On errors-in-variables estimation with unknown noise variance ratio

Sabine Van Huffel, Ivan Markovsky, and Alexander Kukush

K.U.Leuven, ESAT-SISTA

Outline

Classic static EIV model

$$D = \overline{D} + \widetilde{D}, \qquad \operatorname{vec}(\widetilde{D}) \sim \mathrm{N}(0, \sigma^2 I), \quad \operatorname{col\,span}(\overline{D}) \subset \overline{\mathscr{B}}$$

- $\overline{D} \in \mathbb{R}^{\mathrm{d} \times N}$ — "true" data matrix
- $\mathrm{d} := \operatorname{row\,dim}(D)$ — number of variables
- $N := \operatorname{col\,dim}(D)$ — number of data points ($N > \mathrm{d}$)
- $\overline{\mathscr{B}}$ — "true" linear static model (a subspace of $\mathbb{R}^{\mathrm{d}}$)
- $\widetilde{D}$ — measurement errors (zero mean, i.i.d., Gaussian)
- D — measured data matrix

$\overline{D}$ satisfies a linear static model $\overline{\mathscr{B}} \iff \overline{D}$ is low-rank

$\mathscr{B}$ is a subspace of $\mathbb{R}^{\mathrm{d}}$ with dimension $m := \operatorname{rank}(\overline{D}) < \mathrm{d}$

Classic total least squares method

$$\{\widehat{R}_{\mathrm{tls}}, \widehat{D}_{\mathrm{tls}}\} := \arg\min_{R, \widehat{D}} \|D - \widehat{D}\|_{\mathrm{F}}^2 \ \text{ subject to } \ RR^\top = I_p, \ R\widehat{D} = 0$$

- $\widehat{D}_{\mathrm{tls}}$ — TLS estimate of true data matrix $\overline{D}$
- $\widehat{\mathscr{B}}_{\mathrm{tls}} := \ker(\widehat{R}_{\mathrm{tls}})$ — TLS estimate of the true model $\overline{\mathscr{B}}$

Notes:

- Typically we are interested in $\widehat{\mathscr{B}}_{\mathrm{tls}}$, not $\widehat{D}_{\mathrm{tls}}$.
- $\widehat{\mathscr{B}}_{\mathrm{tls}}$ is maximum likelihood estimate of $\overline{\mathscr{B}}$ in the EIV model.
- Without the Gaussianity assumption, $\widehat{\mathscr{B}}_{\mathrm{tls}}$ is not ML but still a consistent estimator in the EIV model.
- The i.i.d. assumption is strong and we aim to relax it.

Generalized EIV model and TLS method

$$D = \bar{D} + \tilde{D}, \qquad \mathrm{vec}(\tilde{D}) \sim \mathrm{N}(0, \sigma^2 W), \quad \mathrm{colspan}(\bar{D}) \subset \bar{\mathscr{B}}$$

- $V := \sigma^2 W$ — measurement error covariance matrix

Maximum likelihood estimate — weighted TLS

$$\{\widehat{R}_{\mathrm{tls}}, \widehat{D}_{\mathrm{tls}}\} := \arg\min_{R,\widehat{D}} \mathrm{vec}^\top(D - \widehat{D}) W^{-1} \mathrm{vec}(D - \widehat{D})$$

$$\text{subject to} \quad RR^\top = I_p \quad \text{and} \quad R\widehat{D} = 0$$

- V should be known up to a scaling factor.
- This is a restrictive assumption and we aim to relax it.

EIV model with unknown noise variance ratio

$$\mathbf{E}\tilde{D}\tilde{D}^\top = \begin{bmatrix} \mathbf{E}\tilde{D}_{\mathrm{i}}\tilde{D}_{\mathrm{i}}^\top & \mathbf{E}\tilde{D}_{\mathrm{i}}\tilde{D}_{\mathrm{o}}^\top \\ \mathbf{E}\tilde{D}_{\mathrm{o}}\tilde{D}_{\mathrm{i}}^\top & \mathbf{E}\tilde{D}_{\mathrm{o}}\tilde{D}_{\mathrm{o}}^\top \end{bmatrix} =: \begin{bmatrix} \bar{\lambda}_{\mathrm{i}} W_{\mathrm{i}} & 0 \\ 0 & \bar{\lambda}_{\mathrm{o}} W_{\mathrm{o}} \end{bmatrix}$$

where

- $W_{\mathrm{i}} \in \mathbb{R}^{m\times m}$, $W_{\mathrm{i}} > 0$ and $W_{\mathrm{o}} \in \mathbb{R}$, $W_{\mathrm{o}} > 0$ are known
- $\bar{\lambda}_{\mathrm{i}}$ and $\bar{\lambda}_{\mathrm{o}}$ are unknown positive scalars

This model is not identifiable, so additional assumptions are needed in order to make the estimation problem well defined.

Such assumptions are, *e.g.*,

- several independent data sets are available for estimation, or
- the true data can be clustered.

Derivation of the estimator

Define the noise variance ratio $\bar{\lambda}_{\mathrm{i}}/\bar{\lambda}_{\mathrm{o}} =: \bar{\mu}$ and let

$$W(\mu) := \begin{bmatrix} \mu W_{\mathrm{i}} & 0 \\ 0 & W_{\mathrm{o}} \end{bmatrix}, \qquad \text{so that} \quad \mathbf{E}\tilde{D}\tilde{D}^\top = \bar{\lambda}_{\mathrm{o}} W(\bar{\mu}).$$

Assuming $\bar{\mu}$ is known, we can solve the weighted TLS problem

$$\min_{R,\widehat{D}} \left\| W^{-1/2}(\bar{\mu})(D - \widehat{D}) \right\|_{\mathrm{F}}^2 \quad \text{subject to} \quad R\widehat{D} = 0,$$

or equivalently the nonlinear system of equations

$$R\left(DD^\top - \lambda_{\mathrm{o}} W(\bar{\mu})\right) = 0,$$

where we aim at a solution corresponding to a minimal λ_{o}.

Computationally, we solve a generalized SVD problem.

Derivation of the estimator (cont.)

With unknown $\bar{\mu}$, we consider two estimating equations

$$R\left(D^k (D^k)^\top - \lambda_{\mathrm{o}} W(\mu)\right) = 0, \quad \text{for } k = 1, 2,$$

corresponding to two disjoint subsets D^1 and D^2 of the data

$$D\Pi =: \begin{bmatrix} D^1 & D^2 \end{bmatrix} =: \begin{array}{c} \begin{bmatrix} D_{\mathrm{i}}^1 & D_{\mathrm{i}}^2 \\ D_{\mathrm{o}}^1 & D_{\mathrm{o}}^2 \end{bmatrix} \end{array} \begin{array}{c} m \\ 1 \end{array}, \qquad \Pi \text{ — permutation matrix.}$$

The clustering problem is

$$\max_{\substack{\text{permutation} \\ \text{matrix } \Pi}} \left(\min_{j=1,\dots,m} \left| \lambda_j \left(D_{\mathrm{i}}^1 (D_{\mathrm{i}}^1)^\top - D_{\mathrm{i}}^2 (D_{\mathrm{i}}^2)^\top \right) \right| \right),$$

where $\lambda_1(A), \dots, \lambda_{\dim(A)}(A)$ are the eigenvalues of A.

Derivation of the estimator (cont.)

Aim: find a common generalized eigenvalue-eigenvector for

$$\left(D^k (D^k)^\top, W(\mu)\right), \quad k = 1,2$$

Nonlinear least squares-type approximate solution

$$\widehat{\mu} = \arg\min_{\mu} \left(\left(\lambda_{\rm o}^1 - \lambda_{\rm o}^2\right)^2 + C \sin^2\left(\angle(R^1, R^2)\right) \right),$$

- $(\lambda_{\rm o}^k, R^k)$ minimal eigenvalue-eigenvec. of $\left(D^k (D^k)^\top, W(\mu)\right)$
- C — regularization parameter
- $\angle(R^1, R^2)$ — angle between the vectors R^1 and R^2
- $\left(\lambda_{\rm o}^1 - \lambda_{\rm o}^2\right)^2$ makes both eigenvalues close to each other
- $\sin^2\left(\angle(R^1, R^2)\right)$ makes the corresponding eigenvec. close

Summary of the proposed estimation algorithm

1. Cluster the data using, *e.g.*, the K-means algorithm.
2. Compute the noise variance ratio estimate $\widehat{\mu}$ by solving

$$\widehat{\mu} = \arg\min_{\mu} \left(\left(\lambda_{\rm o}^1 - \lambda_{\rm o}^2\right)^2 + C \sin^2\left(\angle(R^1, R^2)\right) \right),$$

   for the clusters identified on step 1.
3. Solve the weighted TLS problem for the estimated value of μ on step 2.

Simulation example

$\overline{D} \in \mathbb{R}^{3 \times 2N'}$ is a random rank-2 matrix, with $N' = 10, \ldots, 500$

and two clusters — the first N' and the last N' columns of $\overline{D}$

apply the algorithm for 500 noise realizations with
$\lambda_{\rm i} = 0.01$ and $\lambda_{\rm o} = 0.04$

average relative estimation error of estimation

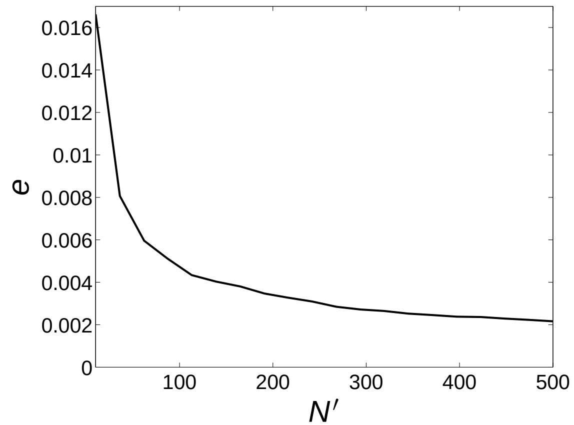
$$e := \frac{1}{500} \sum_{k=1}^{500} \frac{\|\overline{X} - \widehat{X}^{(k)}\|}{\|\overline{X}\|}$$

where $\overline{\mathscr{B}} =: \ker(\begin{bmatrix} \overline{X}^\top & -1 \end{bmatrix})$ and $\widehat{R} =: \begin{bmatrix} \widehat{X}^\top & -1 \end{bmatrix}$ (normalization)

$\widehat{X}^{(i)}$ — estimate of $\overline{X}$ in the ith repetition of the experiment

Simulation example (cont.)

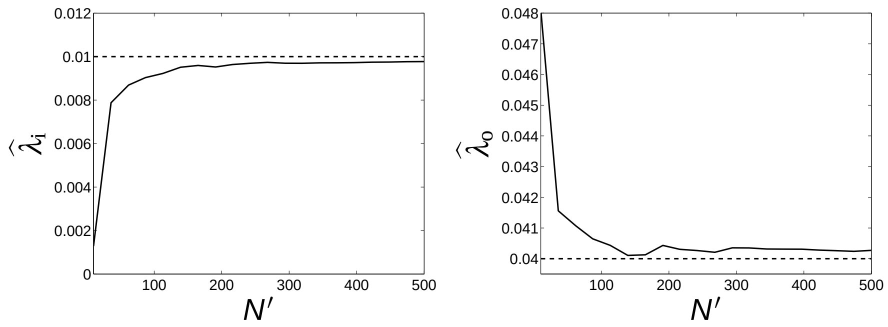
Relative error e as a function of half the sample size N'.



Simulation example (cont.)

Average values of $\widehat{\lambda}_{\rm i}$ and $\widehat{\lambda}_{\rm o}$ as functions of N'.

dashed lines — the true values



Conclusions

- EIV model with error cov. known up to two parameters
- identifiable if the data has two distinct clusters
- estimation procedure:
  1. cluster the data
  2. solve a univariate optimization problem for μ
  3. solve a weighted TLS problem for $\widehat{\mu}$
- generalizes to problems with more than two parameters:
  as many clusters are needed as there are parameters
  the optimization on step 2, however, becomes multidim.
- with Hankel structured data matrix the method is applicable to system identification

Thank you