Likelihood Based Uncertainty Bounding in Prediction Error Identification using ARX models

A Simulation Study

ECC 2007, Kos, Greece

Arjan den Dekker, Xavier Bombois and Paul Van den Hof

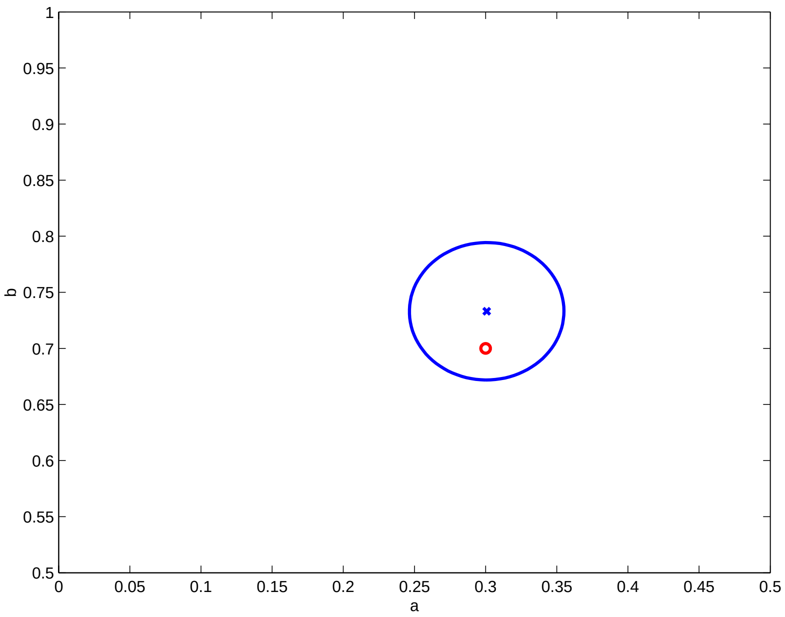
July 4, 2007

DCSC
Delft Center for Systems and Control

TU Delft
Delft University of Technology

Confidence regions

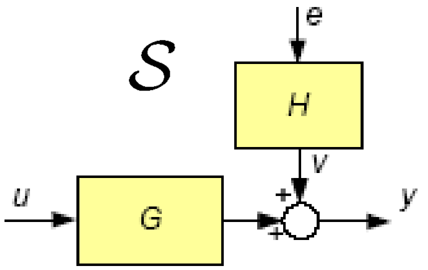
A 95% confidence region is a region in parameter space that attempts to cover the "true" parameter with probability 0.95.



Outline

- Data generating system and predictor model
- Statistical inference in PE identification
- Confidence regions and hypothesis tests
- Test statistics
- ARX modelling
- Simulation results
- Conclusions

Data generating system and predictor model



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

- input (deterministic): $u^N = \{u(t)\}_{t=1,\ldots,N}$
- output (stochastic): $y^N = \{y(t)\}_{t=1,\ldots,N}$
- $e(t)$ Gaussian white noise

Predictor model set ($\mathcal{S} \in \mathcal{M}$):

$$\mathcal{M} \;:=\; \{(G(q,\theta), H(q,\theta))|\theta \in \Theta \subset \mathbb{R}^n\}$$

$$\exists\theta_0 \quad \ni \quad G(q,\theta_0) = G_0(q) \text{ and } H(q,\theta_0) = H_0(q)$$

Statistical inference

One-step ahead predictor:

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

Prediction errors:

$$\begin{aligned} \epsilon(t,\theta) &= y(t) - \hat{y}(t|t-1;\theta) \\ \epsilon(t,\theta_0) &= e(t) \sim \mathcal{N}(0,\sigma^2) \end{aligned}$$

Joint probability density function of y^N:

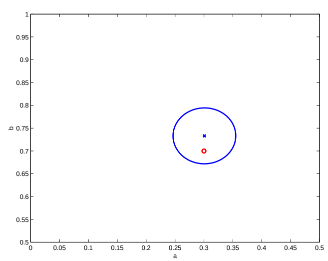
$$f_y(y^N;\theta_0) = \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (\underbrace{y(t) - \hat{y}(t|t-1;\theta_0)}_{\epsilon(t,\theta_0)})^2 \right]$$

Statistical inference

$$S(\theta) \;=\; \frac{\partial \log f_y(y^N;\theta)}{\partial\theta} = \frac{-N}{2\sigma^2}\frac{\partial V_N(\theta)}{\partial\theta}$$

$$F(\theta) \;=\; -\mathbb{E}\left[\frac{\partial^2 \log f_y(y^N;\theta)}{\partial\theta^2}\right] = \frac{N}{2\sigma^2}\mathbb{E}\left[\frac{\partial^2 V_N(\theta)}{\partial\theta^2}\right]$$

$$\hat{\theta}_N = \arg\max_{\theta} f_y(\theta; y^N) = \arg\min_{\theta} \underbrace{\frac{1}{N}\sum_{t=1}^{N}\epsilon(t,\theta)^2}_{V_N(\theta)}$$



Confidence region for θ_0?

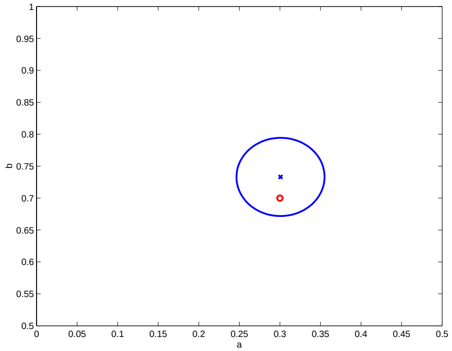
Construction of confidence regions

$$H_0 : \theta_0 = \theta$$
$$H_1 : \theta_0 \neq \theta$$

To test H_0 against H_1 at significance level α, choose a test statistic $T(y^N, \theta)$ with a known distribution under H_0 and decide H_1 if:

$$T(y^N, \theta) > c(\alpha)$$

with $Pr[T(y^N, \theta) > c(\alpha)] = \alpha$ under H_0.



$100(1-\alpha)\%$ confidence region for θ_0:

$$\left\{\theta | T(y^N, \theta) \leq c(\alpha)\right\}$$

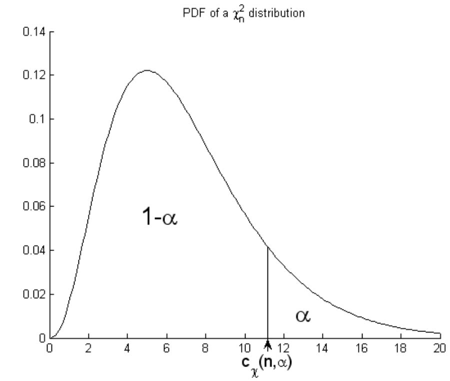
Test statistics

$\hat{\theta}_N \overset{as.}{\sim} \mathcal{N}(\theta_0, F^{-1}(\theta_0))$	$(\hat{\theta}_N - \theta_0)^T F(\theta_0)(\hat{\theta}_N - \theta_0) \overset{as.}{\sim} \chi^2_n$
$S(\theta_0) \overset{as.}{\sim} \mathcal{N}(0, F(\theta_0))$	$S(\theta_0)^T F^{-1}(\theta_0) S(\theta_0) \overset{as.}{\sim} \chi^2_n$
$-2\log \frac{f_y(\theta_0; y^N)}{f_y(\hat{\theta}_N; y^N)} = \frac{N}{\sigma^2}(V_N(\theta_0) - V_N(\hat{\theta}_N)) \overset{as.}{\sim} \chi^2_n$:	

$$\left.\begin{aligned} T_T(y^N, \theta) &= (\hat{\theta}_N - \theta)^T F(\theta)(\hat{\theta}_N - \theta) \\ T_W(y^N, \theta) &= (\hat{\theta}_N - \theta)^T F(\hat{\theta}_N)(\hat{\theta}_N - \theta) \\ T_R(y^N, \theta) &= S(\theta)^T F^{-1}(\theta) S(\theta) \\ T_{LR}(y^N, \theta) &= \frac{N}{\sigma^2}(V_N(\theta) - V_N(\hat{\theta}_N)) \end{aligned}\right\} \overset{as.}{\sim} \chi^2_n \text{ under } H_0$$

$(100 - \alpha)\%$ confidence regions for θ_0:

$$\left\{\theta \,|\, T(y^N, \theta) \leq c_\chi(n, \alpha)\right\}$$



ARX modelling

$$
\begin{aligned}
G(q,\theta) &= \frac{q^{-n_k}(b_0 + b_1 q^{-1} + \cdots + b_{n_b-1} q^{-n_b+1})}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}} \\
H(q,\theta) &= \frac{1}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a})} \\
\theta^T &= [a_1 \cdots a_{n_a} b_0 \cdots b_{n_b-1}] \\
\varphi^T(t) &= [-y(t-1) \cdots - y(t-n_a) u(t-n_k) \cdots u(t-n_k-n_b+1)] \\
\Phi &= \begin{pmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{pmatrix} \\
\mathbf{y}^T &= [y(1) \cdots y(N)] \\
& N(V_N(\theta) - V_N(\hat{\theta}_N)) = (\theta - \hat{\theta}_N)^T \Phi \Phi^T (\theta - \hat{\theta}_N) \\
S(\theta) &= \frac{1}{\sigma^2} \Phi^T (\mathbf{y} - \Phi\theta); \qquad F(\theta_0) = \frac{1}{\sigma^2} \underbrace{\mathbb{E}\left[\Phi^T \Phi\right]}_{R(\theta_0)}
\end{aligned}
$$

Confidence regions for ARX parameters

$100(1-\alpha)\%$ confidence regions for θ_0:
$$\left\{\theta | (\theta - \widehat{\theta}_N)^T X (\theta - \widehat{\theta}_N) \leq \sigma^2 c_\chi(n,\alpha)\right\}$$

Test statistic	X
T_T	$R(\theta)$
T_W (Wald)	$R(\widehat{\theta}_N)$
T_R (Rao)	$\Phi^T \Phi R^{-1}(\theta) \Phi^T \Phi$
T_{LR} (Likelihood Ratio)	$\Phi^T \Phi$

$$R(\theta_0) = \mathbb{E}\left[\Phi^T \Phi\right].$$

Result from asymptotic theory (θ_0 known)

Asymptotically valid expression for $cov(\widehat{\theta}_N)$:

$$P_{\theta_0} = \frac{\sigma^2}{N}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}_{ue}\left[\varphi(t)\varphi^T(t)\right]\right)^{-1}$$

$100(1-\alpha)\%$ confidence region for θ_0

$$\left\{\theta | (\widehat{\theta}_N - \theta)^T P_{\theta_0}^{-1}(\widehat{\theta}_N - \theta) \leq c_\chi(n,\alpha)\right\}$$

Simulation experiment

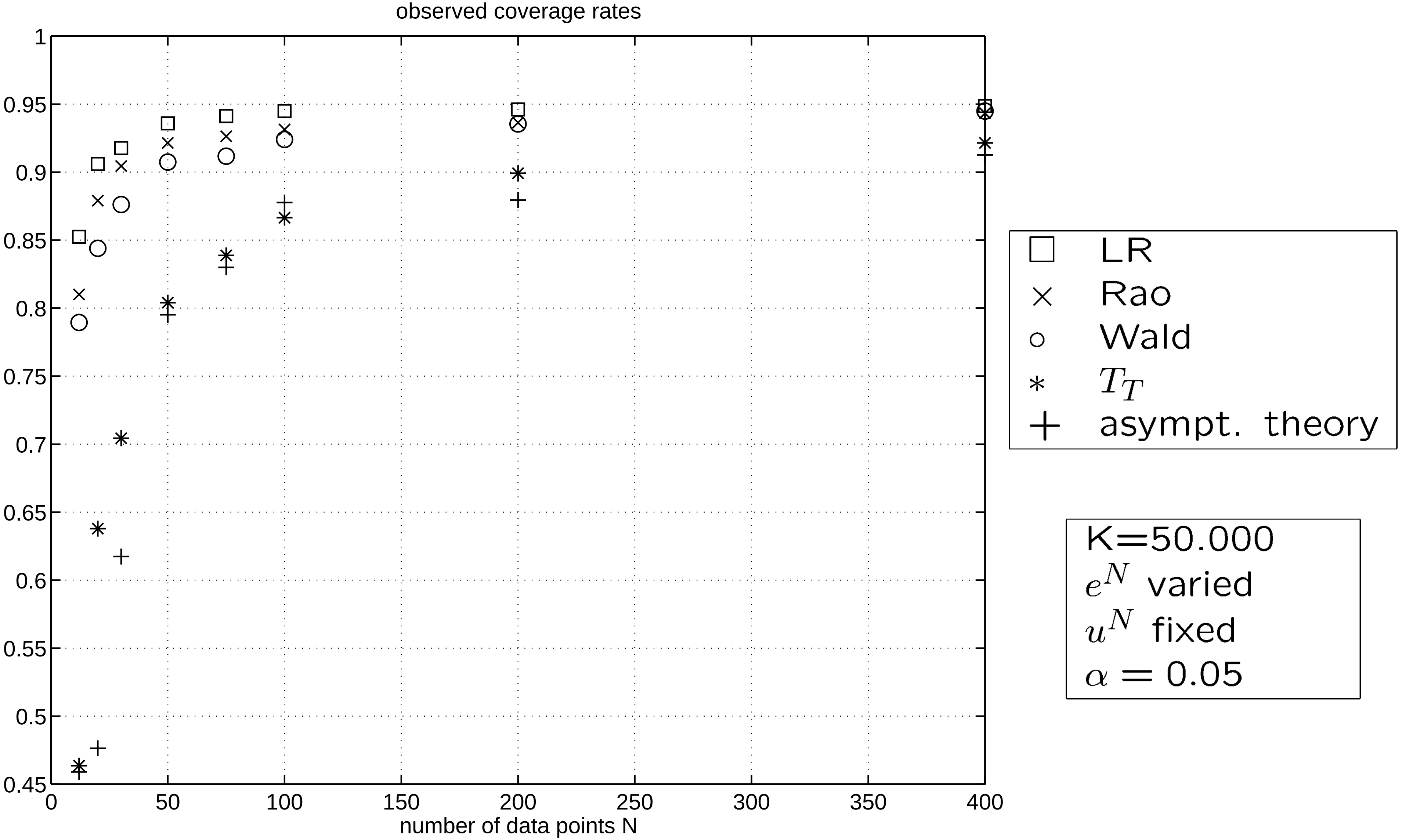
50.000 data sets (y^N, u^N)

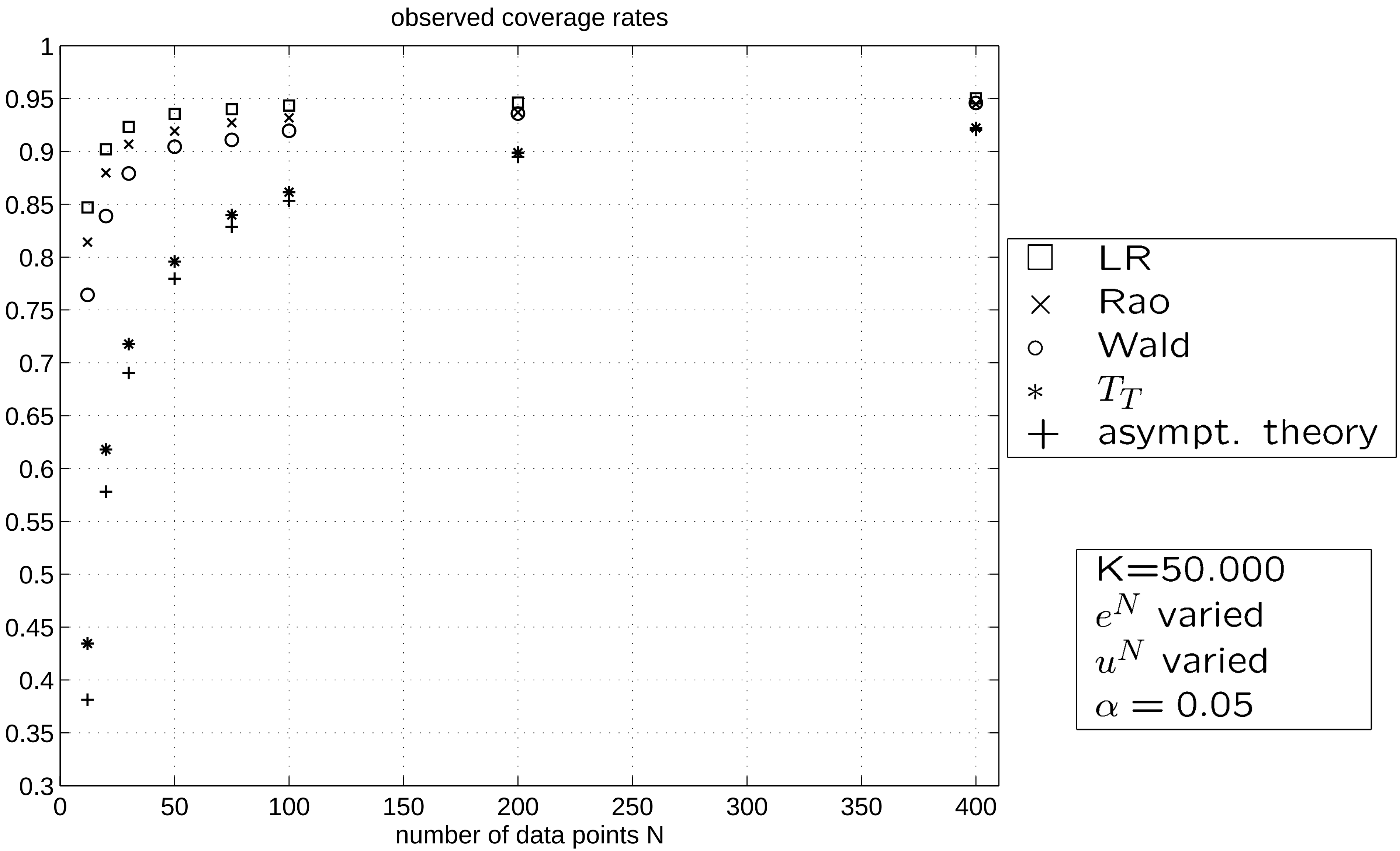
Data generating system $\mathcal{S}$:

$$y(t) - 1.5578y(t-1) + 0.5769y(t-2) = \\ 0.1047u(t-1) + 0.0872u(t-2) + e(t)$$

u^N (known) and e^N (unknown) realizations of Gaussian white noise proces, $\sigma_u^2 = 1, \sigma_e^2 = 0.5$.

For each data set, the model was identified and it was recorded if the constructed confidence regions contained θ_0.


observed coverage rates
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0
50
100
150
200
250
300
350
400
number of data points N
LR
Rao
Wald
T_T
asympt. theory
K=50.000
e^N varied
u^N fixed
$\alpha = 0.05$


observed coverage rates
number of data points N
LR
Rao
Wald
T_T
asympt. theory
K=50.000
e^N varied
u^N varied
$\alpha = 0.05$

Conclusions

$100(1-\alpha)\%$ confidence regions for ARX:

$$\left\{\theta | (\theta - \widehat{\theta}_N)^T X (\theta - \widehat{\theta}_N) \leq \sigma^2 c_\chi(n, \alpha)\right\}$$

- Confidence regions that incorporate information on the particular noise realization e^N in X (via Φ or $\widehat{\theta}_N$) are most reliable.

- Confidence region based on LR test statistic ($X = \Phi^T\Phi$) performs best (in simulations).

ARX modelling

$$\varphi(t,\theta_0) = s_u(t,\theta_0) + s_e(t,\theta_0),$$

$$s_u(t,\theta_0) = \frac{\Lambda_G(q^{-1},\theta_0)}{H(q^{-1},\theta_0)}u(t); \qquad s_e(t,\theta_0) = \frac{\Lambda_H(q^{-1},\theta_0)}{H(q^{-1},\theta_0)}e(t),$$

with $\Lambda_G(q^{-1},\theta)$ and $\Lambda_H(q^{-1},\theta)$ the $n \times 1$ gradient vectors of the transfer function $G(q^{-1},\theta)$ and $H(q^{-1},\theta)$ with respect to θ, respectively.

$$R(\theta_0) = \sum_{t=1}^{N} s_u(t,\theta_0)s_u^T(t,\theta_0) + \sum_{t=1}^{N} \mathbb{E}\left[s_e(t,\theta_0)s_e^T(t,\theta_0)\right]$$

Result from asymptotic theory (θ_0 known)

Asymptotically valid expression for $cov(\widehat{\theta}_N)$:

$$P_{\theta_0} = \frac{\sigma^2}{2\pi N}\left(\int_{-\pi}^{\pi}\left(\Gamma_G(e^{i\omega},\theta_0)\Phi_u(\omega) + \Gamma_H(e^{i\omega},\theta_0)\sigma^2\right)d\omega\right)^{-1}$$

$$\Gamma_G(e^{i\omega},\theta_0) = \frac{\Lambda_G(e^{i\omega},\theta_0)\Lambda_G^*(e^{i\omega},\theta_0)}{H(e^{i\omega},\theta_0)H^*(e^{i\omega},\theta_0)},$$

$$\Gamma_H(e^{i\omega},\theta_0) = \frac{\Lambda_H(e^{i\omega},\theta_0)\Lambda_H^*(e^{i\omega},\theta_0)}{H(e^{i\omega},\theta_0)H^*(e^{i\omega},\theta_0)}$$

with $\Phi_u(\omega)$ the power spectrum of $u(t)$.

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