An Alternative Paradigm for Probabilistic Uncertainty Bounding in Prediction Error Identification

Sippe Douma and Paul Van den Hof

2005 ERNSI Workshop, Louvain-la-Neuve, Belgium, 19-21 September 2005

The current status

- In prediction error identification, model uncertainty bounds are necessary (reliability, robustness)
- Probabilistic bounds based on covariance P of estimator, together with (asymptotic) normal distribution
- For linear parametrizations (FIR,ARX) and $\mathcal{S} \in \mathcal{M}$ exact and explicit expression for P is available
- For nonlinear parametrizations (OE,BJ) approximations are necessary (e.g. Taylor expansion)
- For $\mathcal{S} \notin \mathcal{M}$ results can be obtained only for linear parametrizations (FIR, ORTFIR)

The message

- In prediction error identification, quantified model uncertainty is usually based on
  pdf of estimator $z \rightarrow \hat{\theta}$
- "Exact" probabilistic expressions on $\hat{\theta}_N - \theta_0$ are approximated by:
  - Employing asymptotic Gaussian distribution
  - Obtaining P through Taylor approximation (OE/BJ)
  - Replacing covariance matrix by estimate
- Probabilistic parameter uncertainty regions can be obtained without specifying the estimator pdf, with attractive results even for nonlinear estimators

Contents

Example

Data generating system: $\mathbf{y} = \theta_0 \boldsymbol{x}_1 + \boldsymbol{x}_2$

$\boldsymbol{x}_2 \in \mathcal{N}(0,2)$; $\boldsymbol{x}_1$ correlated with $\boldsymbol{x}_2$

Estimator: $\boldsymbol{\theta} = \mathbf{y}/\boldsymbol{x}_1 = \theta_0 + \boldsymbol{x}_2/\boldsymbol{x}_1$

pdf of $\boldsymbol{\theta}$ is very hard to analyze

However: $\boldsymbol{x}_1(\boldsymbol{\theta} - \theta_0) = \boldsymbol{x}_2 \in \mathcal{N}(0,2)$

After one experiment we have realizations: $x_1, \hat{\theta}$ of $\boldsymbol{x}_1, \boldsymbol{\theta}$
Then $x_1(\hat{\theta} - \theta_0)$ is a realization of $\boldsymbol{x}_2 \in \mathcal{N}(0,2)$.
Based on test statistic $x_1(\hat{\theta} - \tilde{\theta})$ we select all $\tilde{\theta}$ that are within the α-probability level of $\boldsymbol{x}_2$:

$\theta_0 \in \left\{\theta \mid (\hat{\theta} - \theta) x_1^2 (\hat{\theta} - \theta) \leq 2 c_\chi(\alpha, 1)\right\}$ w.p. α

Probabilistic parameter bounding without pdf of estimator

Employ statistical properties of random variable

$$\boldsymbol{x}_1(\boldsymbol{\theta} - \theta_0) = \boldsymbol{x}_2 \in \mathcal{N}(0,2)$$

rather than those of

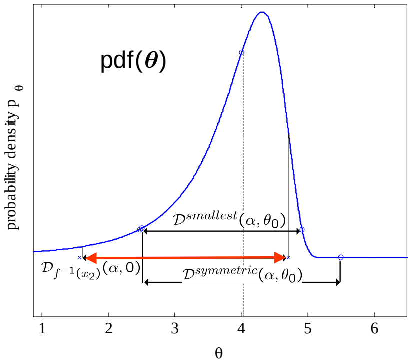
$$\boldsymbol{\theta} - \theta_0$$

Benefit = simplicity of expression

classical:
Fixed (symmetric) interval is estimated from data

interval varies with experiment

Cost: not necessarily smallest interval



ARX modelling

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

With

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

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

If $\mathcal{S} \in \mathcal{M}$: $\mathbf{y} = \Phi\theta_0 + \mathbf{e}$

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

ARX modelling

If $\mathcal{S} \in \mathcal{M}$: $\hat{\theta}_N - \theta_0 = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{e}$

Classical approach:

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

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

$$\theta_0 \in \left\{\theta \mid (\hat{\theta}_N - \theta)P_{arx}^{-1}(\hat{\theta}_N - \theta) \leq c_\chi(\alpha, n)/N\right\} \quad \text{w.p.} \quad \alpha$$

Requires:

- (asymptotic) normality of $(\Phi^T\Phi)^{-1}\Phi^T\mathbf{e}$
- Replacement of P_{arx} by an estimate $\hat{P}_{arx}$

ARX modelling

If $\mathcal{S} \in \mathcal{M}$: $\hat{\theta}_N - \theta_0 = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{e}$

Alternative:

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

$$\to \mathcal{N}(0, Q) \quad Q = \mathbb{E}[\frac{1}{N}\Phi^T\Phi] \cdot \sigma_e^2$$

ARX modelling

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

$$\to \mathcal{N}(0, Q) \quad Q = \mathbb{E}[\frac{1}{N}\Phi^T\Phi] \cdot \sigma_e^2$$

Result

$$\theta_0 \in \left\{\theta \mid (\hat{\theta}_N - \theta)^T P_{arx,n}^{-1}(\hat{\theta}_N - \theta) \leq \frac{c_\chi(\alpha, n)}{N}\right\} \text{w.p.}\alpha$$

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

Requires:

- (asymptotic) normality of $\Phi^T\mathbf{e}$
- Replacement of Q by an estimate

ARX modelling

Implementable scheme:

Replace $Q = \mathbb{E}[\frac{1}{N}\Phi^T\Phi] \cdot \sigma_e^2$ by $\frac{1}{N}\Phi^T\Phi\hat{\sigma}_e^2$

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

Same expression as used in the classical situation

Result is related to likelihood method, determined by

$$\left\{\theta \mid V_N(\theta) - V_N(\hat{\theta}_N) \leq c_\chi(\alpha, n)/N\right\}$$

ARX modelling

Implementable scheme:

Replace $Q = \mathbb{E}[\frac{1}{N}\Phi^T\Phi] \cdot \sigma_e^2$ by $\frac{1}{N}\Phi^T\Phi\hat{\sigma}_e^2$

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

Same expression as used in the classical situation

Conclusion

Classical results with P_{arx} approximated by sample estimates, has stronger theoretical support than often considered.

Simulation example:

First order ARX system:

$$y(t) = \frac{0.5}{1+0.9q^{-1}}u(t) + \frac{1}{1+0.9q^{-1}}e(t)$$

identified with 1st order ARX model.

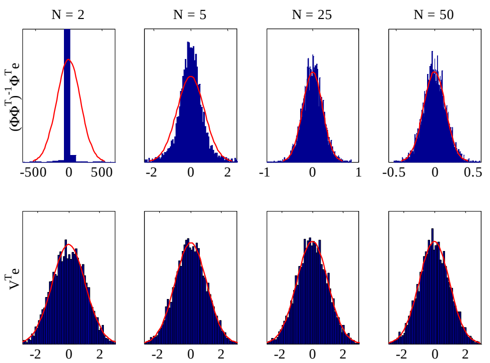
Compare empirical distributions of

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

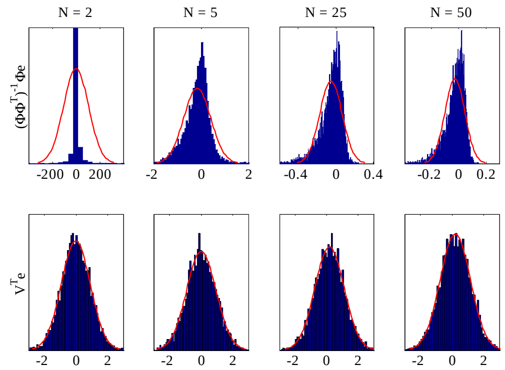
and $\frac{1}{\sqrt{N}}\Phi^T\mathbf{e}$

for different values of N,
on the basis of 5000 Monte Carlo simulations

Component related to numerator parameter:



Component related to denominator parameter:



OE modelling

$$\hat{y}(t|t-1;\theta) = \frac{B(q,\theta)}{F(q,\theta)}u(t)$$

Then $V_N'(\hat{\theta}_N) = 0$ can be written as

$$\frac{1}{N}\sum_{t=1}^{N}[y(t) - \frac{B(q,\hat{\theta}_N)}{F(q,\hat{\theta}_N)}u(t)] \cdot \psi(t,\hat{\theta}_N) = 0$$

$$\psi(t,\theta) = \frac{\partial}{\partial\theta}\hat{y}(t|t-1;\theta)$$

and $$\frac{1}{N}\sum_{t=1}^{N}[F(q,\hat{\theta}_N)y_F(t) - B(q,\hat{\theta}_N)u_F(t)] \cdot \psi(t,\hat{\theta}_N) = 0$$

with $y_F(t) = F(q,\hat{\theta}_N)^{-1}y(t); \quad u_F(t) = F(q,\hat{\theta}_N)^{-1}u(t)$

OE modelling

Linear regression type of equation; solution satisfies

$$\hat{\theta}_N = (\Psi^T\Phi)^{-1}\Psi^T\mathbf{y}_F$$

with

$$\Phi^T = \left[\varphi_F^T(1,\hat{\theta}_N), \cdots \varphi_F^T(N,\hat{\theta}_N)\right]; \Psi^T = \left[\psi^T(1,\hat{\theta}_N) \cdots \psi^T(N,\hat{\theta}_N)\right]$$

$$\varphi_F^T(t,\hat{\theta}_N) = [-y_F(t-1) \cdots -y_F(t-n_f)\ u_F(t) \cdots u_F(t-n_b+1)]$$

$$\frac{1}{N}\sum_{t=1}^{N}[F(q,\hat{\theta}_N)y_F(t) - B(q,\hat{\theta}_N)u_F(t)] \cdot \psi(t,\hat{\theta}_N) = 0$$

with $y_F(t) = F(q,\hat{\theta}_N)^{-1}y(t); \quad u_F(t) = F(q,\hat{\theta}_N)^{-1}u(t)$

OE modelling

Linear regression type of equation; solution satisfies

$$\hat{\theta}_N = (\Psi^T\Phi)^{-1}\Psi^T\mathbf{y}_F$$

Not fit for parameter estimation, since r.h.s. is parameter-dependent.
However since r.h.s. is known once $\hat{\theta}_N$ is determined, similar uncertainty analysis can be made as for ARX

With $\mathbf{y}_F = \Phi\theta_0 + \mathbf{e}_F$

$$\frac{1}{\sqrt{N}}(\Psi^T\Phi)(\hat{\theta}_N - \theta_0) \quad = \quad \frac{1}{\sqrt{N}}\Psi^T\mathbf{e}_F$$

OE modelling

$$\theta_0 \in \left\{\theta |\ (\hat{\theta}_N - \theta)^T P_{oe,n}^{-1}(\hat{\theta}_N - \theta) \leq \frac{c_\chi(\alpha,n)}{N}\right\} \text{ w.p. } \alpha$$

$$P_{oe,n} = (\frac{1}{N}\Psi^T\Phi)^{-1}Q(\frac{1}{N}\Phi^T\Psi)^{-1}, \qquad Q = \sigma_{\mathbf{e}_F}^2\mathbb{E}[\frac{1}{N}\Psi^T\Psi]$$

Requires:

- (asymptotic) normality of $\Psi^T\mathbf{e}_F/\sqrt{N}$
- Replacement of Q by an estimate
- No 1st order Taylor approximation involved

Classical: $P_{oe} = \sigma_e^2[\mathbb{E}\frac{1}{N}\Psi(\theta_0)^T\Psi(\theta_0)]^{-1}$

Simulation example:

First order OE system:

$$y(t) = \frac{0.5}{1 + 0.9q^{-1}}u(t) + e(t)$$

identified with 1st order OE model.

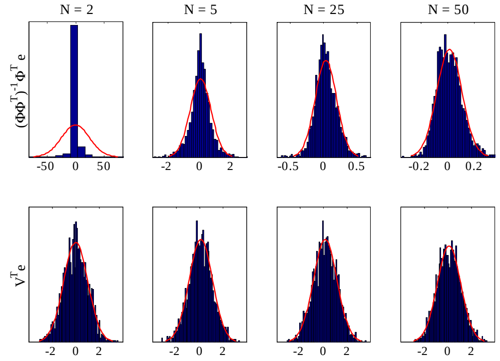
Compare empirical distributions of

$$\hat{\theta}_N - \theta_0 \;=\; (\Psi^T\Phi)^{-1}\Psi^T e_F$$

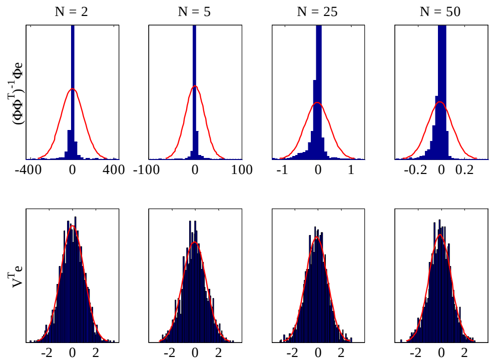
and $\frac{1}{\sqrt{N}}\Psi^T e_F$

for different values of N,
on the basis of 5000 Monte Carlo simulations

Component related to numerator parameter:



Component related to denominator parameter:



Summary

- There is an alternative paradigm for parameter uncertainty bounding, without constructing pdf of estimator
- Applicable to ARX, OE and also BJ models
- Leading to simpler and less approximative expressions
- Can be extended to OE models, even when $\mathcal{S} \notin \mathcal{M}$
- Relation with Bayesian and likelihood based uncertainty intervals needs to be explored

One step further

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

With svd: $\Psi^T = U \Sigma V^T$ it follows that

$$\Sigma^{-1} U^T (\Psi^T \Phi)(\hat{\theta}_N - \theta_0) = V^T \mathbf{e}_F$$

Lemma:

If V^T unitary and random, and e Gaussian with cov(e)=σ^2 I, and V^T and e independent, then $V^T e$ is Gaussian with Cov = σ^2 I.

This would suggest that $V^T \mathbf{e}_F$ is Gaussian for any value of N.

Only (2nd order) effect: V^T and $\mathbf{e}_F$ both depend on $\hat{\theta}_N$