PROBABILISTIC MODEL UNCERTAINTY BOUNDING: AN APPROACH WITH FINITE-TIME PERSPECTIVES

Sippe G. Douma [1] and **Paul M.J. Van den Hof** [2]

Delft Center for Systems and Control,
Delft University of Technology,
Mekelweg 2, 2628 CD Delft, The Netherlands.
Tel. +31-15-2784509; Fax: +31-15-2786679; E-mail:
p.m.j.vandenhof@dcsc.tudelft.nl.

Abstract: In prediction error identification model uncertainty bounds are generally derived from the statistical properties of the parameter estimator, i.e. asymptotic normal distribution of the estimator, and availability of the covariance information. When the primal interest of the identification is in a-posteriori quantifying the uncertainty in an estimated parameter, alternative parameter confidence bounds can be constructed. Probabilistic parameter confidence bounds are studied for ARX models which are generated by computationally more simple expressions, and which have the potential of being less dependent on asymptotic approximations and assumptions. It is illustrated that the alternative bounds can be powerful for quantifying parameter confidence regions for finite-time situations.

Keywords: model validation, system identification, model uncertainty, parameter confidence bounds.

1. INTRODUCTION

In the mainstream approach of system identification, i.e. prediction error identification, model uncertainty quantification is based on the covariance matrix of the parameter estimator, in conjunction with a presumed (or asymptotically achieved if the number of data tends to infinity) Gaussian probability density function, see e.g., (Ljung, 1999*b*). This description leads to probabilistic confidence bounds on estimated parameters, from which also probabilistic confidence bounds on estimated frequency responses can be constructed, with any pre-chosen level of probability.

In classical prediction error identification explicit and exact expressions for the parameter covariance matrix are available for model structures that are linear-in-the-parameters in the situation that the model structures are correct, i.e. the data generating system is part of the model set, $\mathcal{S} \in \mathcal{M}$. For linear regression models with deterministic regressors (such as FIR and generalized FIR (Heuberger *et al.*, 2005)) this holds for finite data length; for ARX models this holds asymptotically. For general model structures, and under the assumption $\mathcal{S} \in \mathcal{M}$, the parameter covariance matrix can be approximated by using first order Taylor expansions. However, in this situation exact system knowledge is also required to compute these approximate expressions for the covariance matrix.

[1] Now with Shell International Exploration and Production, Rijswijk, The Netherlands.
[2] Corresponding author.

Only in case of linear parametrizations results are available for model uncertainty bounding when the model structures are not correct ($\mathcal{S} \notin \mathcal{M}$), see e.g. Hakvoort and Van den Hof (1997), Ljung (1999*a*) and Heuberger *et al.* (2005), Chapter 7.
For Gaussian distributions parameter confidence bounds that are constructed on the basis of the (exact) covariance matrix of the parameter estimator, lead to the smallest possible parameter uncertainty regions for a given probability level. However, usually the exact covariance matrix is not available, and a replacement has to be made with an estimated covariance.
In this paper, some alternatives are studied, where the aim is to specify parameter uncertainty regions that do not (or at least not as much) rely on asymptotic assumptions but for which exact probabilistic expressions can be made. It will be shown that the quantification of parameter uncertainty on the basis of only *one experiment* can be done without the full analysis of the parameter estimator. This will be shown to facilitate uncertainty bounding in several ways, as well as give rise to results that show potentials for application in finite-time analysis. Finite-time analysis of estimated parameters is an important problem, however with few results so far. For some results see e.g. Campi and Weyer (2002) and Weyer and Campi (2002).
After presenting the principle concepts of the paradigm that was introduced in Douma and Van den Hof (2005), parameter uncertainty regions are derived for ARX models, for the situation that $\mathcal{S} \in \mathcal{M}$. The presented approach can also be applied to nonlinearly parametrized models, as well as to the situation $\mathcal{S} \notin \mathcal{M}$ along the lines as presented in Douma and Van den Hof (2005). Due to space limitations this is reported elsewhere, see e.g. (Douma, 2006).

2. ESTIMATOR PROPERTIES AND UNCERTAINTY REGIONS

It is standard practice to base the characterization of the quality of a parameter estimate $\hat{\theta}$ on the (statistical) properties of the estimator, where the estimator is defined as a mapping:

$$\boldsymbol{\theta} = g(\mathbf{z})$$

where $\mathbf{z}$ indicates the measurements. Boldface symbols are generally used to distinguish random variables from realizations thereof. The classical way of arriving at a model uncertainty bound is:

- Assume knowledge of the pdf $p_{\boldsymbol{\theta}}(\theta)$ of $\boldsymbol{\theta}$, given by prior information and/or by application of the Central Limit Theorem;
- Assume that the estimator is unbiased, i.e. $\theta_0 = \mathbb{E}\boldsymbol{\theta}$;
- Then knowing that every estimate $\hat{\theta}$ is a realization of the random variable with pdf $p_{\boldsymbol{\theta}}(\theta)$, this pdf with expected value θ_0 can be used to statistically bound the difference $|\hat{\theta} - \theta_0|$.
- Given this statistic, construct an uncertainty region for θ_0 composed of all θ for which $|\hat{\theta} - \theta|$ satisfies the α-bound of the test statistic (hypothesis-test).

Having in mind this classical way of determining uncertainty bounds, the following example is considered (Douma and Van den Hof, 2005).

Example 1. Consider the data generating system $\mathbf{y} = \theta_0 \boldsymbol{x}_1 + \boldsymbol{x}_2$, and one available measurement $\{y, x_1\}$ of $\mathbf{y}$ and $\boldsymbol{x}_1$. It is given that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are Gaussian distributed and correlated. We consider the following estimator for θ_0:

$$\boldsymbol{\theta} = \frac{\mathbf{y}}{\boldsymbol{x}_1}. \tag{1}$$

Under the above conditions the estimator (1) satisfies

$$\boldsymbol{\theta} = \frac{\mathbf{y}}{\boldsymbol{x}_1} = \theta_0 + \frac{\boldsymbol{x}_2}{\boldsymbol{x}_1}. \tag{2}$$

Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are correlated, the probability density function of this estimator will generally not be Gaussian [3]. Therefore, evaluation of parameter uncertainty regions on the basis of $p_{\boldsymbol{\theta}}$ will generally be cumbersome.
However since $\boldsymbol{x}_1(\boldsymbol{\theta} - \theta_0) = \boldsymbol{x}_2$, and a particular pair $x_1, \hat{\theta}$ is available from the measurement (y, x_1), it can easily be verified that

$$x_1(\hat{\theta} - \theta_0) = x_2 \tag{3}$$

where the term on the right hand side is unknown. Using the prior information that x_2 is a realization of the random variable $\boldsymbol{x}_2$ it simply follows that

$$(\hat{\theta} - \theta_0)x_1^2(\hat{\theta} - \theta_0) \leq \sigma^2_{\boldsymbol{x}_2} c_\chi(\alpha, 1) \quad \text{w.p. } \alpha, \tag{4}$$

where $c_\chi(\alpha, 1)$ corresponds to a probability level α in the Chi-squared distribution with one degree of freedom, i.e. the α probability region under a one-dimensional Gaussian distribution.
Since the distribution of the right hand side of (3) is known, we now consider the test statistic

$$x_1(\hat{\theta} - \tilde{\theta})$$

and select all the values of $\tilde{\theta}$ that lead to an $x_1(\hat{\theta} - \tilde{\theta})$ that is within the α probability level of the Gaussian distribution of $\boldsymbol{x}_2$. This set is exactly given by

$$\mathcal{D}(\alpha, \hat{\theta}) = \left\{ \theta \; |x_1(\hat{\theta} - \theta)|^2 \leq \sigma^2_{\boldsymbol{x}_2} c_\chi(\alpha, 1) \right\} \tag{5}$$

and it holds that

$$\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}) \text{ w.p. } \alpha.$$

The interpretation of this probabilistic expression is that when we construct the uncertainty region

[3] It is plotted in Figure 1 for $\boldsymbol{x}_2 \in \mathcal{N}(0, 2)$ and $\boldsymbol{x}_1 = 3 + \frac{0.5}{\boldsymbol{x}_2}$

$\mathcal{D}(\alpha, \hat{\theta})$ for n experiments, i.e. n realizations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the constructed region (5) will contain the true parameter only a number of αn times if $n \to \infty$.

The result for Example 1 is sketched in Figure 1. In the classical approach the parameter uncertainty region is determined on the basis of $\mathbb{E}\hat{\boldsymbol{\theta}}$ and $cov(\hat{\boldsymbol{\theta}})$. For a Gaussian distribution one then arrives at the smallest possible parameter uncertainty regions corresponding to a fixed probability level. However the above quantities need to be known. The alternative paradigm does not require full analysis of the pdf of the parameter estimator, at the possible cost of delivering larger parameter uncertainty sets, but with exact probabilistic expressions connected to it. In the next section it will be shown how the presented approach can be applied to ARX models.

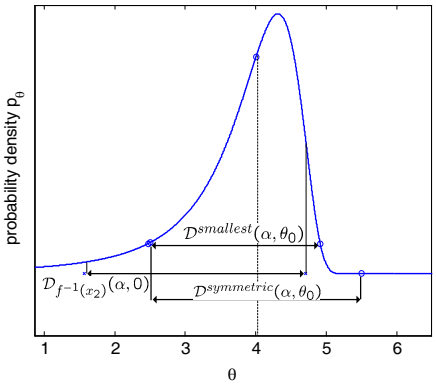

Fig. 1. Probability density function of $\boldsymbol{\theta}$ (example 1) and three uncertainty regions each corresponding to a probability of $\alpha = .9$. The symmetric and smallest 90% regions are tied to the pdf of $\boldsymbol{\theta}$. The computed 90% region $\mathcal{D}_{f^{-1}(x_2)}(\alpha, 0)$ corresponds to all $\hat{\theta}$ for which $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta})$. This region is based on a 90% probability region of random variable $\boldsymbol{x}_2$.

3. ARX MODELLING

In prediction error identification with ARX models a one-step-ahead predictor is considered of the format

$$\hat{y}(t|t-1;\theta) = \varphi^T(t)\theta \tag{6}$$

with $\varphi^T(t) = [-y(t-1)\cdots -y(t-n_a)\ u(t)\cdots u(t-n_b+1)]$, and $\theta^T = [a_1 \cdots a_{n_a}\ b_0 \cdots b_{n_b-1}]$, both having dimensions $n = n_a + n_b$. The parameter estimate is obtained by minimizing the quadratic prediction error criterion

$$\hat{\theta}_N = \arg\min_\theta V_N(\theta); \quad V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\theta)^2$$

with $\varepsilon(t,\theta) = y(t) - \hat{y}(t|t-1;\theta)$. By denoting

$$\boldsymbol{\Phi} = \begin{pmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{pmatrix} \text{ and } \mathbf{y} = [y(1)\cdots y(N)]^T$$

it follows that $\hat{\boldsymbol{\theta}}_N = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{y}$.
If the data generating system belongs to the model class ($\mathcal{S} \in \mathcal{M}$) then it holds that $\mathbf{y} = \boldsymbol{\Phi}\theta_0 + \mathbf{e}$ with $\mathbf{e}$ an N-vector of samples from a white noise process, and so

$$\hat{\boldsymbol{\theta}}_N = \theta_0 + (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{e}. \tag{7}$$

3.1 Classical approach

When analyzing the statistical properties of the estimator, it is generally derived [4] that, for $N \to \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \theta_0) \to \mathcal{N}(0, P_{arx})$$

with

$$P_{arx} = (\mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}])^{-1} \cdot Q \cdot (\mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}])^{-1} \tag{8}$$

with $Q = \mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}] \cdot \sigma_e^2$ and σ_e^2 the variance of the white noise, leading to

$$P_{arx} = (\mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}])^{-1} \cdot \sigma_e^2. \tag{9}$$

This leads to the expression that, asymptotically in N, $\theta_0 \in \mathcal{D}_{arx}(\alpha, \hat{\theta}_N)$ w.p. α, with

$$\mathcal{D}_{arx}(\alpha, \hat{\theta}_N) := \{\theta \mid (\theta - \hat{\theta}_N)^T P_{arx}^{-1}(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\}.$$

One of the problems in the latter expression is that P_{arx} cannot be computed, since σ_e^2 and $\mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}]$ are not available directly. Therefore in practice, the exact covariance matrix P_{arx} is commonly replaced by an estimate

$$\hat{P}_{arx} = \frac{1}{N}\Phi^T\Phi \cdot \hat{\sigma}_e^2, \tag{10}$$

where the estimate $\hat{\sigma}_e^2$ is usually determined on the basis of $\varepsilon^2(t, \hat{\theta}_N)$.

3.2 New approach

Alternatively we can use expression (7) to analyze the expression

$$\boldsymbol{\beta} := \frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\boldsymbol{\Phi}(\hat{\boldsymbol{\theta}}_N - \theta_0) = \frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\mathbf{e}.$$

The unknown term on the right hand side of the equation is known to satisfy

$$\boldsymbol{\beta} = \frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\mathbf{e} \in \mathcal{N}(0, Q), \quad Q = \mathbb{E}[\frac{1}{N}\boldsymbol{\Phi}^T\boldsymbol{\Phi}]\sigma_e^2$$

[4] For this derivation it is required that both the terms $(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}$ and $\boldsymbol{\Phi}^T\mathbf{e}$ as well as their product converge almost surely.

where the Gaussian distribution is reached asymptotically in N as a result of the Central Limit Theorem.

For realizations of $\boldsymbol{\beta} = \frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\mathbf{e}$, the following uncertainty bound can be specified asymptotically in N:

$$\boldsymbol{\beta} \in \mathcal{D}_\beta(\alpha, 0) \text{ w.p.}\alpha, \text{ with}$$
$$\mathcal{D}_\beta(\alpha, 0) := \left\{\beta \mid \beta^T Q^{-1}\beta \leq c_\chi(\alpha, n)\right\}$$

As this probabilistic expression is also valid for the particular estimate $\beta = \frac{1}{\sqrt{N}}\Phi^T\Phi(\hat{\theta}_N - \theta_0)$ on the basis of one single experiment, it follows that

$$(\hat{\theta}_N - \theta_0)^T \frac{1}{N}\Phi^T\Phi Q^{-1}\frac{1}{N}\Phi^T\Phi(\hat{\theta}_N - \theta_0) \leq$$
$$\leq \frac{c_\chi(\alpha, n)}{N} \text{ w.p. } \alpha$$

and consequently

$$\theta_0 \in \mathcal{D}_{arx}(\alpha, \hat{\theta}_N) \text{ w.p. } \alpha, \text{ with}$$
$$\mathcal{D}_{arx}(\alpha, \hat{\theta}_N) := \tag{11}$$
$$\{\theta \mid (\theta - \hat{\theta}_N)^T P_{arx,n}^{-1}(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\}$$

with

$$P_{arx,n} = (\frac{1}{N}\Phi^T\Phi)^{-1}Q(\frac{1}{N}\Phi^T\Phi)^{-1}. \tag{12}$$

Note that this expression is very close to the classical expression (8). However the two expected value expressions in this equation are simply replaced by computable data-based expressions. The current expression only requires the a.s. convergence of $\frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\mathbf{e}$. Again, as in the classical case, since σ_e^2 and $\mathbb{E}[\frac{1}{N}\Phi^T\Phi]$ are unknown, they are replaced by their estimates $\hat{\sigma}_e^2$ and $\frac{1}{N}\Phi^T\Phi$, leading to (11) with

$$\hat{P}_{arx,n} = (\frac{1}{N}\Phi^T\Phi)^{-1}\hat{\sigma}_e^2. \tag{13}$$

Whereas in the classical approach P_{arx} has the interpretation of covariance matrix of the parameter estimator, this interpretation is not applicable to the matrix $P_{arx,n}$. The latter expression only serves as a basis for the parameter uncertainty region.

3.3 Evaluation

In its implementable form (13) the alternative approach is seen to result in exactly the same uncertainty region as is practically used in the classical approach (10), based on the theoretical result (8). However, when comparing the two theoretical expressions (8) and (12) it appears that, besides the replacement of σ_e^2 by an estimate, the latter approach requires only the replacement of Q by a computable estimate, while the former requires three substitutions to be made. Summarizing, through the new paradigm the commonly used uncertainty region based on (10) has a stronger theoretical support than is generally acknowledged.

In terms of satisfying the (asymptotic) Gaussian distribution of the test statistic, it seems to result from Monte Carlo simulations that the term $\frac{1}{\sqrt{N}}\boldsymbol{\Phi}^T\mathbf{e}$ becomes Gaussian even for very small data length N. The term $(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{e}$ in the standard approach generally requires a longer data length to approximate the Gaussian distribution.

It is easily verified that the set (11) with (12) is a parameter uncertainty set that can equivalently be described by those values of θ that satisfy

$$V_N(\theta) - V_N(\hat{\theta}) \leq c_\chi(\alpha, n)/N.$$

This relates to the so-called likelihood method of quantifying parameter uncertainty (Donaldson and Schnabel, 1987).

4. EXTENSION OF THE ARX RESULTS

The probabilistic expressions in the previous section, are based on the statistical properties of the expression $\boldsymbol{\beta} := \frac{1}{\sqrt{N}}\Phi^T\mathbf{e}$. In order to further isolate the role of the noise distribution, the following Lemma will appear to be instrumental. A proof is added in the Appendix.

Lemma 1. Consider random vectors $\mathbf{z}, \mathbf{e} \in \mathbb{R}^{N\times 1}$ and a random matrix $\mathbf{V} \in \mathbb{R}^{N\times N}$ related through

$$\mathbf{z} = \mathbf{V}^T\mathbf{e}.$$

If the following properties are satisfied:

(1) $\mathbf{e}$ has independent identical Gaussian distributed entries, $\mathcal{N}(0, \sigma^2)$, and
(2) $\mathbf{e}$ and $\mathbf{V}$ are independent, and
(3) $\mathbf{V}$ is unitary, i.e. $\mathbf{V}^T\mathbf{V} = I$

then the vector elements of $\mathbf{z}$ are independent identically distributed with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. □

The result of this Lemma is quite remarkable. Irrespective of the *pdf* of the elements of matrix $\mathbf{V}$, the resulting random variable $\mathbf{z}$ has a Gaussian distribution. We can now formulate the following result that is relevant for ARX models.

Proposition 1. Given a random matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N\times n}$ and a random vector $\mathbf{e} \in \mathbb{R}^{N\times 1}$. The vector $\mathbf{e}$ contains independent identically distributed Gaussian random variables with distribution $\mathcal{N}(0, \sigma^2)$. If $\boldsymbol{\Phi}$ and $\mathbf{e}$ are statistically independent, then for every realization (z, Φ) of the random variables $(\mathbf{z}, \boldsymbol{\Phi})$ with

$$z := (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{e}$$

it holds that

$$z^T(\Phi^T\Phi)z \leq c_\chi(\alpha, n) \quad \text{is true w.p. } \alpha.$$

This result is exact for all values of N.

Proof Utilizing the singular value decomposition of $\mathbf{\Phi}^T$: $\mathbf{\Phi}^T = \mathbf{U\Sigma V}^T$, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = I$, it follows that $\mathbf{\Phi}^T\mathbf{\Phi z} = \mathbf{U\Sigma V}^T\mathbf{e}$.
We now denote

$$\boldsymbol{\beta} := \mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{\Phi}^T\mathbf{\Phi z} = \mathbf{V}^T\mathbf{e}. \tag{14}$$

If $\mathbf{V}$ and $\mathbf{e}$ are independent (upon assumption) then according to Lemma 1 $\boldsymbol{\beta} \in \mathcal{N}(0, \sigma^2 I_n)$ and consequently for every realization β corresponding to (z, Φ) with $\Phi = U\Sigma V^T$ it follows that

$$\beta^T\beta \leq c_\chi(\alpha, n) \quad \text{w.p. } \alpha.$$

Substituting β in the left hand side then delivers

$$z^T\Phi^T\Phi U\Sigma^{-1}\Sigma^{-1}U^T\Phi^T\Phi z$$

which by using $\Phi^T\Phi = U\Sigma^2U^T$, simplifies to $z^T(\Phi^T\Phi)z$ which proves the result. □

When applying Proposition 1 to estimated ARX models, the suggested consequence is that for all N with probability α:

$$(\hat{\theta}_N - \theta_0)^T \frac{1}{N}\Phi^T\Phi(\hat{\theta}_N - \theta_0) \leq \frac{c_\chi(\alpha, n)}{N}$$

and consequently $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$, w.p. α,

$$\text{with } \mathcal{D}(\alpha, \hat{\theta}_N) := \tag{15}$$
$$\{\theta \mid (\theta - \hat{\theta}_N)^T \frac{1}{N}\Phi^T\Phi(\theta - \hat{\theta}_N) \leq \frac{c_\chi(\alpha, n)}{N}\}.$$

This - very strong looking - result however is not exactly true, due to the fact that Proposition 1 requires statistical independence of $\mathbf{V}$ and $\mathbf{e}$, a condition that is not satisfied for ARX models. However in many situations it appears that failure to meet this condition hardly affects the Gaussian character of the test statistic $\boldsymbol{\beta}$. This is illustrated in the following example.

Example 2. A first-order data generating system is modelled with an ARX model of the form

$$\varepsilon(t, \theta) = (1 + \theta_f q^{-1})y(t) + \theta_b u(t),$$

such that $\mathcal{S} \in \mathcal{M}$. Experimental data is simulated driving the data-generating system with an input $u(t)$ and noise disturbance $e(t)$ that are independent Gaussian distributed white noise sequences with variance $\sigma_u^2 = \sigma_e^2 = 1$. The system coefficients are $\theta_b = 0.5$, $\theta_f = 0.9$. The parameters θ_b and θ_f are estimated with a least-squares identification criterion.
The top of Figure 2 depicts the histogram of the second element of $(\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{\Phi}^T\mathbf{e}$ corresponding with $\hat{\theta}_b$ as a function of data length N and for 5000 Monte Carlo simulations. It is the empirical distribution of the test statistic, related to the classical ARX estimate (7). The bottom depicts the distribution of the second element of $\mathbf{V}^T\mathbf{e}$, being the empirical distribution related to the test statistic (14). Clearly, the bottom row is indistinguishable from the Gaussian distribution, while the top approaches the Gaussian slowly. Similar results are presented in Figure 3 for the first elements of the parameter vector, corresponding with $\hat{\theta}_f$. Clearly, the bottom row is indistinguishable from the Gaussian distribution, while the top approaches the Gaussian distribution very slowly.

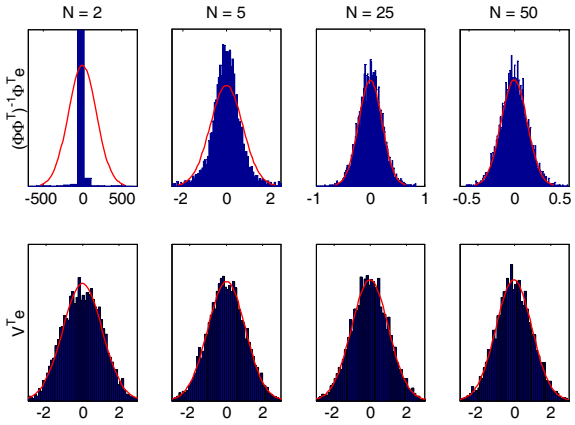

Fig. 2. Distribution of parameters in ARX structure. Top: second element of $(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{e}$ corresponding to $\hat{\theta}_b$ for data length $N = 2, 5, 25, 50$. Bottom: the distribution of the second element of $\mathbf{V}^T\mathbf{e}$ corresponding to $\hat{\theta}_b$.

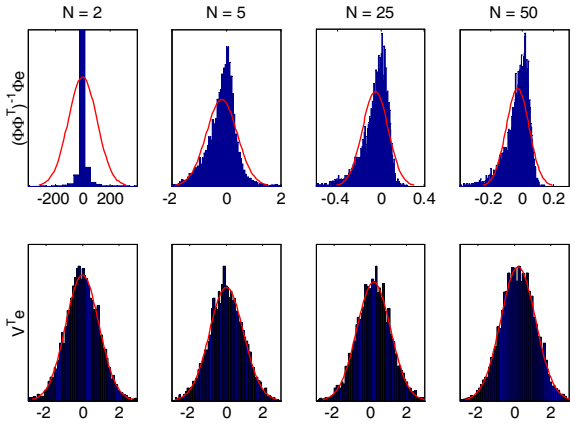

Fig. 3. Similar simulation results as in Figure 2 but then for the test statistic related to the denominator parameter $\hat{\theta}_f$.

The results of the example suggest that the existing correlation between $\mathbf{V}$ and $\mathbf{e}$ hardly affects the normality of the test statistic. This would allow to apply the Gaussian distribution to finite time signals. The interesting thing is that -along the lines sketched in Douma and Van den Hof (2005) similar results can be obtained for (nonlinearly parametrized) Output Error model structures.

5. CONCLUSIONS

The standard approach of formulating probabilistic parameter bounds on the basis of the statistical properties of the parameter estimator is discussed. Alternative approaches are presented that lead to exact probabilistic parameter bounding expressions for ARX models. Remarkably this alternative reasoning leads to bounding expressions that are similar to the current -implemented- expressions, in which unknown quantities in the theoretical expressions are replaces by data-based estimates. A particular ARX result is presented that suggests the applicability of the parameter bounding approach to finite-time data. This suggestion, supported by simulation results, will need to be substantiated in future work. Applicability of similar approaches to nonlinearly parametrized (Output Error) model structures is presented elsewhere.

ACKNOWLEDGEMENTS

The authors acknowledge fruitful discussions with Robert Bos, Arjan den Dekker and Xavier Bombois.

REFERENCES

Campi, M. and E. Weyer (2002). Finite sample properties of system identification methods. *IEEE Trans. Autom. Control*, **47**(8), 1329-1334.

de Vries, D.K. and P.M.J. Van den Hof (1995). Quantification of uncertainty in transfer function estimation: A mixed probabilistic-worst-case approach. *Automatica* **31**(4), 543–557.

Donaldson, J.R. and R.B. Schnabel (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics* **29**(1), 67–82.

Douma, S.G. and P.M.J. Van den Hof (2005). An alternative paradigm for probabilistic uncertainty bounding in prediction error identification. *Proc. 2005 CDC/ECC*, December 12-15, 2005, Seville, Spain.

Douma, S.G. (2006). *From Data to Robust Control - System Identification Uncertainty and Robust Control Design.* Dr. Dissertation, Delft Univ. Technology, to appear.

Goodwin, G.C., M. Gevers and B.M. Ninness (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Autom. Control* **37**, 913–928.

Hakvoort, R.G. and P.M.J. Van den Hof (1997). Identification of probabilistic uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Automatic Control* **42**(11), 1516–1528.

Heuberger, P.S.C., Van den Hof, P.M.J. and Wahlberg, B., Eds. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions.* Springer Verlag, 2005.

Ljung, L. (1999*a*). Model validation and model error modeling. In: *The Åström Symposium on Control* (B. Wittenmark and A. Rantzer, Eds.). Studentlitteratur. Lund, Sweden. pp. 15–42.

Ljung, L. (1999*b*). *System Identification — Theory for the User*, second ed. Prentice-Hall, Upper Saddle River, NJ.

Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Acad. Press, London, UK, 1981.

Weyer, E. and M.C. Campi (2002). Non-asymptotic confidence ellipsoids for the least-squares estimate. *Automatica*, **38**, 1539-1547.

APPENDIX: Proof of Lemma 1.

Define the vector valued function $g(\cdot) : \mathbb{R}^{(n+n^2)\times 1} \to \mathbb{R}^{(n+n^2)\times 1}$, defined by

$$g(\mathbf{z},\mathbf{v}) := \begin{bmatrix} \mathbf{e} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{V} & 0 \\ 0 & I^{n^2\times n^2} \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{v} \end{bmatrix}, \quad \text{(A.1)}$$

with $\mathbf{v} = col(\mathbf{V}^T)$ a vector containing all elements of $\mathbf{V}^T$. When denoting $\mathbf{e}' := [\mathbf{e}^T\ \mathbf{v}^T]^T$ and $\mathbf{z}' := [\mathbf{z}^T\ \mathbf{v}^T]^T$ it follows that the pdf

$$p_{\mathbf{z}}(z) = \int_v p_{\mathbf{z}'}(z')dv.$$

Using the mapping from $\mathbf{z}'$ to $\mathbf{e}'$ it follows from standard theory on the transformation of random variables (Priestley, 1981) that

$$p_{\mathbf{z}'}(z') = p_{\mathbf{e}'}(g(z')) \cdot \det(J(g(z')))$$

with the Jacobian given by

$$J(g(z')) = \begin{bmatrix} V & Z \\ 0 & I^{n^2\times n^2} \end{bmatrix},$$

and Z containing the partial derivatives of $V\mathbf{z}$ to $\mathbf{v}$. Consequently

$$\begin{aligned} p_{\mathbf{z}}(z) &= \int_v p_{\mathbf{e}'}(g(z')) \cdot \det(J)dv \\ &= \int_v p_{\mathbf{e}}(Vz)p_{\mathbf{v}}(v) \cdot \det(J)dv \end{aligned}$$

where the latter equation follows from the fact that $\mathbf{e}$ and $\mathbf{V}$ are independent. Using the Gaussian distribution of $\mathbf{e}$ and the fact that $\det(J) = \det(V) = 1$ it follows that

$$\begin{aligned} p_{\mathbf{z}}(z) &= \int_v \frac{1}{\sigma\sqrt{2\pi}} e^{-\sigma^{-2} z^T V^T V z}\ p_{\mathbf{v}}(v)dv \\ &= \int_v \frac{1}{\sigma\sqrt{2\pi}} e^{-\sigma^{-2} z^T z}\ p_{\mathbf{v}}(v)dv \\ &= p_{\mathbf{e}}(z) \int_v p_{\mathbf{v}}(v)dv = p_{\mathbf{e}}(z). \end{aligned}$$

□