

System identification for the errors-in-variables problem

Torsten Söderström

Department of Information Technology
Uppsala University, Sweden

Seminar, UTFSM, Valparaiso, November 2012 – p.1/62

Contents

- Background and motivation
 - Problem formulation

Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

Contents

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

Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

EIV in system identification

Work done in collaboration (discussions and joint publications) with J. C. Agüero, Th. Anderson, K. Burnham, M. Cedervall, H. F. Chen, M. Deistler, R. Diversi, M. Ekman, H. Garnier, M. Gilson, G. Goodwin, R. Guidorzi, Ch. Heij, H. Hjalmarsson, M. Hong, A. Karimi, E. Karlsson, A. Kukush, T. Larkowski, E. K. Larsson, J. Linden, K. Mahata, I. Markovsky, M. Mossberg, R. Pintelon, A. Rensfelt, W. Scherrer, J. Schoukens, V. Šimonyte, J. Sorelius, U. Soverini, P. Stoica, S. Thil, K. van Heusden, S. Van Huffel, K. Wada, W. X. Zheng

Seminar, UTFSM, Valparaiso, November 2012 – p.2/62

Contents

- Background and motivation
 - Problem formulation
 - Line fitting

Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

Contents

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

Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

Contents

- Background and motivation

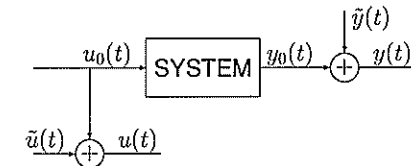
Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

Contents

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

Seminar, UTFSM, Valparaiso, November 2012 – p.3/62

1.1 Problem formulation

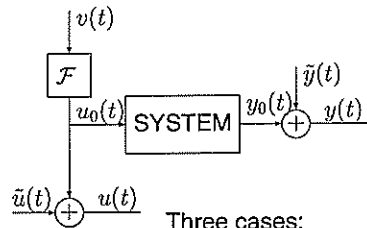


$\tilde{u}(t)$, $\tilde{y}(t)$ measurement noise.

Determine the system transfer function.

Seminar, UTFSM, Valparaiso, November 2012 – p.4/62

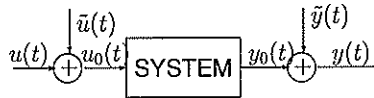
1.1 Problem formulation EIV, cont'd



Three cases:

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

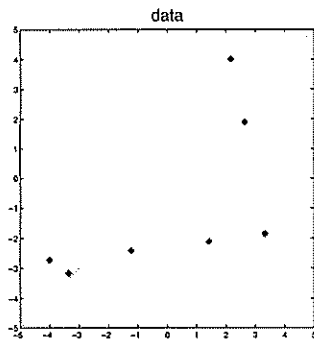
1.1 A related case



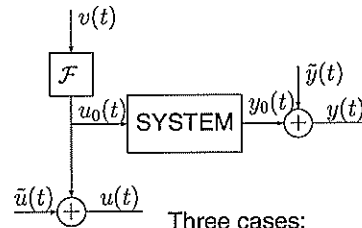
This is not an EIV problem! Why?

- $\tilde{u}(t)$ effects $y(t)$ [process noise!]
- $u(t)$ and $u_0(t)$ influence $y_0(t)$ in the same way

1.2 Line fitting, cont'd: Data



1.1 Problem formulation EIV, cont'd



Three cases:

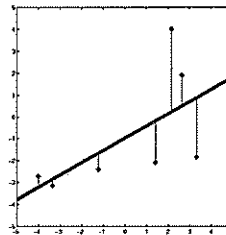
- v and \mathcal{F} unknown: True EIV situation
- v under control, \mathcal{F} unknown (repeated exp.)

1.1 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]

1.2 Line fitting, cont'd

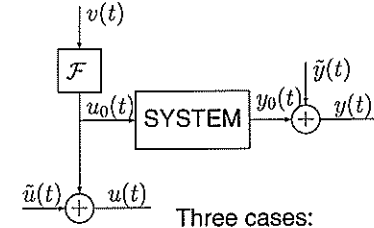
Least squares estimate



$$(\hat{a}, \hat{b}) = \arg \min_{a,b} V_1(a, b)$$

$$V_1(a, b) = \sum (y_i - ax_i - b)^2$$

1.1 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

1.2 Line fitting

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

Model

$$y_i = y_{0i} + \tilde{y}_i,$$

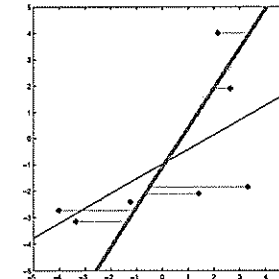
$$x_i = x_{0i} + \tilde{x}_i, \quad i = 1, \dots, n.$$

$$y_{0i} = a_0 x_{0i} + b_0,$$

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

1.2 Line fitting, cont'd

Data least squares estimate

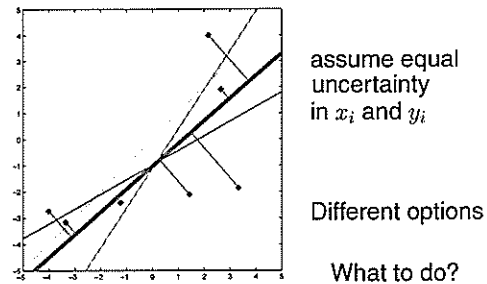


$$(\hat{a}, \hat{b}) = \arg \min_{a,b} V_2(a, b)$$

$$V_2(a, b) = \sum (x_i - \frac{y_i}{a} + \frac{b}{a})^2$$

1.2 Line fitting, cont'd

Total least squares estimate (orthogonal regression)



Seminar, UTFSM, Valparaiso, November 2012 – p.12/62

1.2 Line fitting, generalizations

- Multivariable model

$$y = b + \sum a_j x_j$$

Seminar, UTFSM, Valparaiso, November 2012 – p.15/62

1.2 Line fitting, generalizations

- Multivariable model

$$y = b + \sum a_j x_j$$

- Time series model

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

- Regressor variables will be correlated
- Uncertainties in different regressor variables are related!!

Seminar, UTFSM, Valparaiso, November 2012 – p.15/62

1.2 Line fitting, identifiability analysis

Use first and second order moments. Assume $E(x_{0i}) = m$, $\text{var}(x_{0i}) = \sigma^2$.

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

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

Seminar, UTFSM, Valparaiso, November 2012 – p.13/62

1.2 Line fitting, generalizations

- Multivariable model

$$y = b + \sum a_j x_j$$

- Time series model

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

Seminar, UTFSM, Valparaiso, November 2012 – p.15/62

1.2 Consequences of input noise

The presence of input noise $\tilde{u}(t)$ will cause most identification methods to give biased estimates. This may be tolerable in closed-loop design:

- G_c is usually insensitive to variations in G_0 (Examples: OP amplifiers, approximate linearization by feedback, etc)
- The robustness condition

$$\| \Delta_G T \|_\infty < 1$$

gives tolerance for deviations of open loop transfer function.

Seminar, UTFSM, Valparaiso, November 2012 – p.16/62

Line fitting, cont'd

Modified problem: Assume known noise variance ratio

$$\lambda_y / \lambda_x = r \text{ (known)}$$

This problem is feasible:

- Geometrically: Scale so that $\lambda_y = \lambda_x$. Use orthogonal regression.
- Algebraically: Use total least squares.
- Statistically: ML loss is well behaved and has a maximum for true parameter values of a and b . Seminar, UTFSM, Valparaiso, November 2012 – p.14/62

1.2 Line fitting, generalizations

- Multivariable model

$$y = b + \sum a_j x_j$$

- Time series model

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

- Regressor variables will be correlated

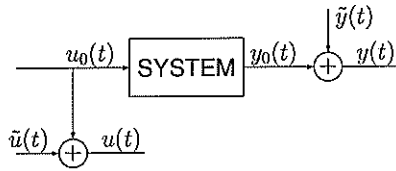
Seminar, UTFSM, Valparaiso, November 2012 – p.15/62

Contents

- Background and motivation
- Identifiability
 - Problem formulation, basic assumptions
 - Nonparametric models
 - How to handle lack of identifiability
 - Parametric models
- Estimators
- Comparisons and conclusions

Seminar, UTFSM, Valparaiso, November 2012 – p.17/62

2.1 Problem formulation EIV cont'd



Determine the system transfer function

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

Seminar, UTFSM, Valparaiso, November 2012 – p.18/62

2.3 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 possible estimates. [Set membership estimation]

Seminar, UTFSM, Valparaiso, November 2012 – p.21/62

2.3 Identifiability, cont'd

4. Use more than one experiment. [Assume the user can control the signal $v(t)$]
 - $\phi_{u_0}(\omega)$ differs between the different experiments,
 - or
 - $u_0(t)$ is (well) correlated between experiments, but $\tilde{y}(t)$, $\tilde{u}(t)$ are uncorrelated between experiments.

Seminar, UTFSM, Valparaiso, November 2012 – p.23/62

2.1 Identifiability: Assumptions

The available signals are time-discrete

$$\begin{aligned} u(t) &= u_0(t) + \tilde{u}(t), \\ y(t) &= y_0(t) + \tilde{y}(t). \end{aligned}$$

- The system is linear [causal] and asymptotically stable.
- $\tilde{u}(t)$, $\tilde{y}(t)$ are uncorrelated stationary processes, with zero means and spectra $\phi_{\tilde{u}}(\omega)$ and $\phi_{\tilde{y}}(\omega)$, respectively.
- $u_0(t)$ is persistently exciting and uncorrelated with $\tilde{u}(t)$ and $\tilde{y}(t)$.

Seminar, UTFSM, Valparaiso, November 2012 – p.19/62

2.3 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 possible estimates. [Set membership estimation]
2. Impose more detailed, parametric models of $u_0(t)$, $\tilde{u}(t)$, $\tilde{y}(t)$, say ARMA processes of specified orders.

Seminar, UTFSM, Valparaiso, November 2012 – p.21/62

2.4 Identifiability parametric models

Model $\tilde{u}(t)$, $\tilde{y}(t)$, $u_0(t)$ as ARMA processes (or white noise as a special case), and analyze identifiability.

Identifiability is often, but not always achieved.

Example:

$$\hat{G}\hat{\phi}_{u_0} \equiv G\phi_{u_0}$$

Assume that $G = B/A$ does not contain any pair of zeros, reflected in the unit circle, in order to avoid ambiguities.

Seminar, UTFSM, Valparaiso, November 2012 – p.24/62

2.2 Identifiability nonparametric models

Use second order statistics of $\zeta(t) = (y(t) \ u(t))^T$:

$$\begin{aligned} \Phi_{\zeta} &= \begin{pmatrix} GG^* & G \\ G^* & 1 \end{pmatrix} \phi_{u_0} + \begin{pmatrix} \phi_{\tilde{y}} & 0 \\ 0 & \phi_{\tilde{u}} \end{pmatrix} \\ &= \begin{pmatrix} \hat{G}\hat{G}^* & \hat{G} \\ \hat{G}^* & 1 \end{pmatrix} \hat{\phi}_{u_0} + \begin{pmatrix} \hat{\phi}_{\tilde{y}} & 0 \\ 0 & \hat{\phi}_{\tilde{u}} \end{pmatrix}. \end{aligned}$$

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. Cf static cases!

Seminar, UTFSM, Valparaiso, November 2012 – p.20/62

2.3 Identifiability, cont'd

3. Modify at least one of the assumptions on Gaussian distributed data $(u_0, \tilde{u}, \tilde{y})$. Use higher order statistics to gain additional information. Deistler(1986), Tugnait(1992).

Seminar, UTFSM, Valparaiso, November 2012 – p.22/62

2.4 Identifiability parametric models, cont'd

Assume $\phi_{\tilde{u}}, \phi_{\tilde{y}}$ general, G no reflected pair of zeros.

$$\begin{aligned} \hat{G}\hat{\phi}_{u_0} &\equiv G\phi_{u_0} \\ \Rightarrow \hat{G} &= \alpha G, \quad \hat{\phi}_{u_0} = \phi_{u_0}/\alpha \\ &\alpha > 0 \text{ (constant)} \end{aligned}$$

(cf static case)

Conditions

$$\hat{\phi}_{\tilde{u}}(\omega) \geq 0, \quad \hat{\phi}_{\tilde{y}}(\omega) \geq 0, \quad \forall \omega$$

$\Rightarrow \alpha \in$ a small interval around $\alpha = 1$.

Seminar, UTFSM, Valparaiso, November 2012 – p.25/62

Contents

- Background and motivation
- Identifiability
- Example of estimators
 - Least squares (LS), Instrumental variables (IV), Bias-compensated LS (BCLS), The Frisch scheme, Total least squares (TLS)
 - Prediction error method (PEM) and maximum likelihood (ML) method
 - Accuracy aspects
- Comparisons and conclusions

Seminar, UTFSM, Valparaiso, November 2012 – p.26/62

3.1 Notations for parametric estimators, cont'd

Denote covariance matrices and their estimates as

$$\mathbf{R}_\varphi = \mathbb{E} \{ \varphi(t) \varphi^T(t) \}, \quad \hat{\mathbf{R}}_\varphi = \frac{1}{N} \sum_{t=1}^N \varphi(t) \varphi^T(t).$$

Conventions:

- θ_0 denotes the true parameter vector
 $\hat{\theta}$ denotes its estimate.
- $\varphi_0(t)$ denotes the noise-free part of the regressor vector.
- $\tilde{\varphi}(t)$ denotes the noise-contribution to the regressor vector. ($\varphi(t) = \varphi_0(t) + \tilde{\varphi}(t)$)

Seminar, UTFSM, Valparaiso, November 2012 – p.29/62

3.3 Relation to factor models

$$\mathbf{R}_{\tilde{\varphi}} = \mathbf{R}_{\tilde{\varphi}_0} + \mathbf{R}_{\tilde{\varphi}}$$

$\tilde{y}(t), \tilde{u}(t)$ white $\Rightarrow \mathbf{R}_{\tilde{\varphi}}$ diagonal
 $\mathbf{R}_{\tilde{\varphi}_0}$ singular (co-rank = 1)

Factor model

- $\mathbf{R}_{\tilde{\varphi}_0}$ rank-deficient ('low rank') - due to latent variables
- $\mathbf{R}_{\tilde{\varphi}}$ diagonal

Generalizations: include dynamics!

Seminar, UTFSM, Valparaiso, November 2012 – p.32/62

3.1 Notations for parametric estimators

System description

$$A(q^{-1})y_0(t) = B(q^{-1})u_0(t),$$

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \\ B(q^{-1}) &= b_1q^{-1} + \dots + b_{n_b}q^{-n_b}. \end{aligned}$$

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

$$\begin{aligned} \theta &= (a_1 \dots a_{n_a} \ b_1 \dots b_{n_b})^T, \\ \varphi(t) &= (-y(t-1) \dots -y(t-n_a) \\ &\quad u(t-1) \dots u(t-n_b))^T. \end{aligned}$$

Seminar, UTFSM, Valparaiso, November 2012 – p.27/62

3.2 The least squares estimate is biased

Model

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

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

The least squares (LS) estimate

$$\begin{aligned} \hat{\theta}_{LS} &= \hat{\mathbf{R}}_\varphi^{-1} \hat{\mathbf{r}}_{\varphi y} \rightarrow \mathbf{R}_{\varphi}^{-1} \mathbf{r}_{\varphi y}, \quad N \rightarrow \infty \\ &= (\mathbf{R}_{\varphi_0} + \mathbf{R}_{\tilde{\varphi}})^{-1} \mathbf{r}_{\varphi_0 y_0} = (\mathbf{R}_{\varphi_0} + \mathbf{R}_{\tilde{\varphi}})^{-1} \mathbf{R}_{\varphi_0} \theta_0 \end{aligned}$$

Bias due to $\mathbf{R}_{\tilde{\varphi}}$.

Seminar, UTFSM, Valparaiso, November 2012 – p.30/62

3.4 Instrumental variable (IV) methods

The IV estimate can be defined as

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

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

= errors = – p.33/62

3.1 Notations for parametric estimators, cont'd

System description

$$\begin{aligned} &A(q^{-1})y(t) - B(q^{-1})u(t) \\ &= A(q^{-1})y_0(t) - B(q^{-1})u_0(t) \\ &+ A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned} \left. \vphantom{\begin{aligned} &A(q^{-1})y(t) - B(q^{-1})u(t) \\ &= A(q^{-1})y_0(t) - B(q^{-1})u_0(t) \\ &+ A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned}} \right\} = 0 \left. \vphantom{\begin{aligned} &A(q^{-1})y(t) - B(q^{-1})u(t) \\ &= A(q^{-1})y_0(t) - B(q^{-1})u_0(t) \\ &+ A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned}} \right\} \triangleq \varepsilon(t)$$

Hence, the system can be written as a linear regression

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

Note that $\varphi(t)$ and $\varepsilon(t)$ are correlated.

Seminar, UTFSM, Valparaiso, November 2012 – p.28/62

3.3 Relation to factor models

Extend regressor and parameter vectors

$$\tilde{\varphi}(t) = \begin{pmatrix} -y(t) \\ \varphi(t) \end{pmatrix}, \quad \tilde{\theta} = \begin{pmatrix} 1 \\ \theta \end{pmatrix}$$

System dynamics

$$\tilde{\varphi}_0^T \tilde{\theta}_0 = 0$$

$$\mathbf{R}_{\tilde{\varphi}} = \mathbf{R}_{\tilde{\varphi}_0} + \mathbf{R}_{\tilde{\varphi}}$$

Note $\mathbf{R}_{\tilde{\varphi}_0}$ is singular.

Seminar, UTFSM, Valparaiso, November 2012 – p.31/62

3.4 IV methods, properties

- Applicable under fairly general noise conditions.
- Inexpensive from a computational point of view.
- Often poor accuracy of $\hat{\theta}$.
- The matrix $\mathbf{R}_{\mathbf{z}\varphi}$ has to be full rank [a p.e. like condition on $u_0(t)$].

vari(errorS)ables \rightarrow – p.34/62

3.4 IV and higher-order statistics

Several high-order statistics methods can be interpreted as IV methods, with $z(t)$ formed by a simple nonlinear transformation of the data.

Example: Assume that $u_o(t)$ has a skewed distribution, $E u_o^3(t) \neq 0$.

$$z(t) = \begin{pmatrix} u^2(t-1) & \dots & u^2(t-na-nb) \end{pmatrix}^T$$

Cf Thil et al(2008).

Seminar UTFSM, Valparaiso, November 2012 – p.35/62

3.5 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 → low dimensional optimization problem

Seminar UTFSM, Valparaiso, November 2012 – p.38/62

3.6 GIVE, cont'd

Special cases

- Bias-eliminating least squares (Zheng, and others)
- Different Frisch schemes (Bologna group, and others)
- Extended compensated least squares (Ekman)

Role of equations vs. choice of algorithm

Seminar UTFSM, Valparaiso, November 2012 – p.41/62

3.5 Bias-compensating least squares, BCLS

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

$$\underbrace{\left(\hat{R}_{\varphi} - \begin{pmatrix} \hat{\lambda}_y I_{n_n} & 0 \\ 0 & \hat{\lambda}_u I_{n_b} \end{pmatrix} \right)}_{\hat{R}_{\varphi_0}} \hat{\theta} = \hat{r}_{\varphi y}$$

Many possibilities exist.

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

Seminar UTFSM, Valparaiso, November 2012 – p.36/62

3.6 Generalized IV estimate (GIVE)

Introduce a noise parameter vector ρ :

- \tilde{y} white:

$$\rho = \begin{pmatrix} \lambda_y & \lambda_u \end{pmatrix}^T$$

- \tilde{y} correlated:

$$\rho = \begin{pmatrix} r_{\tilde{y}}(0) & \dots & r_{\tilde{y}}(n_a) & \lambda_u \end{pmatrix}^T$$

Seminar UTFSM, Valparaiso, November 2012 – p.39/62

3.6 GIVE: Algorithmic aspects

Total parameter vector

$$\vartheta = \begin{pmatrix} \theta^T & \rho^T \end{pmatrix}^T$$

Biased-compensated IV equations

$$f(\vartheta) \triangleq \frac{1}{N} \sum_{t=1}^N z(t, \theta) \varepsilon(t, \theta) - r_{z\varepsilon}(\theta, \rho) \approx 0$$

Criterion

$$V(\theta, \rho) = \|f(\vartheta)\|_{W(\vartheta)}^2$$

$$\hat{\theta}, \hat{\rho} = \arg \min_{\theta, \rho} V(\theta, \rho)$$

Seminar UTFSM, Valparaiso, November 2012 – p.42/62

3.5 BCLS, cont'd

There are many variants

- $\bar{u}(t), \bar{y}(t)$ 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), Beghelli et al(1990), Diversi et al(2003, 2006), Ekman(2005), Ekman et al(2006), TS(2008)

Interrelations: Hong et al(2009)

Seminar UTFSM, Valparaiso, November 2012 – p.37/62

3.6 GIVE, cont'd

'Biased-compensated IV equations'

$$\left(\hat{R}_{z\rho} - R_{z\tilde{y}}(\rho) \right) \theta = \hat{r}_{z\tilde{y}} - r_{z\tilde{y}}(\rho)$$

where

$$z(t) = \begin{pmatrix} y(t) \\ \vdots \\ y(t-n_a-p_y) \\ u(t-1) \\ \vdots \\ u(t-n_b-p_u) \end{pmatrix}$$

Seminar UTFSM, Valparaiso, November 2012 – p.40/62

3.6 GIVE: Algorithmic aspects, cont'd

Assume W does not depend on θ : A variable projection algorithm is possible:

$$\theta(\rho) \triangleq \arg \min_{\theta} V(\theta, \rho)$$

$$\bar{V}(\rho) = V(\theta(\rho), \rho)$$

$$\hat{\rho} = \arg \min_{\rho} \bar{V}(\rho)$$

$$\hat{\theta} = \theta(\hat{\rho})$$

Seminar UTFSM, Valparaiso, November 2012 – p.43/62

3.7 Covariance matching

Define

$$z_0(t) = \frac{1}{A(q^{-1})} u_0(t)$$

$$a_0 = 1, \quad r_0(\tau) = E\{z_0(t+\tau)z_0(t)\}$$

Covariance functions of the measured data

$$r_u(\tau) = \sum_i \sum_j a_i a_j r_0(\tau - i + j), \quad (\tau > 0)$$

$$r_y(\tau) = \sum_i \sum_j b_i b_j r_0(\tau - i + j), \quad (\tau > 0)$$

$$r_{yu}(\tau) = \sum_i \sum_j b_i a_j r_0(\tau - i + j)$$

Seminar UTFSM, Valparaiso, November 2012 – p.44/62

3.7 Covariance matching, cont'd

Variable projection algorithm

$$\{\hat{\theta}, \hat{r}_z\} = \arg \min_{\theta, r_z} \|\hat{r} - F(\theta)r_z\|_W^2$$

$$\Rightarrow \hat{r}_z = (F^T(\theta)WF(\theta))^{-1}F^T(\theta)W\hat{r}$$

$$\hat{\theta} = \arg \min_{\theta} \left[\hat{r}^T W \hat{r} - \hat{r}^T W F(\theta) \right. \\ \left. \times (F^T(\theta)WF(\theta))^{-1}F^T(\theta)W\hat{r} \right]$$

Seminar UTFSM, Valparaiso, November 2012 – p.48/62

3.8 PEM and ML, cont'd

Prediction errors

$$\varepsilon(t, \theta) = \zeta(t) - \hat{\zeta}(t|t-1; \theta)$$

$$= H^{-1}(q^{-1}; \theta)\zeta(t).$$

PEM (prediction error method) estimate

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

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

Seminar UTFSM, Valparaiso, November 2012 – p.50/62

3.7 Covariance matching, cont'd

Step 1 Estimate the covariance vectors

$$r_y = \begin{pmatrix} r_y(1) \\ \vdots \\ r_y(p_y) \end{pmatrix} \quad r_u = \begin{pmatrix} r_u(1) \\ \vdots \\ r_u(p_u) \end{pmatrix} \quad r_{yu} = \begin{pmatrix} r_{yu}(p_1) \\ \vdots \\ r_{yu}(p_2) \end{pmatrix}$$

from data.

Step 2 Determine θ and

$$r_z = \begin{pmatrix} r_0(0) \\ \vdots \\ r_0(k) \end{pmatrix}$$

Seminar UTFSM, Valparaiso, November 2012 – p.45/62

3.8 Maximum likelihood

Model noise and noise-free input as well as the system.

Example with $\tilde{y}(t)$, $\tilde{u}(t)$ white, $u_0(t)$ ARMA process:

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

$\theta = \text{coeff}(A, B, C, D, \lambda_u, \lambda_y, \lambda_v)$

Seminar UTFSM, Valparaiso, November 2012 – p.48/62

3.8 PEM and ML, cont'd

ML estimate (Gaussian data)

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

with

$$\ell(\varepsilon, \theta, t) = \frac{1}{2} \log \det Q(\theta) + \frac{1}{2} \varepsilon^T(t, \theta) Q^{-1}(\theta) \varepsilon(t, \theta),$$

$$Q(\theta) = E \{ \varepsilon(t, \theta) \varepsilon^T(t, \theta) \}.$$

Seminar UTFSM, Valparaiso, November 2012 – p.51/62

3.7 Covariance matching, cont'd

Model

$$r \triangleq \begin{pmatrix} r_y \\ r_u \\ r_{yu} \end{pmatrix} = \begin{pmatrix} F_y(\theta) \\ F_u(\theta) \\ F_{yu}(\theta) \end{pmatrix} r_z \triangleq F(\theta)r_z$$

Seminar UTFSM, Valparaiso, November 2012 – p.46/62

3.8 ML, innovations form

$$\Phi_{\zeta}(\omega) \equiv H(e^{i\omega}; \theta)Q(\theta)H^*(e^{i\omega}; \theta)$$

H monic; $H(q^{-1})$, $H^{-1}(q^{-1})$ asymptotically stable

Consequence: Optimal one-step prediction errors

$$\varepsilon(t, \theta) = \zeta(t) - \hat{\zeta}(t|t-1; \theta)$$

$$= H^{-1}(q^{-1}; \theta)\zeta(t)$$

(can be computed using a Kalman filter; spectral factorization using an algebraic Riccati equation)

Seminar UTFSM, Valparaiso, November 2012 – p.49/62

3.8 ML, cont'd

The ML estimate can alternatively be computed in the frequency domain, Pintelon-Schoukens(2005), Goodwin et al(2010), [some differences in how transient effects are handled]

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

Seminar UTFSM, Valparaiso, November 2012 – p.52/62

3.8 ML, cont'd

General properties:

- (Very) high accuracy.
- The numerical optimization procedure is, in general, quite complex.
- Computations in the frequency domain is possible.
- The procedure may fail to give good results if only poor initial parameter estimates are available.
- A prediction error method leads to another criterion, and somewhat degraded accuracy.

Seminar UTFSM, Valparaiso, November 2012 – p.53/62

3.9 How good can the estimates be?

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

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\text{dist}} \mathcal{N}(0, P),$$

The covariance matrix P depends on

- the method (and the user parameters),
- the system,
- the dynamics for $u_0(t)$, $\tilde{u}(t)$, $\tilde{y}(t)$.

Seminar UTFSM, Valparaiso, November 2012 – p.54/62

Contents

- Background and motivation
- Identifiability
- Estimators
- Comparisons and conclusions
 - Example: Computational load
 - Example: Statistical accuracy
 - Some comparisons
 - Further issues (some open)

Seminar UTFSM, Valparaiso, November 2012 – p.57/62

3.9 How good can the estimates be?

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

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\text{dist}} \mathcal{N}(0, P),$$

The covariance matrix P depends on

- the method (and the user parameters),

Seminar UTFSM, Valparaiso, November 2012 – p.54/62

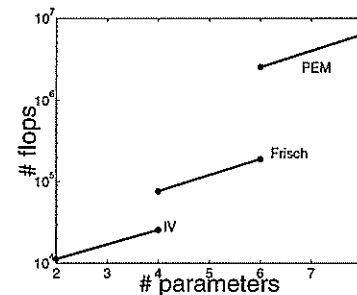
3.9 How good can the estimates be?, cont'd

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), ECLS, GIVE, Söderström(2011), CM, Söderström(2010).
- Prediction error method and maximum likelihood method, Ljung(1999), Söderström(2006).

4.1 Some comparisons - computational load

First and second order systems;



Seminar UTFSM, Valparaiso, November 2012 – p.58/62

3.9 How good can the estimates be?

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

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\text{dist}} \mathcal{N}(0, P),$$

The covariance matrix P depends on

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

Seminar UTFSM, Valparaiso, November 2012 – p.54/62

3.9 How good can the estimates be?, cont'd

The Cramér-Rao lower bound P_{CRLB}

$$\text{cov}(\hat{\theta} - \theta_0) \geq P_{\text{CRLB}} = J^{-1} = P_{\text{ML}},$$

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

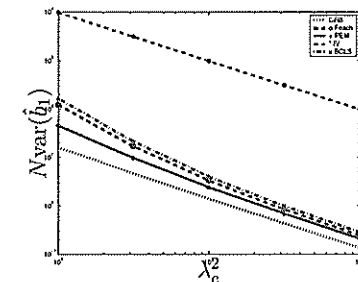
where $L(\theta)$ is the likelihood function. The matrix J is the Fisher information matrix.

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

Seminar UTFSM, Valparaiso, November 2012 – p.56/62

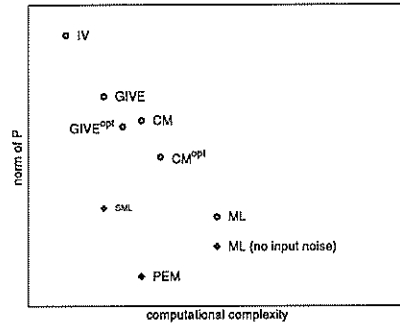
4.2 Some comparisons - performance

A second order system; $N_{\text{var}}(\hat{b}_1)$ vs. λ_c^2 ; other parameters behave similarly.



Seminar UTFSM, Valparaiso, November 2012 – p.59/62

4.3 Some comparisons - performance



Seminar UTFSM, Valparaiso, November 2012 – p.60/62

4.4 Some further issues (partly open)

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

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

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

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

- Undermodeling

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

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

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

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

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

- Undermodeling
- More of unification and relation between methods

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

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

Seminar UTFSM, Valparaiso, November 2012 – p.61/62

4.4 Some further issues (partly open)

- Undermodeling
- More of unification and relation between methods
- Extensions to the multivariate case
- Modeling in continuous-time
- Model order determination
- Recursive algorithms
- Nonlinear models
- Closed-loop experiments

Seminar UTFSM, Valparaiso, November 2012 – p.61/62