

# Alternative paradigms for probabilistic model uncertainty bounding in prediction error identification

Paul M.J. Van den Hof

Sippe Douma, Arjan den Dekker, Xavier Bombois

Delft Center for Systems and Control  
Delft University of Technology  
The Netherlands

July 3, 2008

# Outline

Introduction

PE uncertainty bounds and hypothesis testing

ARX models

Reflection on finite-time perspectives

A likelihood perspective

Output error models

Conclusions

# Introduction

## Motivation

- ▶ General interest in bounding model uncertainty, motivated e.g. by “identification for control”
- ▶ Almost all results in PE uncertainty bounding are asymptotic in  $N$  (at least), require  $S \in \mathcal{M}$ , and the exact covariance matrix  $P_0$  to be known
- ▶ Finite time results, as well as handling of  $S \notin \mathcal{M}$  only available for FIR-like model structures
- ▶ Some alternative results available from Campi and Weyer (2002,2006).

## Main question

Explore the possibilities for probabilistic (parametric) model uncertainty bounding beyond the classical situation that relies heavily on asymptotics and consistent parameter estimates

# PE uncertainty bounds and hypothesis testing

## The set-up

Data-generating system:

$$y(t) = G_0(q)u(t) + H_0(q)e(t)$$

with  $e$  a Gaussian white noise (for the moment).

One-step ahead prediction error:

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

and the parameter estimator:

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta, Z^N)$$

$$\text{with } V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta),$$

Note:

boldface  $\hat{\theta}_N$  denotes random variable, as opposed to single realization  $\hat{\theta}_N$ .

## Reasoning to arrive at uncertainty bounds

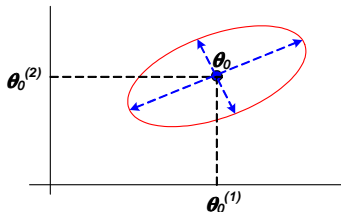
Start with asymptotic distribution:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, P_0)$$

and in quadratic form:

$$N(\hat{\theta}_N - \theta_0)^T P_0^{-1} (\hat{\theta}_N - \theta_0) \xrightarrow{N \rightarrow \infty} \chi_d^2$$

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



with

$$\mathcal{D}(\alpha, \theta_0) := \left\{ \theta \mid N(\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0) \leq \chi_{d,\alpha}^2 \right\}$$

However; we want to say something about  $\theta_0$ !

Since for every realization  $\hat{\theta}_N$ :

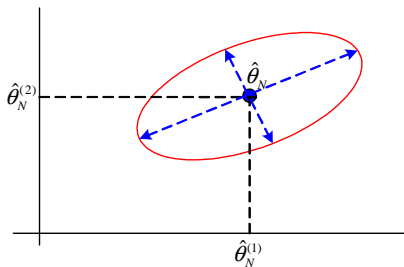
$$\hat{\theta}_N \in \mathcal{D}(\alpha, \theta_0) \Leftrightarrow \theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$$

it follows that

$$\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N) \text{ with probability } \alpha,$$

with

$$\mathcal{D}(\alpha, \hat{\theta}_N) := \left\{ \theta \mid N(\hat{\theta}_N - \theta)^T P_0^{-1} (\hat{\theta}_N - \theta) \leq \chi_{d,\alpha}^2 \right\}.$$



Note that this is an uncertainty set that varies with every realization  $\hat{\theta}_N$  of  $\hat{\theta}_N$ .

## Interpretation in terms of hypothesis test

Test the null hypothesis  $H_0: \theta = \theta_0$   
 against the alternative hypothesis  $H_1: \theta \neq \theta_0$   
 on the basis of the test statistic:

$$N(\hat{\theta}_N - \theta)^T P_0^{-1} (\hat{\theta}_N - \theta)$$

that under  $H_0$  is known to have a (asymptotic)  $\chi_d^2$  distribution.

## Standard hypothesis test

For a given realization  $\hat{\theta}_N$ , select those values of  $\theta$  that are within an  $\alpha$  confidence bound of the test statistic.

This leads to exactly the same ellipsoid

Test statistic is directly based on the estimator pdf. Is this necessary?

# ARX models

Consider e.g ARX models:

$$G(q, \theta) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)}, \quad H(q, \theta) = \frac{1}{A(q^{-1}, \theta)}$$

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

with  $\varphi^T(t) = [-y(t-1) \cdots -y(t-n_a) \ u(t) \cdots u(t-n_b+1)]$ .

By denoting

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

and using  $\mathcal{S} \in \mathcal{M}$ , it follows that

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

Following classical PE analysis:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, P_0)$$

with

$$P_0 = \left( \lim_{N \rightarrow \infty} \mathbb{E} \left[ \frac{1}{N} \Phi^T \Phi \right] \right)^{-1} \cdot \sigma_e^2.$$

According to the hypothesis testing procedure discussed before:

## Classical Result 1

On the basis of the test statistic

$$N(\hat{\theta}_N - \theta)^T P_0^{-1}(\hat{\theta}_N - \theta)$$

it follows that asymptotically in  $N$ ,  $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$  w.p.  $\alpha$ , with

$$\mathcal{D}(\alpha, \hat{\theta}_N) := \{\theta \mid N(\hat{\theta}_N - \theta)^T P_0^{-1}(\hat{\theta}_N - \theta) \leq \chi_{d,\alpha}^2\}. \quad (1)$$

This result is built on the asymptotic normality of the term  $(\Phi^T \Phi)^{-1} \Phi^T \mathbf{e}$ .

**Note:**  $P_0$  will generally be unknown.

## Looking for an alternative test-statistic

$$\hat{\theta}_N - \theta_0 = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{e}$$

With svd:  $\Phi^T = \mathbf{U}\Sigma\mathbf{V}^T$ , this can be written as

$$\Sigma\mathbf{U}^T(\hat{\theta}_N - \theta_0) = \mathbf{V}^T\mathbf{e}.$$

With the central limit theorem:

$$\mathbf{V}^T\mathbf{e} \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \sigma_e^2 I)$$

leading to the test statistic:

$$\begin{aligned} \frac{1}{\sigma_e^2}(\hat{\theta}_N - \theta)^T \mathbf{U}\Sigma^2\mathbf{U}^T(\hat{\theta}_N - \theta) &= \\ &= \frac{1}{\sigma_e^2}(\hat{\theta}_N - \theta)^T \Phi^T \Phi(\hat{\theta}_N - \theta) \end{aligned}$$

known to be (asymptotically)  $\chi_d^2$  distributed under  $\theta = \theta_0$ .

## Alternative Result 2

On the basis of the test statistic

$$\frac{1}{\sigma_e^2} (\hat{\theta}_N - \theta)^T \Phi^T \Phi (\hat{\theta}_N - \theta)$$

it follows that asymptotically in  $N$ ,  $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$  w.p.  $\alpha$ , with

$$\mathcal{D}(\alpha, \hat{\theta}_N) := \left\{ \theta \mid \frac{1}{\sigma_e^2} (\hat{\theta}_N - \theta)^T \Phi^T \Phi (\hat{\theta}_N - \theta) \leq \chi_{d,\alpha}^2 \right\}. \quad (2)$$

This result is built on the asymptotic normality of the term  $\mathbf{V}^T \mathbf{e}$ .

**Note:** All terms in the uncertainty set are known, except  $\sigma_e^2$ .

# Intermediate discussion

- ▶ If unknown  $P_0$  (Result 1) is replaced by sample-estimate, both results become exactly the same!
- ▶ In Result 1, this replacement is a compromise, in Result 2 it is not.
- ▶ When  $\sigma_e^2$  is to be estimated also, the  $\chi_d^2$  distributions need to be replaced by  $\mathcal{F}$  distributions.

# Reflection on finite-time perspectives

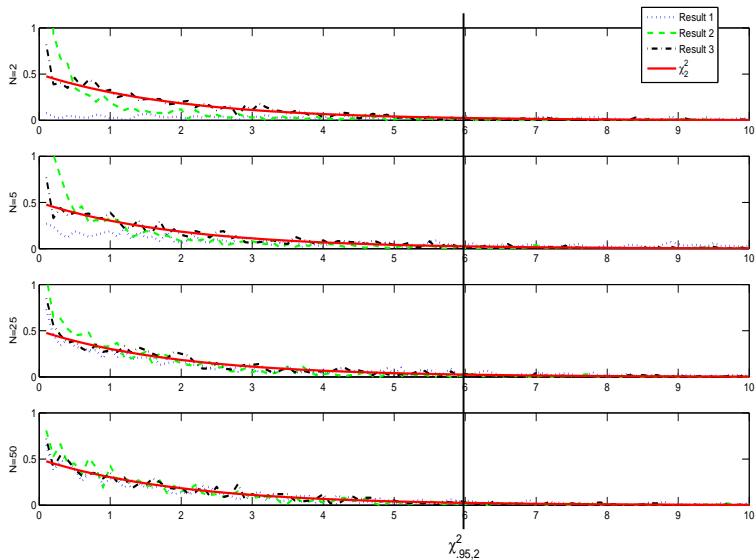
The two results rely on asymptotic normality of:

$$\begin{array}{ll} (\Phi^T \Phi)^{-1} \Phi^T \mathbf{e} & \text{Result 1} \\ \mathbf{V}^T \mathbf{e} & \text{Result 2} \end{array}$$

Verification in simulation example of 1st order ARX model:

$$\varepsilon(t, \theta) = (1 + 0.5q^{-1})y(t) + 0.9u(t)$$

Input signal and noise signal are unit variance Gaussian white noise. Evaluate, for different values of  $N$ , 5000 Monte Carlo results of the relevant random variables above, and show the test statistics:

$\chi^2$  test statistics:

Coverage rates for  $\alpha = 0.95$ :

	$N = 2$	$N = 5$	$N = 25$	$N = 50$
Result 1	0.1810	0.5530	0.8320	0.8760
Result 2	0.9610	0.9500	0.9590	0.9550

## Further justification of this result

### Lemma

If  $\mathbf{V}$  and  $\mathbf{e}$  are independent,  $\mathbf{V}$  unitary and  $e(t) \in \mathcal{N}(0, \sigma_e^2)$ , then  $\mathbf{V}^T \mathbf{e} \in \mathcal{N}(0, \sigma_e^2)$  for any value of  $N$ .

**Note:** Independence of  $\mathbf{V}$  and  $\mathbf{e}$  is formally not satisfied for ARX models (due to correlation between  $\Phi$  and  $\mathbf{e}$ ). However the deteriorating effect of this seems minimal.

## Observation

Replacing (covariance) matrices by sample estimates in parameter uncertainty bounds is not at all a compromise, but rather well justified from a statistical and a practical point of view.

# A likelihood perspective

When  $\mathbf{e}$  is Gaussian (as assumed here) we can phrase the results in a likelihood framework, with the log-likelihood:

$$\log f_{\mathbf{y}}(\theta; \mathbf{y}^N) = -\frac{N}{2} \log(2\pi) - N \log \sigma_e - \frac{N}{2\sigma_e^2} V_N(\theta).$$

Define the **generalized likelihood ratio**  $L_G(\theta)$  as (Kay, 1998):

$$L_G(\theta) = \frac{f_{\mathbf{y}}(\theta; \mathbf{y}^N)}{\sup_{\theta} f_{\mathbf{y}}(\theta; \mathbf{y}^N)} = \frac{f_{\mathbf{y}}(\theta; \mathbf{y}^N)}{f_{\mathbf{y}}(\hat{\theta}_N; \mathbf{y}^N)}$$

with  $\hat{\theta}_N$  the maximum likelihood estimator.

Then (Kay, 1998):

$$-2 \log L_G(\theta_0) \rightarrow \chi_d^2$$

As a result,  $-2 \log L_G(\theta)$  can be used as a test statistic for quantifying the model uncertainty.

## Proposition

For the ARX models considered,

$$\begin{aligned} -2 \log L_G(\theta_0) &= \frac{N}{\sigma_e^2} \left[ V_N(\theta_0) - V_N(\hat{\theta}_N) \right] \\ &= \frac{N}{\sigma_e^2} (\hat{\theta}_N - \theta_0)^T \frac{1}{N} \Phi^T \Phi (\hat{\theta}_N - \theta_0) \end{aligned}$$

## Observation

Uncertainty bounding Result 2 is equivalent to a generalized likelihood ratio test, and comes down to considering a level set of the identification cost function  $V_N(\theta)$ , i.e.  $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta}_N)$  w.p.  $\alpha$  with

$$\mathcal{D}(\alpha, \hat{\theta}_N) = \left\{ \theta \mid \frac{N}{\sigma_e^2} [V_N(\theta) - V_N(\hat{\theta}_N)] \leq \chi_{d, \alpha}^2 \right\}.$$

# Output error models

Can the same mechanisms be used for nonlinearly parametrized (Output error) models?

$$\varepsilon(t, \theta) = y(t) - \frac{B(q^{-1}, \theta)}{F(q^{-1}, \theta)} u(t)$$

## Classical result:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow \mathcal{N}(0, P_{oe})$$

with

$$P_{oe} = \sigma_e^2 \cdot \left[ \lim_{N \rightarrow \infty} \frac{1}{N} \Psi(\theta_0)^T \Psi(\theta_0) \right]^{-1}$$

with  $\Psi(\theta)$  a matrix with rows  $\psi^T(t, \theta) = \frac{\partial}{\partial \theta} \hat{y}(t|t-1, \theta)$ .

An ellipsoidal confidence bound for  $\theta_0$  is obtained on the basis of the test statistic

$$\frac{1}{N} (\hat{\theta}_N - \theta)^T P_{oe}^{-1} (\hat{\theta}_N - \theta)$$

Result relies on Taylor approximation:  $(\hat{\theta}_N - \theta_0) \approx -[\bar{V}_N''(\theta_0)]^{-1} [V_N'(\theta_0)]$

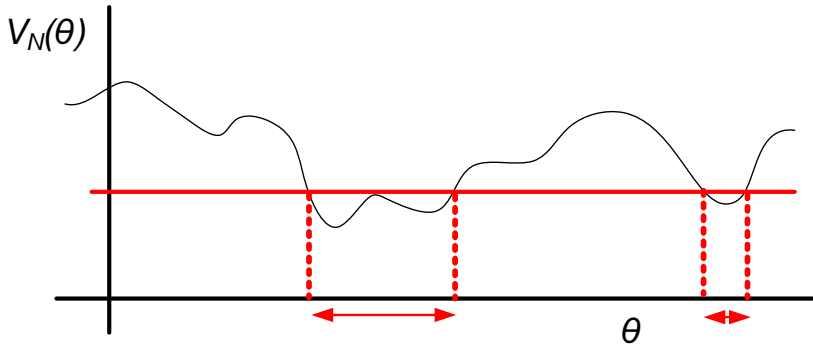
# Alternatives

- ▶ Similar step as for ARX:  $\Psi^T(\hat{\theta}_N)\Psi(\hat{\theta}_N)[\hat{\theta}_N - \theta_0] = \Psi^T(\hat{\theta}_N)\mathbf{e}$   
(Taylor approxim)  
with  $\Psi$  containing predictor derivatives w.r.t.  $\theta$
- ▶ **Likelihood ratio** approach:

$$-2 \log L_G = \frac{N}{\sigma_e^2} [V_N(\theta) - V_N(\hat{\theta}_N)] \rightarrow \chi_d^2$$

under  $\theta = \theta_0$ .

This leads to level sets of the identification cost function (no ellipsoid).



Uncertainty set on the basis of “level sets” can easily lead to disconnected regions,  
but are very appealing from an engineering perspective.

# Alternatives (contin.)

- ▶ The cost function derivative:

$$V'_N(\theta_0) = \boldsymbol{\Psi}^T(\theta_0)\mathbf{e} \in \mathcal{N}(\mathbf{0}, \mathbf{J}(\theta_0)); \quad \mathbf{J}(\theta_0) = \frac{1}{\sigma_e^2} \boldsymbol{\Psi}^T(\theta_0)\boldsymbol{\Psi}(\theta_0)$$

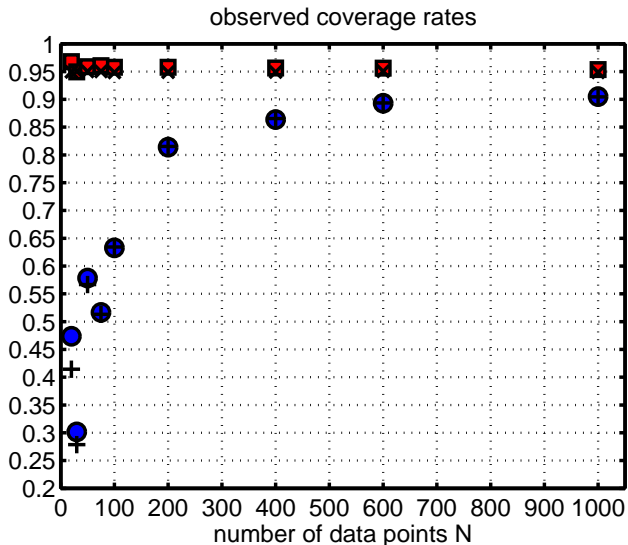
## Fisher score or Rao test.

As in OE models  $\boldsymbol{\Psi}(\theta_0)$  is not correlated with  $\mathbf{e}$  the distribution holds for finite-time also (provided that  $\mathbf{e}$  is Gaussian).

Related test statistic:

$$\frac{1}{\sigma_e^2} \boldsymbol{\varepsilon}^T(\theta) \boldsymbol{\Psi}(\theta) [\boldsymbol{\Psi}^T(\theta) \boldsymbol{\Psi}(\theta)]^{-1} \boldsymbol{\Psi}^T(\theta) \boldsymbol{\varepsilon}(\theta) \in \chi_d^2$$

This is no ellipsoid!

Coverage rates for a second order OE model ( $\alpha = 0.95$ ,  $M = 50,000$ )

Classical (blue circles); LR (red  $\square$ ); Rao-test (black  $\times$ )

# Conclusions

- ▶ We have explored a degree of freedom in formulating model uncertainty bounds
- ▶ Bounds on the basis of data-based “covariance matrices” seem better than the theoretical ones
- ▶ Finite-time results are available for OE models (but not in ellipsoidal form)
- ▶ In nonlinearly parametrized model sets (OE), linearization seems to be a severe limitation
- ▶ Simply extendable to IV methods

# References



S.G. Douma and P.M.J. Van den Hof.

An alternative paradigm for probabilistic uncertainty bounding in prediction error identification. In *Proc. 44th IEEE Conf. Decision and Control and European Control Conf., CDC-ECC'05*, pages 4970–4975, Seville, Spain, December 2005.



S.G. Douma.

*From Data to Performance - System Identification Uncertainty and Robust Control Design*. PhD thesis, Delft Univ. Technology, Delft, The Netherlands, 2006.



S.G. Douma and P.M.J. Van den Hof.

Probabilistic model uncertainty bounding: an approach with finite-time perspectives. In *Prepr. 14th IFAC Symp. System Identification*, pages 1021–1026, Newcastle, NSW, Australia, March 2006.



S.G. Douma and P.M.J. Van den Hof.

Probabilistic uncertainty bounding in output error models with unmodelled dynamics. In *Proc. 2006 American Control Conf.*, Minneapolis, MA, pp. 1677–1682.



A.J. den Dekker, X. Bombois, and P.M.J. Van den Hof.

Likelihood based uncertainty bounding in prediction error identification using ARX models: a simulation study. In *Proc. 2007 European Control Conf.*, Kos, Greece, pp. 2879–2886.



A.J. den Dekker, X. Bombois, and P.M.J. Van den Hof.

Finite sample confidence regions for parameters in prediction error identification using output error models. In *Proc. 2008 IFAC World Congress*, Seoul, 6-11 July 2008.