UPPSALA UNIVERSITET

Informationsteknologi

Errors-in-variables methods in system identification

T. Söderström

Department of Information Technology

Uppsala University, Sweden

EIV in system identification

Work done in collaboration (discussions and joint publications) with

Juan Carlos Agüero, Theodore Anderson, Mats Cedervall, Manfred Deistler, Roberto Diversi, Mats Ekman, Graham Goodwin, Roberto Guidorzi, Christiaan Heij, Mei Hong, Erlendur Karlsson, Alexander Kukush, Erik K. Larsson, Kaushik Mahata, Ivan Markovsky, Magnus Mossberg, Rik Pintelon, Wolfgang Scherrer, Johan Schoukens, Virginija Šimonyte, Joachim Sorelius, Umberto Soverini, Petre Stoica, Sabine Van Huffel, Kiyoshi Wada, Wei Xing Zheng.

UPPSALA UNIVERSITET

Informationsteknologi

Contents

- Background and motivation

Contents

- Background and motivation
  - Problem formulation

UPPSALA UNIVERSITET

Informationsteknologi

Contents

- Background and motivation
  - Problem formulation
  - Line fitting

UPPSALA UNIVERSITET

Informationsteknologi

Contents

- Background and motivation
  - Problem formulation
  - Line fitting
- Identifiability

UPPSALA UNIVERSITET

Informationsteknologi

Contents

- Background and motivation
  - Problem formulation
  - Line fitting
- Identifiability
- Estimators

UPPSALA UNIVERSITET

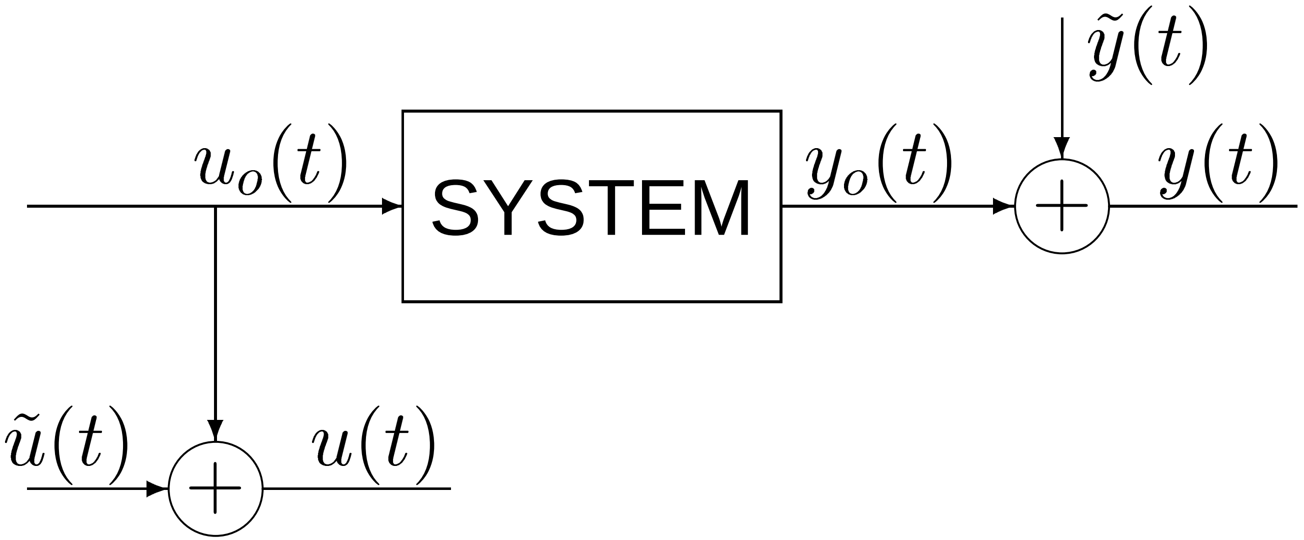
Informationsteknologi

Contents

UPPSALA UNIVERSITET

Informationsteknologi

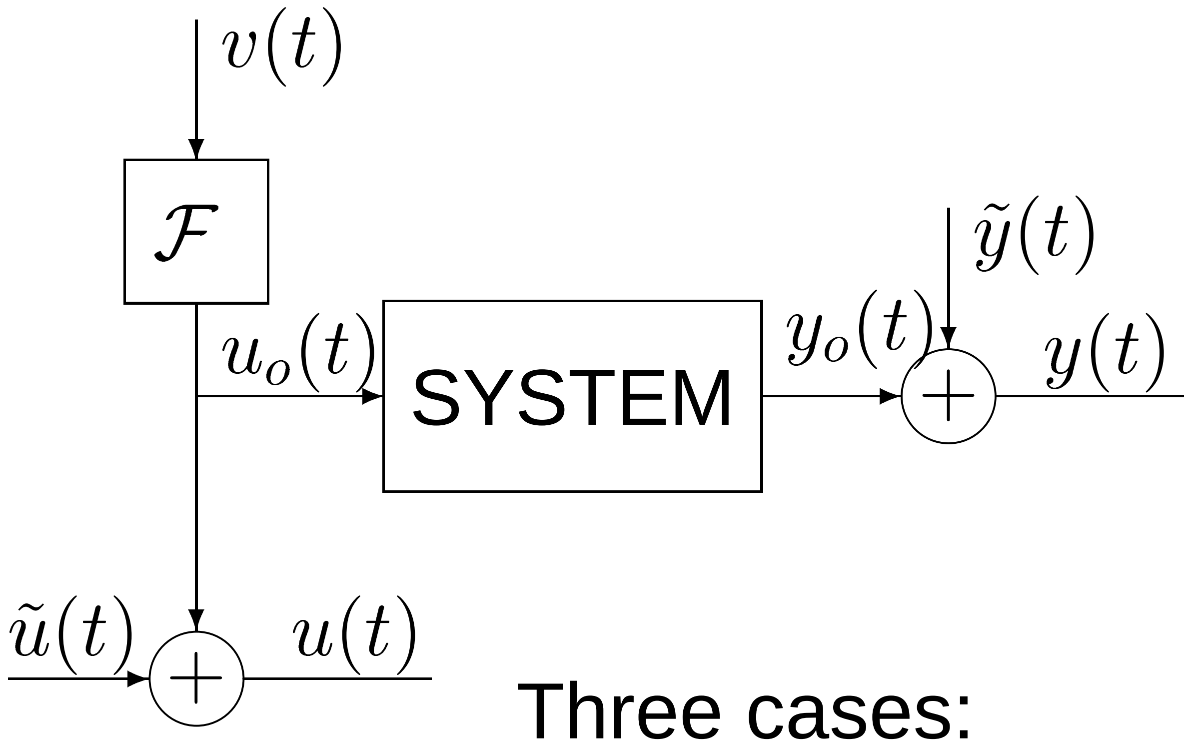
Problem formulation



$\tilde{u}(t)$, $\tilde{y}(t)$ measurement noise.

Determine the system transfer function.

Problem formulation EIV cont'd



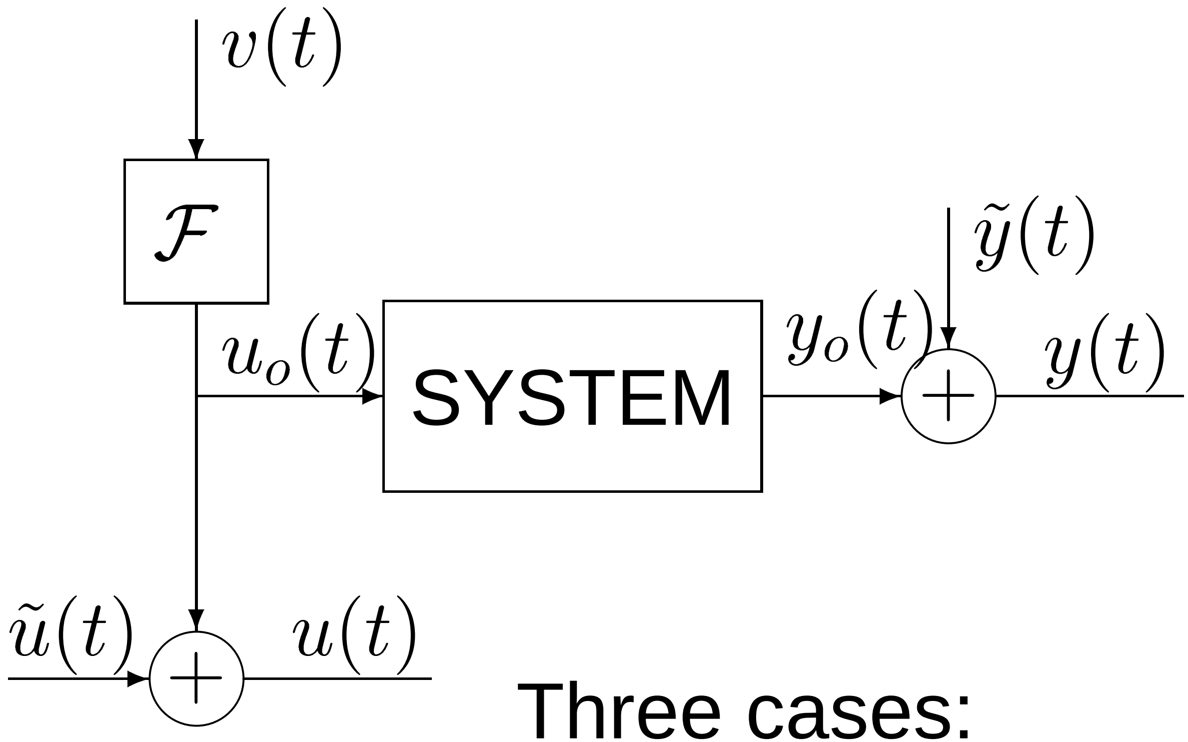
Three cases:

- v and $\mathcal{F}$ unknown: True EIV situation

UPPSALA UNIVERSITET

Informationsteknologi

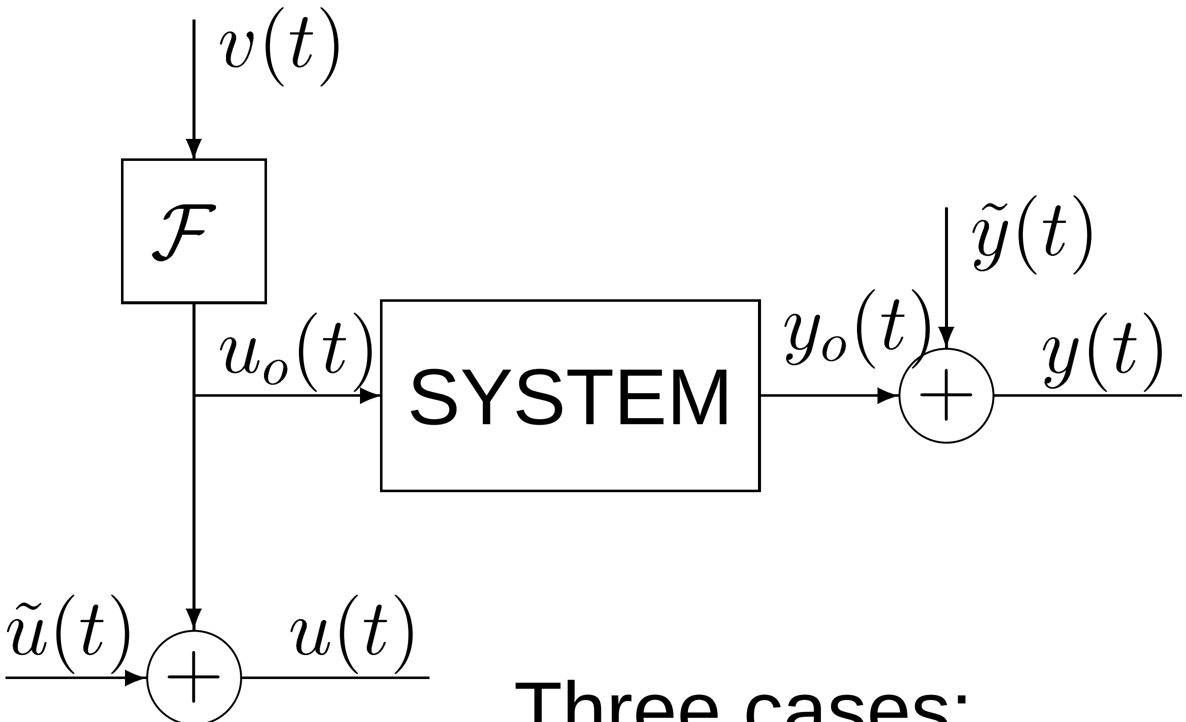
Problem formulation EIV cont'd



Three cases:

- v and $\mathcal{F}$ unknown: True EIV situation
- v under control, $\mathcal{F}$ unknown (repeated exp.)

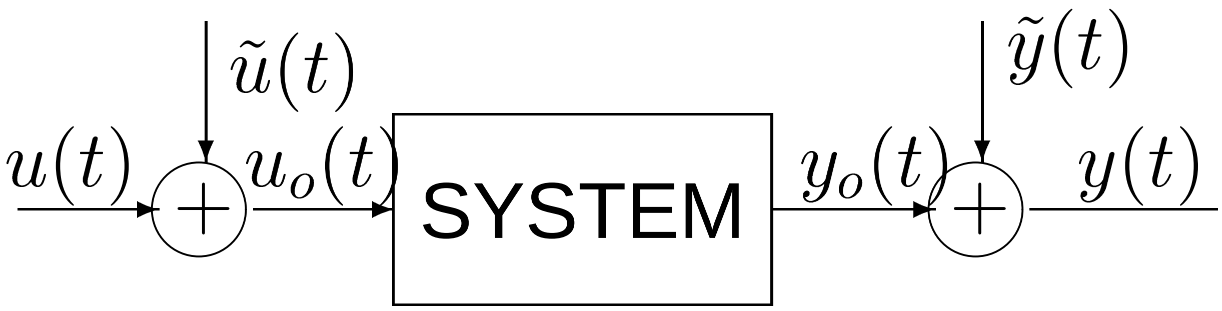
Problem formulation EIV cont'd



Three cases:

- v and $\mathcal{F}$ unknown: True EIV situation
- v under control, $\mathcal{F}$ unknown (repeated exp.)
- v new control variable, not an EIV problem

A related case



This is not an EIV problem! Why?

- $\tilde{u}(t)$ effects $y(t)$ [process noise!]
- $u(t)$ and $u_o(t)$ influence $y_o(t)$ in the same way

UPPSALA UNIVERSITET

Informationsteknologi

Motivations

- Understand the underlying relations (rather than make a good prediction from noisy data). [The ‘classical’ motivation in e.g. econometrics]
- Approximate a high-dimensional data vector by a small number of factors. [The standard motivation for factor analysis]
- Lack of enough information to classify the available signals into inputs and outputs; use a ‘symmetric’ system model. [Cf. the behavioral approach to modeling]

UPPSALA UNIVERSITET

Informationsteknologi

Line fitting

Assume that we have a set of points in the x, y plane, that corresponds to noisy measurements $(x_1, y_1), \ldots (x_n, y_n)$.

Model

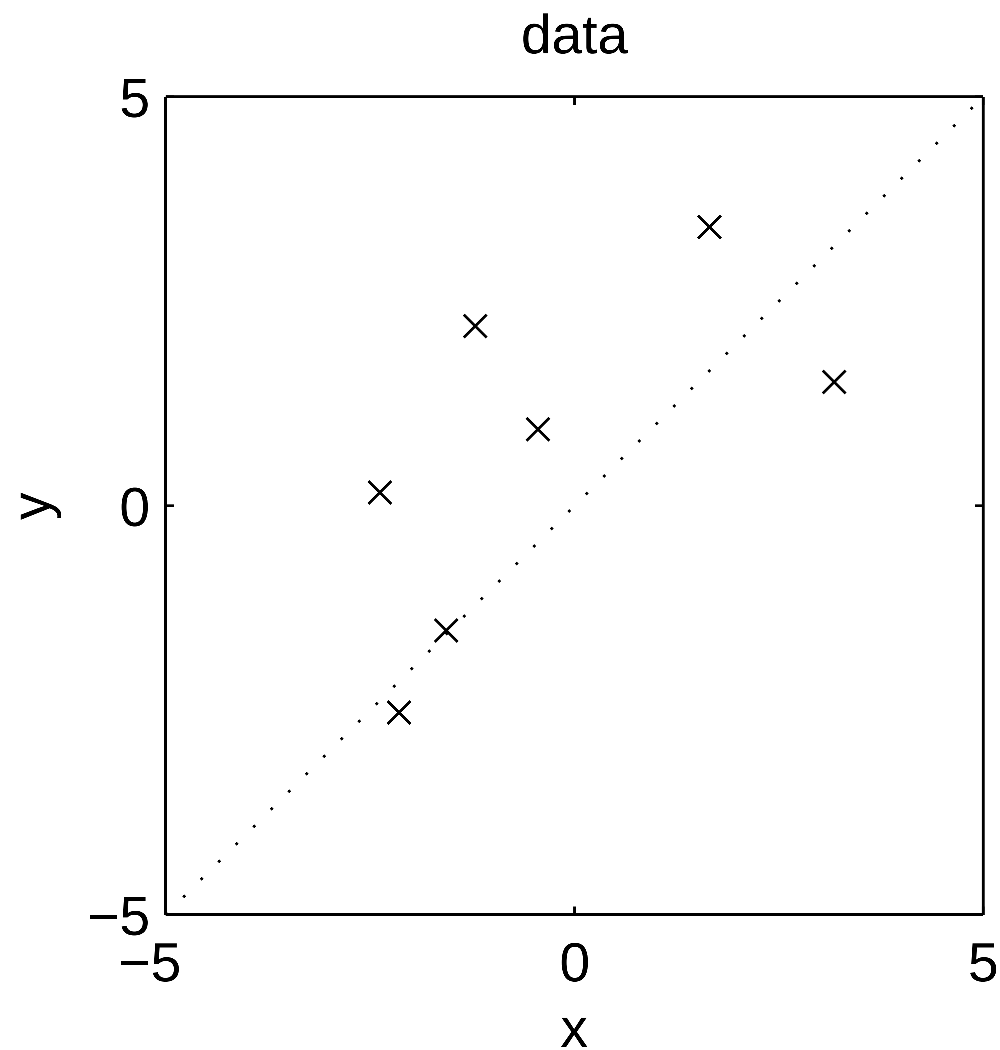
$$\begin{aligned} y_i &= y_{oi} + \tilde{y}_i, \\ x_i &= x_{oi} + \tilde{x}_i, \qquad\qquad i = 1, \ldots, n. \\ y_{oi} &= a_o x_{oi} + b_o, \end{aligned}$$

The measurement errors $\{\tilde{y}_i\}$ and $\{\tilde{x}_i\}$: independent random variables of zero mean and variances λ_y and λ_x, respectively.

UPPSALA UNIVERSITET

Informationsteknologi

Line fitting, cont’d



UPPSALA UNIVERSITET

Informationsteknologi

Line fitting, identifiability analysis

Use first and second order moments. Assume $E(x_{oi}) = m, \;\; \text{var}(x_{oi}) = \sigma^2$.

$$
\begin{array}{ll}
E(x) = m & \text{5 equations} \\
E(y) = am + b & \text{6 unknowns :} \\
\text{var}(x) = \sigma^2 + \lambda_x & a, \; b, \\
\text{var}(y) = a^2\sigma^2 + \lambda_y & m, \; \sigma^2, \; \lambda_x, \; \lambda_y. \\
\text{cov}(x, y) = a\sigma^2 &
\end{array}
$$

No unique solution! Unknown uncertainties in **both** x_i and y_i makes the problem difficult.

UPPSALA UNIVERSITET

Informationsteknologi

Line fitting, cont'd

Maximum likelihood estimation of
$a,\ b,\ \lambda_y,\ \lambda_x,\ \{x_{oi}\}$:

The ML estimate does not exist! (The likelihood function $\to \infty$ for finite parameter values).

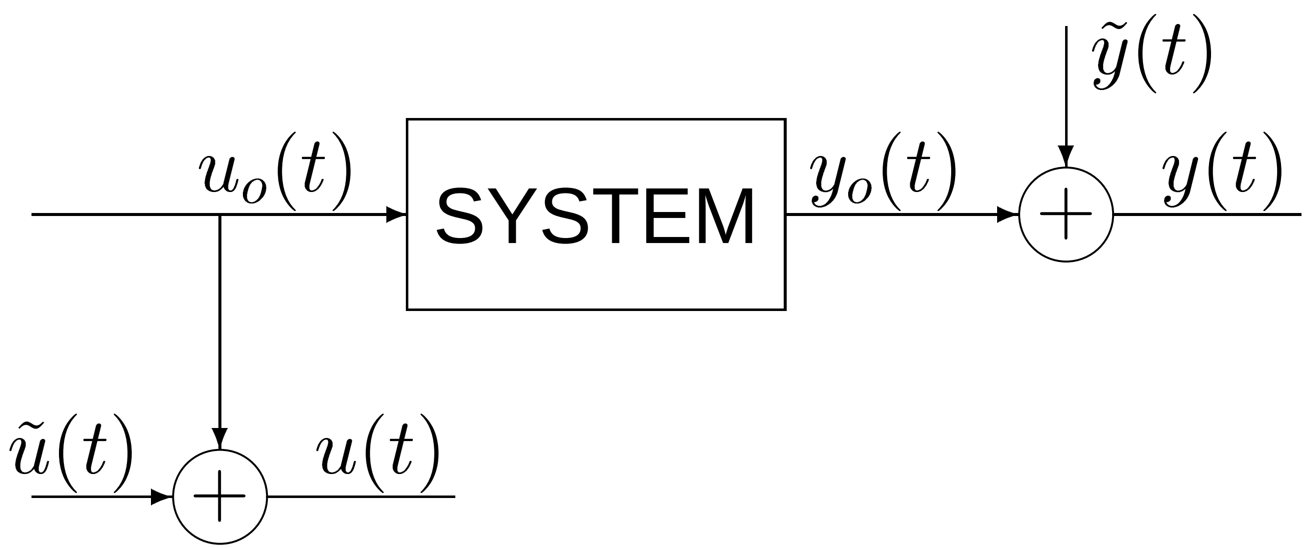
Assume λ_y/λ_x known: Then $\hat{\theta}_{\mathrm{ML}}$ is feasible ($\hat{a}_{\mathrm{ML}}$ and $\hat{b}_{\mathrm{ML}}$ are consistent).

UPPSALA UNIVERSITET

Informationsteknologi

Contents

Problem formulation EIV cont'd



Given noisy data $y(1), u(1), \ldots, y(N), u(N)$, determine the system transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}.$$

UPPSALA UNIVERSITET

Informationsteknologi

Assumptions

The available signals are time-discrete

$$\begin{aligned} u(t) &= u_o(t) + \tilde{u}(t), \\ y(t) &= y_o(t) + \tilde{y}(t). \end{aligned}$$

AS1. The system is linear [causal] and asymptotically stable. □

AN1. $\tilde{u}(t)$, $\tilde{y}(t)$ are uncorrelated stationary processes, with zero means and spectra $\phi_{\tilde{u}}(\omega)$ and $\phi_{\tilde{y}}(\omega)$, respectively. □

AI1. $u_o(t)$ is p.e. and uncorrelated with $\tilde{u}(t)$ and $\tilde{y}(t)$. □

Identifiability nonparametric models

Use second order statistics of $z(t) = (y(t)\ u(t))^\top$:

$$\begin{aligned} \phi_z &= \begin{pmatrix} GG^* & G \\ G^* & 1 \end{pmatrix} \phi_{u_o} + \begin{pmatrix} \phi_{\tilde{y}} & 0 \\ 0 & \phi_{\tilde{u}} \end{pmatrix} \\ &= \begin{pmatrix} \hat{G}\hat{G}^* & \hat{G} \\ \hat{G}^* & 1 \end{pmatrix} \hat{\phi}_{u_o} + \begin{pmatrix} \hat{\phi}_{\tilde{y}} & 0 \\ 0 & \hat{\phi}_{\tilde{u}} \end{pmatrix}. \end{aligned}$$

Note that for each frequency there are 3 equations with 4 unknowns. There is hence one degree of freedom (for each frequency) in the solution.

How to handle the lack of identifiability?

At least four options

1. ‘Accept’ the status. Do not make further assumptions. Instead of looking for a unique estimate, deal with the whole set of estimates. [Set membership estimation]
2. Impose more detailed models of $u_o(t)$, $\tilde{u}(t)$, $\tilde{y}(t)$, say ARMA processes of specified orders.

Identifiability, cont'd

3. Modify at least one of the assumptions **AN2**, **AI2** on Gaussian distributed data. Use higher order statistics to gain additional information. Deistler(1986), Tugnait(1992).
   [time-consuming; may not lead to accurate estimates]

UPPSALA UNIVERSITET

Informationsteknologi

Identifiability, cont'd

4. Use more than one experiment. [Assume the user can control the signal $v(t)$]

- $\phi_{u_o}(\omega)$ differs between the different experiments,

  or

- $u_o(t)$ is (well) correlated between experiments, but $\tilde{y}(t),\ \tilde{u}(t)$ are uncorrelated between experiments.

UPPSALA UNIVERSITET

Informationsteknologi

Identifiability parametric models

Model $\tilde{u}(t)$, $\tilde{y}(t)$, $u_o(t)$ as ARMA processes, and analyze identifiability.

- **AN3a.** Both $\tilde{y}(t)$ and $\tilde{u}(t)$ are ARMA processes. Agüero et al(2005,2006), Nowak(1985,1992), Castaldi-Soverini(1996).
- **AN3b.** $\tilde{y}(t)$ is an ARMA process, while $\tilde{u}(t)$ is white. Söderström(1980), Solo(1986).

UPPSALA UNIVERSITET

Informationsteknologi

Identifiability parametric models, cont'd

- **AN3c.** Both $\tilde{y}(t)$ and $\tilde{u}(t)$ are white noise sequences. Castaldi et al(1996), Söderström(2003), Stoica-Nehorai(1987). [less realistic]

- Generalization to the multivariate case Nowak(1992).

Contents

- Background and motivation
- Identifiability
- Estimators (omit methods for periodic data)
  - Least squares (LS), Instrumental variables (IV), Bias-compensated LS (BCLS), The Frisch scheme, Total least squares (TLS)
  - Frequency domain methods, Prediction error method (PEM) and maximum likelihood (ML) method
  - Accuracy aspects
- Comparisons and conclusions

UPPSALA UNIVERSITET

Informationsteknologi

Notations for parametric estimators

AS5. The system is described as

$$A(q^{-1})y_o(t) = B(q^{-1})u_o(t),$$

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}, \\ B(q^{-1}) &= b_1 + \cdots + b_{nb} q^{-nb+1}. \end{aligned}$$

Parameter vector θ and regressor vector $\varphi(t)$:

$$\begin{aligned} \theta &= (a_1 \ldots \ a_{na} \ b_1 \ldots \ b_{nb})^\top, \\ \varphi(t) &= (-y(t-1) \ldots \ -y(t-na) \\ &\qquad u(t) \ldots \ u(t-nb+1))^\top. \end{aligned}$$

UPPSALA UNIVERSITET

Informationsteknologi

Notations for parametric estimators, cont'd

System description

$$\left.\begin{aligned} & A(q^{-1})y(t) - B(q^{-1})u(t) \\ = \; & A(q^{-1})y_o(t) - B(q^{-1})u_o(t) \\ + \; & A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned}\right. \quad \begin{aligned} &\Big\} = 0 \\ &\Big\} \triangleq \varepsilon(t) \end{aligned}$$

Hence, the system can be written as a linear regression

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t).$$

UPPSALA UNIVERSITET

Informationsteknologi

Notations for parametric estimators, cont'd

Denote covariance matrices and their estimates as

$$R_\varphi = E[\varphi(t)\varphi^\top(t)], \quad \hat{R}_\varphi = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^\top(t).$$

Conventions:

- θ_o denotes the true parameter vector, and $\hat{\theta}$ denotes its estimate.
- $\varphi_o(t)$ denotes the noise-free part of the regressor vector.
- $\tilde{\varphi}(t)$ denotes the noise-contribution to the regressor vector.

The least squares estimate is biased

Model

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t).$$

Assume $\tilde{u}(t)$ and $\tilde{y}(t)$ are white, **(AN3c)**.

The least squares (LS) estimate

$$\begin{aligned}\hat{\theta}_{\mathrm{LS}} &= \hat{R}_\varphi^{-1}\hat{r}_{\varphi y} \to R_\varphi^{-1} r_{\varphi y}, \qquad N \to \infty \\ &= (R_{\varphi_o} + R_{\tilde{\varphi}})^{-1} r_{\varphi_o y_o} = (R_{\varphi_o} + R_{\tilde{\varphi}})^{-1} R_{\varphi_o}\theta_o\end{aligned}$$

Bias due to $R_{\tilde{\varphi}}$.

UPPSALA UNIVERSITET

Informationsteknologi

Instrumental variable methods

The IV estimate can be defined as

$$\left(\frac{1}{N}\sum_{t=1}^{N} z(t)\varphi^{\top}(t)\right)\hat{\theta}_{\mathrm{IV}} = \left(\frac{1}{N}\sum_{t=1}^{N} z(t)y(t)\right).$$

If $\dim(z) > \dim(\varphi)$, solve the equations in a (weighted) least squares sense.

UPPSALA UNIVERSITET

Informationsteknologi

Instrumental variable methods, properties

- Applicable under fairly general noise conditions, **AN3b**.
- Inexpensive from a computational point of view.
- Poor accuracy of $\hat{\theta}$ is often obtained.
- The matrix $R_{z\varphi}$ has to be full rank [a p.e. like condition on $u_o(t)$].

UPPSALA UNIVERSITET

Informationsteknologi

EIV issues

== error**s** == error**s**

$$\text{iniput } u(t),\ \text{outplut } y(t)$$

$u(t),\ y(t)$ **errors** $\tilde{u}(t),\ tildey(t)$

UPPSALA UNIVERSITET

Informationsteknologi

Bias-compensating least squares, BCLS

Idea: Find equations for determining λ_u and λ_y and modify the normal equations to

$$(\hat{R}_\varphi - \underbrace{\begin{pmatrix} \hat{\lambda}_y I_{na} & 0 \\ 0 & \hat{\lambda}_u I_{nb} \end{pmatrix}}_{\text{compensation}})\hat{\theta} = \hat{r}_{\varphi y}$$

Many possibilities exist.

Nonlinear equations with structure (often bilinear equations). Hence iterative schemes are necessary.

BCLS, cont'd

There are many variants

- $\tilde{u}(t), \tilde{y}(t)$ may be white or ARMA
- Different additional equations
- Different algorithms for solving the equations

Zheng(1998,1999,2002), Wada et al(1990), Jia et al(2001), Ikenoue et al(2005), Ekman(2005), Ekman et al(2006).

UPPSALA UNIVERSITET

Informationsteknologi

BCLS, cont'd

Some possibilities for additional equations:

- Minimal LS loss
- LS estimates for an extended model
- Residual covariance function

Some possibilities for algorithms (note equations are often bilinear!)

- Relaxation algorithms (solve repeatedly linear equations)
- Variable projection algorithms $\rightarrow$ low dimensional optimization problem

UPPSALA
UNIVERSITET

Informationsteknologi

The Frisch scheme

Links to static case: Beghelli et al(1990), Scherrer-Deistler(1998).

Aspects for dynamic models: Guidorzi(1996), Söderström et al(2002), Diversi et al(2003).

Can be seen as a special form of BCLS!

UPPSALA UNIVERSITET

Informationsteknologi

The Frisch scheme, notations

Extended parameter vector

$$\overline{\theta} = (1 \;\; a_1 \ldots \; a_{na} \; b_1 \ldots \; b_{nb})^\top .$$

Extended regressor vector

$$\begin{aligned} \overline{\varphi}(t) &= (-y(t) \ldots \; -y(t-na) \\ &\qquad u(t) \ldots \; u(t-nb+1))^\top \\ &= (-y(t) \;\; \varphi^\top(t))^\top \\ &= (-\overline{\varphi}_y^\top(t) \;\; \varphi_u^\top(t))^\top . \end{aligned}$$

The Frisch scheme, cont'd

Note that $R_{\overline{\varphi}_o}$ is singular.

Assume that some estimate $\hat{\lambda}_u$ is available. Then determine $\hat{\lambda}_y$ so that

$$\left(\hat{R}_{\overline{\varphi}} - \hat{R}_{\tilde{\overline{\varphi}}}\right) = \begin{pmatrix} \hat{R}_{\overline{\varphi}_y} - \hat{\lambda}_y I_{na+1} & \hat{R}_{\overline{\varphi}_y \varphi_u} \\ \hat{R}_{\varphi_u \overline{\varphi}_y} & \hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \end{pmatrix}$$

is singular.

Hence, $\hat{\lambda}_y = \hat{\lambda}_y(\hat{\lambda}_u)$.

The Frisch scheme, cont'd

The estimate of the parameter vector θ is determined by solving

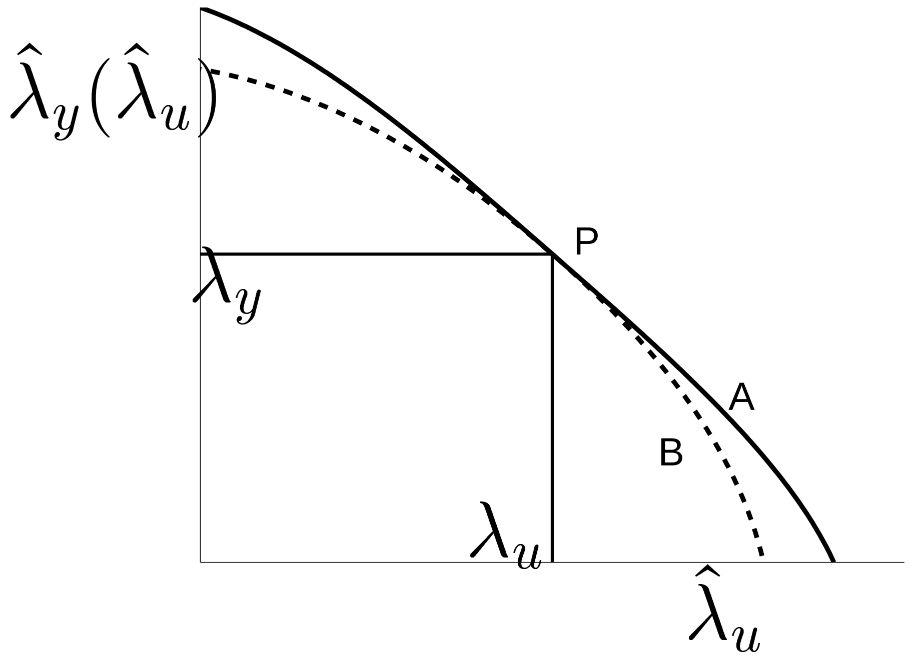
$$\left(\hat{R}_{\varphi} - \begin{pmatrix} \hat{\lambda}_y I_{na} & 0 \\ 0 & \hat{\lambda}_u I_{nb} \end{pmatrix} \right) \hat{\theta} = \hat{r}_{\varphi y},$$

which is indeed the BCLS equations.

What remains is to determine $\hat{\lambda}_u$. Different alternatives have been proposed:

The Frisch scheme, example

The function $\hat{\lambda}_y(\hat{\lambda}_u)$ is evaluated both for the nominal model and for an extended model, Beghelli et al(1990).



UPPSALA UNIVERSITET

Informationsteknologi

Total least squares, TLS

Consider the overdetermined system of equations

$$Ax \approx b.$$

The least squares solution is

$$\hat{x}_{\mathrm{LS}} = (A^\top A)^{-1} A^\top b,$$

and solves the optimization problem

$$\min \parallel \Delta b \parallel^2 \text{ subject to } A\hat{x}_{\mathrm{LS}} = b + \Delta b.$$

UPPSALA UNIVERSITET

Informationsteknologi

Total least squares, cont'd

The TLS problem can be formulated as,

$$\min \| \, [\Delta A \ \Delta b] \, \|_F^2 \text{ s. t. } (A + \Delta A)\hat{x}_{\mathrm{TLS}} = b + \Delta b.$$

The TLS solution gives the ML estimate, *if* the errors in the A and b elements are independent and identically distributed, Gleser(1981).

Is this helpful?

Total least squares, cont'd

For a linear regression model, $t = 1, \ldots, N$,

$$\begin{pmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{pmatrix} \theta = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}.$$

The matrix is block Toeplitz (equal elements along the diagonals). The structured TLS (STLS) solution is more relevant than the basic TLS solution in general.

Total least squares, cont’d

The STLS leads to numerical optimization.

The statistical properties of the solution to a structured TLS problem is considered in several papers, e.g. Kukush et al(2005).

Common assumtion: Either λ_y/λ_u known, or $u_o(t)$ changes character (i.e. more than one experiment).

UPPSALA UNIVERSITET

Informationsteknologi

Frequency domain methods 1

The spectral density of the input-output data satisfies

$$\phi_z - \begin{pmatrix} \lambda_y & 0 \\ 0 & \lambda_u \end{pmatrix} = \begin{pmatrix} G \\ 1 \end{pmatrix} \begin{pmatrix} G^* & 1 \end{pmatrix} \phi_{u_o}.$$

Both sides are singular. It must hold for each frequency $\omega_k, k = 1, 2, \ldots,$ that

$$[\phi_y(\omega_k) - \lambda_y][\phi_u(\omega_k) - \lambda_u] - |\phi_{yu}(\omega_k)|^2 = 0.$$

UPPSALA UNIVERSITET

Informationsteknologi

Frequency domain methods 1, cont'd

This relation is exploited as a linear regression with $\lambda_y, \ \lambda_u, \ \lambda_y \lambda_u$ as three unknowns, to derive an estimate of the noise variances.

Once estimates of λ_y and λ_u are available, it is straightforward to estimate $G(\mathrm{e}^{\mathrm{i}\omega_k})$, for example as

$$\hat{G}(\mathrm{e}^{\mathrm{i}\omega_k}) = \phi_{yu}(\omega_k)/[\phi_u(\omega_k) - \hat{\lambda}_u].$$

Beghelli et al(1997), Söderström et al(2003).

Frequency domain methods 2

Sample maximum likelihood (SML), Schoukens et al(1997).
Periodic data, at least four periods.

Step 1. Estimate $\sigma_u^2(\omega), \sigma_y^2(\omega), \sigma_{yu}^2(\omega)$.

Step 2. Estimate A and B by minimizing

$$V_{\text{SML}} = \frac{1}{N} \sum_{k=1}^{N} \frac{\left| B(\mathrm{e}^{\mathrm{i}\omega_k}, \theta) U(\omega_k) - A(\mathrm{e}^{\mathrm{i}\omega_k}, \theta) Y(\omega_k) \right|^2}{D(\omega_k)}$$

$$\begin{aligned} D(\omega) &= \sigma_u^2(\omega) |B(\mathrm{e}^{\mathrm{i}\omega}, \theta)|^2 + \sigma_y^2(\omega) |A(\mathrm{e}^{\mathrm{i}\omega}, \theta)|^2 \\ &\quad -2 \mathrm{Re} \left[\sigma_{yu}^2(\omega) A(\mathrm{e}^{\mathrm{i}\omega}, \theta) B(\mathrm{e}^{-\mathrm{i}\omega}, \theta) \right] \end{aligned}$$

PEM and maximum likelihood

Model noise and noise-free input as well as the system. Example with $\tilde{y}(t),\ \tilde{u}(t)$ white:

$$z(t) = \begin{pmatrix} y(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} \frac{B(q^{-1})C(q^{-1})}{A(q^{-1})D(q^{-1})} & 1 & 0 \\ \frac{C(q^{-1})}{D(q^{-1})} & 0 & 1 \end{pmatrix} \begin{pmatrix} e(t) \\ \tilde{y}(t) \\ \tilde{u}(t) \end{pmatrix}.$$

Prediction error (PEM) and maximum likelihood (ML) estimates:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta).$$

UPPSALA UNIVERSITET

Informationsteknologi

PEM and ML, cont'd

Prediction errors $\varepsilon(t,\theta) = z(t) - \hat{z}(t|t-1;\theta)$.

PEM estimate

$$V_N(\theta) = \det\left(\frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\theta)\varepsilon^\top(t,\theta)\right).$$

UPPSALA UNIVERSITET

Informationsteknologi

PEM and ML, cont'd

ML estimate

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \ell(\varepsilon(t,\theta),\theta,t),$$

with

$$\begin{aligned}\ell(\varepsilon,\theta,t) &= \frac{1}{2}\log\det Q(\theta) + \frac{1}{2}\varepsilon^\top(t,\theta)Q^{-1}(\theta)\varepsilon(t,\theta), \\ Q(\theta) &= E\varepsilon(t,\theta)\varepsilon^\top(t,\theta).\end{aligned}$$

UPPSALA UNIVERSITET

Informationsteknologi

PEM and ML, cont'd

The ML estimate can alternatively be computed in the frequency domain, Pintelon-Schoukens(2005), [some differences in how transient effects are handled]

The inherent spectral factorization is somewhat easier to carry out in the frequency domain.

PEM and ML, cont'd

General properties:

- (Very) high accuracy.
- The numerical optimization procedure is, in general, quite complex.
- The procedure may fail to give good results if only poor initial parameter estimates are available.

How good can the estimates be?

The asymptotic distribution of $\hat{\theta}$ is known in many cases

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{\text{dist}} \mathbf{N}(0, P),$$

The covariance matrix P depends on

- the method (and the user parameters),

UPPSALA
UNIVERSITET

How good can the estimates be?

The asymptotic distribution of $\hat{\theta}$ is known in many cases

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{\text{dist}} \mathbf{N}(0, P),$$

The covariance matrix P depends on

- the method (and the user parameters),
- the system,

UPPSALA UNIVERSITET

Informationsteknologi

How good can the estimates be?

The asymptotic distribution of $\hat{\theta}$ is known in many cases

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{\text{dist}} \mathbf{N}(0, P),$$

The covariance matrix P depends on

- the method (and the user parameters),
- the system,
- the dynamics for $u_o(t),\ \tilde{u}(t),\ \tilde{y}(t)$.

How good can the estimates be?, cont'd

Example of results

- Instrumental variable (IV) methods, Söderström-Stoica(1983,1989).
- Bias-compensating least squares (BCLS), Hong et al (2006).
- The Frisch scheme, Söderström(2005).
- Prediction error method and maximum likelihood method, Ljung(1999), Söderström(2006).

UPPSALA UNIVERSITET

Informationsteknologi

How good can the estimates be?, cont'd

The Cramér-Rao lower bound P_{CRLB} gives a lower bound for the covariance matrix of any unbiased parameter estimates.

$$\mathrm{cov}(\hat{\theta} - \theta_o) \geq P_{\mathrm{CRLB}} = J^{-1},$$

$$J = E\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^{\top}\left(\frac{\partial \log L(\theta)}{\partial \theta}\right),$$

where $L(\theta)$ is the likelihood function. The matrix J is the Fisher information matrix.

UPPSALA UNIVERSITET

Informationsteknologi

How good can the estimates be?, cont'd

Algorithms exist for computing P_{CRLB}, Söderström(2006).

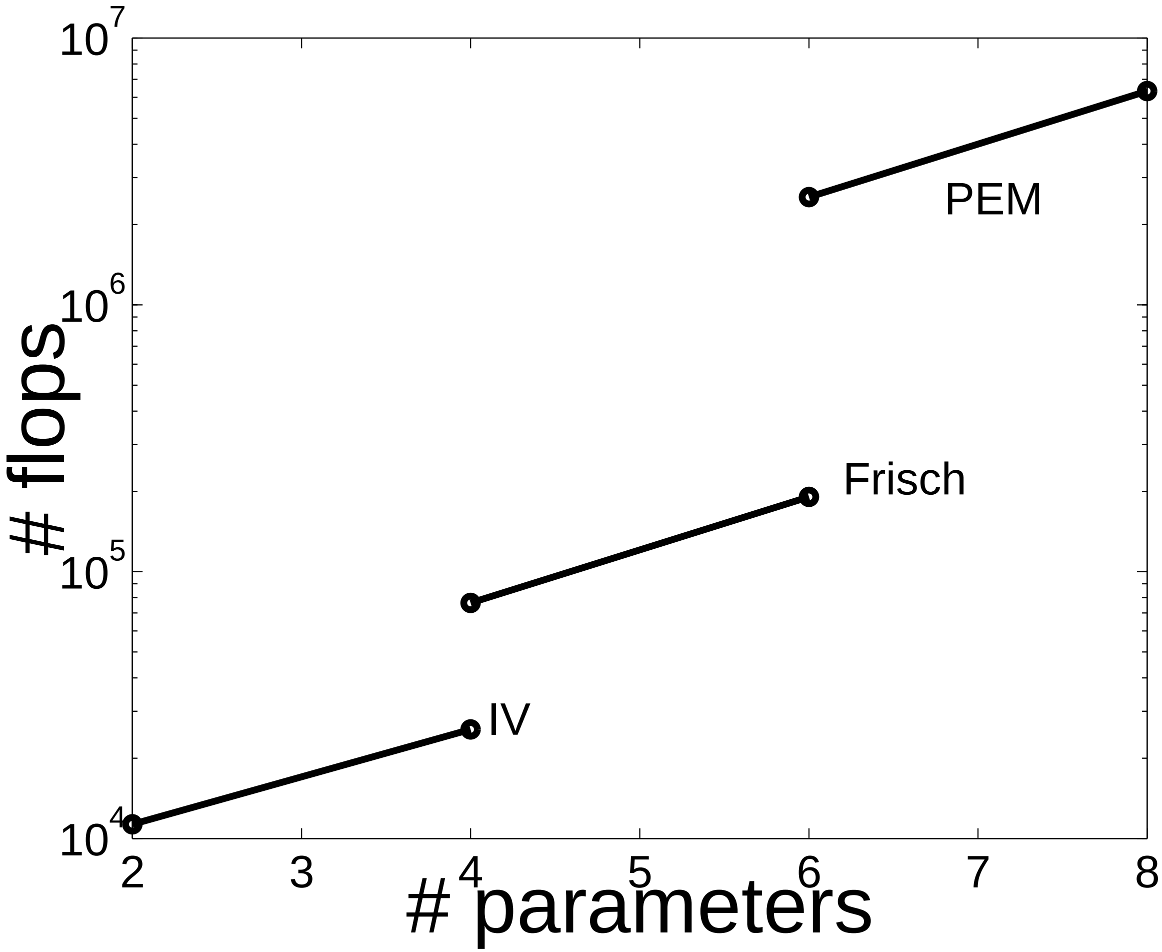
- Assumptions on parameterization of the dynamics for $u_o(t),\ \tilde{u}(t),\ \tilde{y}(t)$ are needed.
- $P \geq P_{\mathrm{CRLB}}$
- $P_{\mathrm{ML}} = P_{\mathrm{CRLB}}$

UPPSALA UNIVERSITET

Informationsteknologi

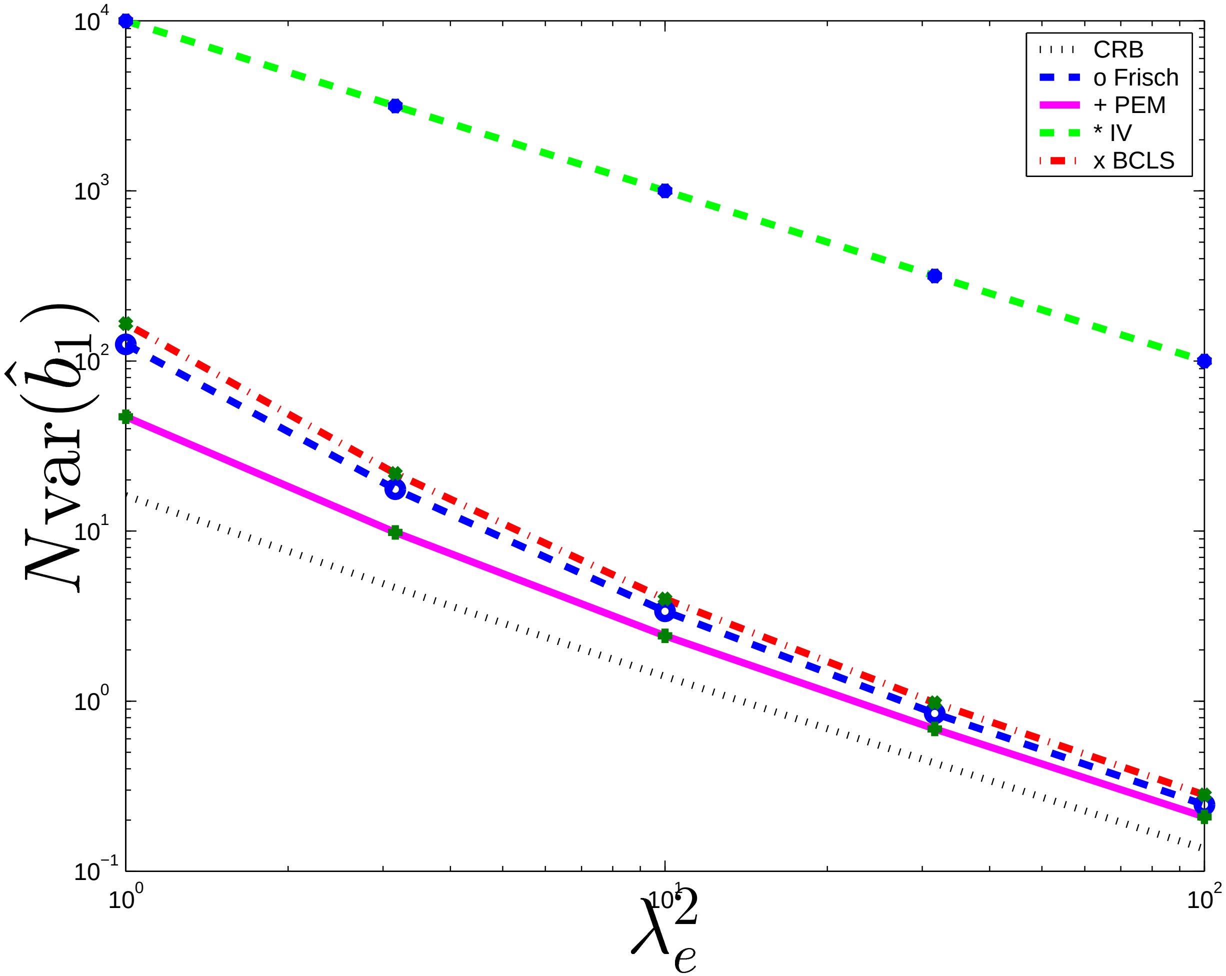
Contents

Some comparisons - computational load



Some comparisons - performance

A second order system; other parameters behave similarly.



Some comparisons – identifiability

Method	$\tilde{u}(t)$	$\tilde{y}(t)$	Experiment.
Basic IV	MA	ARMA	-
IV + WSF	MA	ARMA	-
BCLS	white	white/ARMA	-
Frisch	white	white/ARMA	-
TLS	white	white	> 1, or λ_y/λ_u known
SML	ARMA	ARMA	≥ 4
PEM	ARMA	ARMA	-
ML	ARMA	ARMA	-

UPPSALA UNIVERSITET

Informationsteknologi

Some comparisons – performance

Method	Comp. complexity	Accuracy
Basic IV	very low	low
IV + WSF	medium	medium-high
BCLS	low	medium-high
Frisch	low	medium-high
TLS	medium	medium-high
SML	medium-high	very high
PEM	high	high
ML	high	very high

UPPSALA UNIVERSITET

Informationsteknologi

Some open issues and future work

- Undermodeling

UPPSALA UNIVERSITET

Informationsteknologi

Some open issues and future work

- Undermodeling
- More of unification and relation between methods

UPPSALA UNIVERSITET

Informationsteknologi

Some open issues and future work

- Undermodeling
- More of unification and relation between methods
- Extensions to multivariate case

UPPSALA
UNIVERSITET

Informationsteknologi

Some open issues and future work

- Undermodeling
- More of unification and relation between methods
- Extensions to multivariate case
- Modeling in continuous-time

UPPSALA UNIVERSITET

Informationsteknologi

Some open issues and future work

- Undermodeling
- More of unification and relation between methods
- Extensions to multivariate case
- Modeling in continuous-time
- Model order determination

Thanks for listening !