-1) ; U_1^T(t-1) ; U_2^T(t-1)],$$

with the obvious definitions for the n-dimensional vectors $\mathbf{a}, \mathbf{b_1}, \mathbf{b_2}, Y(t-1), U_1(t-1)$ and $U_2(t-1)$. For an ARX model, the Least Squares parameter estimate $\hat{\theta}_N$ of θ based on N input-output data is unbiased. Its covariance, for large N, can be approximated by (11) where M is now given by

$$M = \begin{pmatrix} R_{YY} & R_{YU_1} & R_{YU_2} \\ R_{U_1Y} & R_{U_1U_1} & 0 \\ R_{U_2Y} & 0 & R_{U_2U_2} \end{pmatrix} \qquad (15)$$

with $R_{YY} \triangleq E\{Y(t-1)Y^T(t-1)\}$, $R_{YU_1} \triangleq E\{Y(t-1)U_1^T(t-1)\}$, etc. The zero blocks in M follow from the independence between the two input signals. In order to apprehend the effect of each of the two input signals on the accuracy of the parameter estimates of the polynomials A, B_1 and B_2, we need to compute, blockwise, the inverse of the information matrix M. We denote:

$$C \triangleq M^{-1} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \qquad (16)$$

Remember that $P_\theta = \sigma_e^2 M^{-1} = \sigma_e^2 C$. We now compare the covariances of the estimates of the parameter vectors $\mathbf{a}, \mathbf{b_1}$ and $\mathbf{b_2}$, based on N input-output data, for the case of one input only ($u_2 = 0$) and the case where two independent input signals, u_1 and u_2, are applied simultaneously. We have $cov(\hat{\mathbf{a}}) \approx \frac{\sigma_e^2}{N} C_{11}$, $cov(\hat{\mathbf{b_1}}) \approx \frac{\sigma_e^2}{N} C_{22}$, and $cov(\hat{\mathbf{b_2}}) \approx \frac{\sigma_e^2}{N} C_{33}$. We shall denote by $C_{ii}^{(1)}, i = 1, 2$ the block-diagonal elements of the matrix C obtained when only input u_1 is applied, and by $C_{ii}^{(2)}, i = 1, 2, 3$ the corresponding elements when the two inputs are applied simultaneously.

When only input u_1 is present, we have:

$$C_{11}^{(1)} = (R_{YY} - R_{YU_1} R_{U_1U_1}^{-1} R_{U_1Y})^{-1}$$
$$C_{22}^{(1)} = (R_{U_1U_1} - R_{U_1Y} R_{YY}^{-1} R_{YU_1})^{-1} \qquad (17)$$

When the two inputs, u_1 and u_2, are present, we have:

$$C_{11}^{(2)} = (R_{YY} - R_{YU_1} R_{U_1U_1}^{-1} R_{U_1Y} - R_{YU_2} R_{U_2U_2}^{-1} R_{U_2Y})^{-1}$$
$$C_{22}^{(2)} = (R_{U_1U_1} \qquad\qquad (18)$$
$$- R_{U_1Y}(R_{YY} - R_{YU_2} R_{U_2U_2}^{-1} R_{U_2Y})^{-1} R_{YU_1})^{-1}$$

We can now establish the following result.

Theorem 1. Consider the identification of the parameter vectors $\mathbf{a}$ and $\mathbf{b_1}$ of the ARX model (13), where the inputs u_1 and u_2 are independent. Then the covariance matrices of the parameter estimates $\hat{\mathbf{a}}$ and $\hat{\mathbf{b_1}}$ are smaller if $u_2 \neq 0$ than if $u_2 = 0$, i.e.

$$C_{11}^{(2)} \leq C_{11}^{(1)} \quad \text{and} \quad C_{22}^{(2)} \leq C_{22}^{(1)}.$$

Proof. We first compare the covariances of the parameter estimates of the A polynomial. In comparing $C_{11}^{(1)}$ and $C_{11}^{(2)}$, we must keep in mind that the covariances R_{YY} are different in these two situations. To take this into account, we explicitly split up the covariance matrix R_{YY} into the sum of its contributions due to u_1, u_2 and e, respectively (see (10)):

$$R_{YY} = R_{YY}(u_1) + R_{YY}(u_2) + R_{YY}(e).$$

Comparing (18) with (17), we then observe that:

$$(C_{11}^{(2)})^{-1} - (C_{11}^{(1)})^{-1} = R_{YY}(u_2) - R_{YU_2} R_{U_2U_2}^{-1} R_{U_2Y}$$

The latter quantity is always nonnegative definite, because the matrix

$$\begin{pmatrix} R_{YY}(u_2) & R_{YU_2} \\ R_{U_2Y} & R_{U_2U_2} \end{pmatrix}$$

is a covariance matrix. Hence $C_{11}^{(2)} \leq C_{11}^{(1)}$. We now turn to the covariance of the parameter estimates of the B_1 polynomial. Comparing $C_{22}^{(1)}$ and $C_{22}^{(2)}$, we observe that they differ by the inverse that appears in the middle of their second term. In the case where $u_2 = 0$, the term to be inverted is $R_{YY}(u_1) + R_{YY}(e)$. In the case where $u_2 \neq 0$, that term becomes $R_{YY}(u_1) + R_{YY}(e) + R_{YY}(u_2) - R_{YU_2} R_{U_2U_2}^{-1} R_{U_2Y}$. For the same reason as above, the latter term is always larger than the former, and hence $C_{22}^{(2)} \leq C_{22}^{(1)}$. $\qquad\square$

Analysis of the ARX case

(1) Addition of a second input signal in an ARX system improves the accuracy of all parameter estimates. While the improved accuracy of the estimate of the A polynomial is easily understood by the increased energy in the $y(t)$ data resulting from the additional input signal, the improvement of the estimate of the polynomial B_1 most probably comes as a surprise to most readers. In fact, the improved accuracy of the B_1 polynomial estimate results from the improved accuracy of the A polynomial estimate and the correlation between the estimates of the A and B_1 polynomial coefficients.

(2) A practical consequence of the last observation is as follows. There are situations where the aim is to identify a SISO ARX model with constraints on the input u, but where it is physically possible to add an additional excitation u_2 going through the same pole dynamics $A(z^{-1})$. Our result suggests that it then pays off to add such a second excitation, because it improves the accuracy of the estimates of the polynomials A and B.

Simulation example

We consider the DISO ARX model structure (13) with $A = 1 - 0.2z^{-1}$, $B_1 = 10z^{-1} + z^{-2}$, $B_2 = 0.1z^{-1} + 4z^{-2}$, with $u_1, u_2,$ and e being three independent white noises, with standard deviations $\sigma_{u_1} = 2, \sigma_{u_2} = 10$, and $\sigma_e = 4$, respectively. We have compared the situation where only input u_1 is used with the situation where the two inputs are used simultaneously. For each case, we have computed the theoretical covariance matrix, and we have also performed a Monte-Carlo simulation to confirm these theoretical results.

Case A: 1 input

The theoretically computed asymptotic covariance matrix P_θ for $\theta = (a_1, b_1^1, b_1^2)$ is:

$$P_\theta = \begin{pmatrix} 0.2954 & 0 & -2.9538 \\ 0 & 4.0000 & 0 \\ -2.9538 & 0 & 33.5385 \end{pmatrix}$$

The variances of the 3 parameter estimates computed by 1000 Monte-Carlo runs are:

0.2849 4.0171 32.7052

Case B: 2 inputs
We now obtain the following theoretical covariance matrix for the vector $\theta = (a_1, b_1^1, b_1^2, b_2^1, b_2^2)$:

$$
P_\theta = \begin{pmatrix}
0.0092 & 0 & -0.0921 & 0 & -0.0009 \\
0 & 4.0000 & 0 & 0 & 0 \\
-0.0921 & 0 & 4.9208 & 0 & 0.0092 \\
0 & 0 & 0 & 0.1600 & 0 \\
-0.0009 & 0 & 0.0092 & 0 & 0.1601
\end{pmatrix}
$$

The variances obtained by Monte-Carlo runs are:

0.0088 4.0185 5.1151 0.1616 0.1580

Comparing the two situations, we observe that the presence of the second input, u_2, not only reduces the variance of a_1 drastically, but it also significantly improves the precision of the estimated coefficient b_2^1. Note that it has no effect on b_1^1 because of sampling delay. The Monte-Carlo simulations confirm the theoretical calculations.

B. FIR models with two inputs

We now consider that the true system is described by the following FIR model structure:
$$y(t) = B_1(z^{-1})u_1(t) + B_2(z^{-1})u_2(t) + e(t). \quad (19)$$

It is a special case of the ARX model in which $A(z^{-1}) = 1$. The parameter vector θ thus specializes to $\theta^T = [b_1^1 \dots b_1^n; b_2^1 \dots b_2^n]$. The covariance analysis for the FIR model structure therefore results from that of the ARX model structure. Thus, the Least Squares parameter estimate $\hat{\theta}_N$ of θ based on N input-output data is unbiased, and its covariance, for large N, is given by (11) where

$$
P_\theta = \sigma_e^2 \begin{pmatrix} R_{U_1 U_1} & 0 \\ 0 & R_{U_2 U_2} \end{pmatrix}^{-1} \quad (20)
$$

In the special case where the inputs are white noises with variance $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$, respectively, we get the following approximate formula:

$$
cov(\hat{\theta}_N) \approx \frac{1}{N} \begin{pmatrix} \frac{\sigma_e^2}{\sigma_{u_1}^2} I_n & 0 \\ 0 & \frac{\sigma_e^2}{\sigma_{u_2}^2} I_n \end{pmatrix} \quad (21)
$$

Analysis of the FIR case

(1) The asymptotic accuracy of the estimates of the b_1^j coefficients is totally independent of the presence of u_2 or of its power. Thus, even if in the output signal y the signal u_2 accounts for 90%, say, of the total power, this has no effect on the quality of the estimates of the B_1 coefficients.

(2) The accuracy of the estimate of each coefficient b_1^j is identical; it depends only on the signal to noise ratio $\frac{\sigma_{u_1}^2}{\sigma_e^2}$.

(3) Given a limited experimentation time of length N, it is best to excite both inputs simultaneously.

4. THE ARMAX MODEL

In the case of an ARMAX system, the variance analysis is not quite as simple, because the predictor is no longer a linear regression of measured signals. Thus, we need to appeal to the frequency domain expression (10), in which the general expressions of the sensitivities are specialized to the ARMAX case. As we shall see, the effect of adding a second input signal in an ARMAX model structure follows immediately from the covariance analysis of a SISO ARMAX model structure. Thus, we consider the ARMAX structure:

$$A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})e(t) \quad (22)$$

with A and B as before and $C(z^{-1}) = 1 + \sum_{j=1}^n c_j z^{-j}$. The parameter vector is defined as:

$$\theta^T \triangleq [\ \mathbf{a}^T \ ; \ \mathbf{b}^T \ ; \ \mathbf{c}^T].$$

For this ARMAX model, the predictor becomes:

$$\hat{y}(t|t-1,\theta) = \frac{B}{C}u(t) + (1 - \frac{A}{C})y(t)$$

Thus, the two terms in the expression (10) of the information matrix specialize to:

$$
\frac{1}{H}\frac{\partial G}{\partial \theta} = \frac{1}{C}\begin{pmatrix} -\frac{B}{A}e^{-j\omega} \\ \vdots \\ -\frac{B}{A}e^{-j\omega n} \\ \hline e^{-j\omega} \\ \vdots \\ e^{-j\omega n} \\ \hline 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \frac{1}{H}\frac{\partial H}{\partial \theta} = \frac{1}{C}\begin{pmatrix} -\frac{C}{A}e^{-j\omega} \\ \vdots \\ -\frac{C}{A}e^{-j\omega n} \\ \hline 0 \\ \vdots \\ 0 \\ \hline e^{-j\omega} \\ \vdots \\ e^{-j\omega n} \end{pmatrix}
$$

With these expressions under our belt, we observe that the information matrix M for an ARMAX model takes the following block-diagonal structure:

$$
M = \begin{pmatrix} M_{11}(u) + M_{11}(e) & M_{12}(u) & M_{13}(e) \\ M_{21}(u) & M_{22}(u) & 0 \\ M_{31}(e) & 0 & M_{33}(e) \end{pmatrix} \quad (23)
$$

where, e.g., $M_{11}(u)$ denotes the contribution of the $(1,1)$-term of the information matrix that is due to the signal u. By computing the inverse $C = M^{-1}$ of (23) we can then study the effect of the input signal u on the precision of each of the components $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ of the parameter estimate $\hat{\theta}$. We have the following result.

Theorem 2. Consider the identification of the parameters of the ARMAX model (22). The covariance of the parameter estimates $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ are all influenced by the input signal. In particular, the covariance matrix of $\hat{\mathbf{c}}$ has the following expression:

$$
P_{\mathbf{c}} = \sigma_e^2\{M_{33}(e) - M_{31}(e)[M_{11}(u) + M_{11}(e) \\
- M_{12}(u)M_{22}^{-1}(u)M_{21}(u)]^{-1}M_{13}(e)\}^{-1} \quad (24)
$$

For the corresponding ARMA model structure (i.e. no input) with the same A and C polynomial, the

covariance of the estimate of the C polynomial is larger than in the ARMAX model structure. We get:

$$P_{\mathbf{c}} = \sigma_e^2 \{M_{33}(e) - M_{31}(e)M_{11}^{-1}(e)M_{13}(e)\}^{-1} \quad (25)$$

Proof. Space limitations prevent us from giving a complete proof, but it is very straightforward. A comparison of the expressions (24) and (25) clearly shows that $P_{\mathbf{c}}(ARMA) \geq P_{\mathbf{c}}(ARMAX)$. $\qquad\square$

Analysis of the ARMAX case
  (1) It follows from the result above that the input signal u has a variance reduction effect on the estimate of the C polynomial.
  (2) It follows from the analysis of the ARX and ARMAX cases that addition of a second input in an ARMAX model structure reduces the covariance of the estimates of all polynomials.

Simulation example
We have simulated the ARMAX model (22) with $A = 1 - 0.2z^{-1}$, $B = 10z^{-1} + z^{-2}$ and $C = 1 - 1.6z^{-1} + 0.64z^{-2}$, and with u and e two independent white noises with $\sigma_u = 1$ and $\sigma_e = 4$, respectively. We have compared, both theoretically and by Monte-Carlo simulations with 1000 runs, the covariances obtained for the parameter estimates for this ARMAX model with those obtained with the ARMA model with same A and C polynomial, i.e. with $u = 0$ in the above model.

ARMA model
The theoretically computed asymptotic covariance matrix P_θ with $\theta = (a_1, c_1, c_2)$ is:

$$P_\theta = \begin{pmatrix} 3.6879 & 2.4837 & -2.3934 \\ 2.4837 & 2.2631 & -2.1879 \\ -2.3934 & -2.1879 & 2.1437 \end{pmatrix}$$

The variances obtained by Monte-Carlo runs are:

$$3.9433 \quad 2.4015 \quad 2.2896$$

ARMAX model
The theoretical covariance matrix for the vector $\theta = (a_1, b_1, b_2, c_1, c_2)$ is: $P_\theta =$

$$\begin{pmatrix} 0.6906 & -1.8792 & 11.2279 & 0.4651 & -0.4482 \\ -1.8792 & 14.5597 & -39.7682 & -1.2656 & 1.2195 \\ 11.2279 & -39.7682 & 191.9959 & 7.5617 & -7.2867 \\ 0.4651 & -1.2656 & 7.5617 & 0.9036 & -0.8778 \\ -0.4482 & 1.2195 & -7.2867 & -0.8778 & 0.8813 \end{pmatrix}$$

The variances of the 5 parameter estimates obtained with 1000 Monte-Carlo runs are:

$$0.6805 \quad 15.8600 \quad 194.6594 \quad 0.9922 \quad 0.9735$$

The Monte-Carlo simulations confirm the theoretical computations. More importantly, we observe that the presence of an input signal in the ARMAX model, in comparison with the ARMA model, not only reduces the variance of the a parameter (by a factor of 5), but perhaps more surprisingly, it also reduces the variance of the c_j parameter estimates (by a factor of 2.5).

5. CONCLUSIONS

We have performed the same variance analysis for the other standard polynomial models, namely the Output-Error (OE) and Box-Jenkins (BJ) models. It follows from our analysis that the following overall conclusions can be drawn.

  • In the case of DISO systems, it is always best to excite the two inputs simultaneously.
  • For ARX and ARMAX structures, that have common parameters in the input-output and noise dynamics, the addition of a second input u_2 improves the accuracy of all parameter estimates, including those of the B_1 polynomial corresponding to u_1, and of the C polynomial of the noise model. Thus, if you can add a second input signal, do so.
  • For FIR, OE, or BJ structures, that have no common parameters between the different transfer functions, the input signal u_2 does not affect the variances of the parameters corresponding to the first input or to the noise model.

Finally, it is worthwhile to extend the results presented here to more general MISO (multi-input single-output) structures.

REFERENCES

Dayal, B. S. and J. F. MacGregor (1997). Multi-output process identification. *Journal of Process Control* **7**(4), 269–282.

Garatti, S., M. C. Campi and S. Bittanti (2004). Assessing the quality of identified models through the asymptotic theory — when is the result reliable?. *Automatica* **40**, 1319–1332.

Klein, A.A.B. and G. Mélard (1994). The information matrix of multiple-input single-output time series models. *Journal of Computational and Applied Mathematics* **51**, 349–356.

Ljung, L. (1999). *System Identification: Theory for the User, 2nd Edition*. Prentice-Hall. Englewood Cliffs, NJ.

Ninness, B. and H. Hjalmarsson (2004). Variance error quantifications that are exact for finite-model order. *IEEE Trans. Automatic Control* **49**(8), 1275–1291.

Xie, L. L. and L. Ljung (2001). Asymptotic variance expressions for estimated frequency functions. *IEEE Trans. Automatic Control* **46**(12), 1887–1889.

Zhu, Y. C. (1998). Multivariable process identification for MPC: the asymptotic method and its applications. *Journal of Process Control* **8**(2), 101–115.

Zhu, Y. C. and T. Backx (1993). Identification of multivariable industrial processes for simulation, diagnosis and control. In: *Advances in Industrial Control* (M. J. Grimble and M. A. Johnson, Eds.). Springer-Verlag. London.