Automatica 66 (2016) 218–230

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Errors-in-variables system identification using structural equation modeling☆

David Kreiberg [a,b], Torsten Söderström [c], Fan Yang-Wallentin [a]

[a] Department of Statistics, Uppsala University, PO Box 513, SE-75120, Uppsala, Sweden
[b] Department of Economics, Norwegian Business School, 0442 Oslo, Norway
[c] Division of Systems and Control, Department of Information Technology, Uppsala University, PO Box 337, SE-75105, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form
20 July 2015
Accepted 20 November 2015
Available online 25 January 2016

Keywords:
System identification
Errors-in-variables models
Linear systems
Structural equation models

ABSTRACT

Errors-in-variables (EIV) identification refers to the problem of consistently estimating linear dynamic systems whose output and input variables are affected by additive noise. Various solutions have been presented for identifying such systems. In this study, EIV identification using Structural Equation Modeling (SEM) is considered. Two schemes for how EIV Single-Input Single-Output (SISO) systems can be formulated as SEMs are presented. The proposed formulations allow for quick implementation using standard SEM software. By simulation examples, it is shown that compared to existing procedures, here represented by the covariance matching (CM) approach, SEM-based estimation provide parameter estimates of similar quality.

1. Introduction

Several estimation methods have been proposed for identifying linear dynamic systems from noise-corrupted output measurements, see for instance Ljung (1999) and Söderström and Stoica (1989). On the other hand, estimation of the parameters of systems in which the input signal is also affected by noise, here referred to as "errors-in-variables" (EIV) models, is recognized as a more delicate problem. Studying such systems is of interest due to their potential usage in the engineering sciences and elsewhere.

Established techniques for handling the EIV problem include the bias-eliminating least squares (Zheng, 1998, 2002), the Frisch estimator (Beghelli, Castaldi, Guidorzi, & Soverini, 1993; Beghelli, Guidorzi, & Soverini, 1990; Diversi & Guidorzi, 2012; Diversi, Guidorzi, & Soverini, 2003, 2004, 2006; Guidorzi, Diversi, & Soverini, 2008; Söderström, 2008) and various forms of bias-compensated least squares (Ekman, 2005; Ekman, Hong, & Söderström, 2006; Mahata, 2007). An overview of EIV system identification, containing various solutions from the literature, can be found in Söderström (1981, 2007, 2012). The topic is also treated from different points of view in the books (Cheng & Van Ness, 1999; Fuller, 2006). A more recent development is represented by the covariance matching (CM) approach introduced in Mossberg and Söderström (2011), Söderström and Mossberg (2011) and Söderström, Mossberg, and Hong (2009). Mossberg and Söderström (2012) and Söderström, Kreiberg, and Mossberg (2014). This approach has been shown to be related to structural equation modeling (SEM) techniques. In Kreiberg, Söderström, and Yang-Wallentin (2013), it is demonstrated how SEM can be applied to the problem of EIV system identification.

The objective of the present study is to further extend and refine the SEM approach. As compared to Kreiberg et al. (2013), we provide a more thorough analysis of how SEM can be applied to the EIV problem. Two different and quite general formulations of the EIV system as SEMs are presented, and their relation is analyzed. To facilitate the SEM implementation of such systems, several extensions of the standard SEM framework are proposed. The suggested formulations are evaluated in terms of statistical and numerical performance using simulated data. Aspects concerning the implementation, which were only briefly considered in Kreiberg et al. (2013), are studied in more detail. In the simulation examples, standard software developed for SEM-based estimation is used.

The study is organized as follows. First, in Section 2, we outline the background of the EIV problem. In Section 3, the standard

☆ This work was supported in part by The Swedish Research Council (VR) under the contract 421-2011-1727. The material in this paper was presented at the 52nd IEEE Conference on Decision and Control December, 10–13, 2013, Florence, Italy. This paper was recommended for publication in revised form by Associate Editor James Lam under the direction of Editor Ian R. Petersen.

E-mail addresses: David.Kreiberg@statistics.uu.se, David.Kreiberg@bi.no (D. Kreiberg), Torsten.Soderstrom@it.uu.se (T. Söderström), Fan.Yang@statistics.uu.se (F. Yang-Wallentin).

http://dx.doi.org/10.1016/j.automatica.2015.12.007
0005-1098/

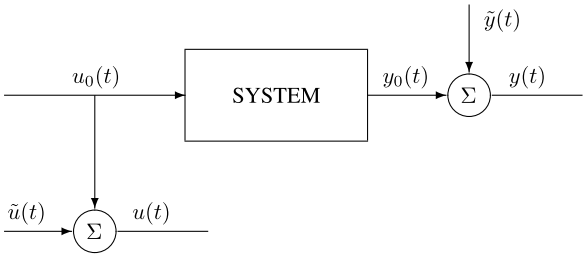

Fig. 1. Basic setup for the dynamic EIV problem.

SEM framework for static systems is reviewed. In Section 4, it is shown how EIV systems can be formulated as SEMs, and in Section 5, simulation examples of the two formulations are presented. Finally, in Section 6, concluding remarks are given.

2. EIV system formulation

First, we define the signals entering the system and then describe the general EIV problem for linear dynamic systems. The usual setup of the EIV problem is illustrated in Fig. 1.

Our interest lies in the linear Single-Input Single-Output (SISO) system described by

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

where $y_0(t)$ and $u_0(t)$ are the noise-free output and input signals, respectively, and $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward shift operator q^{-1} of the form

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}, \tag{2}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}. \tag{3}$$

We allow the noise-free signals to be corrupted by additive measurement noises $\tilde{y}(t)$ and $\tilde{u}(t)$. The available signals are in discrete time and are given by

$$y(t) = y_0(t) + \tilde{y}(t), \tag{4}$$

$$u(t) = u_0(t) + \tilde{u}(t). \tag{5}$$

Since $y_0(t)$ and $u_0(t)$ are not directly observable, the signals are considered *latent*.

The assumptions related to the system and its components are as follows:

- $\mathcal{A}1$. All signals and disturbances are zero mean stationary processes.
- $\mathcal{A}2$. The polynomials $A(q^{-1})$ and $B(q^{-1})$ are coprime and their respective degrees n_a and n_b are known.
- $\mathcal{A}3$. Data records of the noisy output and input signals $\{y(t), u(t)\}_{t=1}^{N}$ are known.
- $\mathcal{A}4$. The noise-free input $u_0(t)$ is unknown as well as its second order properties such as its spectrum $\phi_{u_0}(\omega)$.
- $\mathcal{A}5$. The measurement noises $\tilde{y}(t)$ and $\tilde{u}(t)$ are white and mutually uncorrelated. Moreover, $\tilde{y}(t)$ and $\tilde{u}(t)$ are both uncorrelated with $u_0(t-\tau)$ for all τ. Their unknown variances are denoted $\psi_{\tilde{y}}$ and $\psi_{\tilde{u}}$.

Our concern is to determine the system transfer function described by

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}}. \tag{6}$$

It follows that the parameter vector to be estimated from the noisy data is

$$\boldsymbol{\theta}_0 = \begin{pmatrix} a_1 & \cdots & a_{n_a} & b_1 & \cdots & b_{n_b} \end{pmatrix}^T, \tag{7}$$

where the superscript T denotes the transpose. It may also be of interest to determine other system characteristics such as the measurement noise variances $\boldsymbol{\psi}_0 = (\psi_{\tilde{y}}\ \psi_{\tilde{u}})^T$.

3. Structural equation modeling

In multivariate statistics, SEM is a well established statistical technique which has become popular within many disciplines of social science research. The popularity of SEM stems from its versatility, in which estimation problems involving latent variables and measurement errors can be handled. The versatility is also seen from the fact that numerous types of statistical problems can be formulated within the SEM framework. In what follows, we only briefly summarize the basics of SEM. For a more thorough introduction, see Bartholomew, Knott, and Moustaki (2011) and Bollen (1989).

3.1. Model formulation

The basic framework of SEM is described by the following three equations

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{8}$$

$$\mathbf{x}_1 = \boldsymbol{\Lambda}_1\boldsymbol{\eta} + \boldsymbol{\epsilon}_1, \tag{9}$$

$$\mathbf{x}_2 = \boldsymbol{\Lambda}_2\boldsymbol{\xi} + \boldsymbol{\epsilon}_2. \tag{10}$$

The first equation is referred to as the *structural equation*, while the latter two equations are known as the *measurement equations*. The random vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ consist of unobserved (or latent) quantities, whereas the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ consist of observed quantities.

The structural equation describes the relationship among the latent quantities, wherein $\boldsymbol{\eta}$ is endogenous and $\boldsymbol{\xi}$ is exogenous. The parameter matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$ consist of elements that represent the effect of $\boldsymbol{\eta}$ on $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$, respectively. It is assumed that $\mathbf{I} - \mathbf{B}$ is nonsingular such that $\boldsymbol{\eta}$ can be uniquely determined by $\boldsymbol{\xi}$ and the noise vector $\boldsymbol{\delta}$. It is further assumed that $\boldsymbol{\delta}$ has expectation zero and is mutually uncorrelated with $\boldsymbol{\xi}$.

The measurement equations describe how the observed quantities depend on the latent quantities. The parameter matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are so-called *loading* matrices whose elements represent the effect of $\boldsymbol{\eta}$ on $\mathbf{x}_1$ and $\boldsymbol{\xi}$ on $\mathbf{x}_2$, respectively. The measurement noises $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ may or may not be correlated, but are assumed to be mutually uncorrelated with $\boldsymbol{\eta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$. Note that the measurement equations are modeling devices in their own right. When a measurement equation is implemented without considering the remaining equations, the model is referred to as a *Confirmatory Factor Analysis* (CFA) model. Additional details are given in Bartholomew et al. (2011) and Bollen (1989).

The dimensions of the parameter matrices in (8)–(10) follow from the dimensions of the random vectors. Let n_η, n_ξ, n_{x_1} and n_{x_2} denote the number of elements in $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. The dimensions are then given by

$$\mathbf{B}\left(n_\eta \times n_\eta\right), \qquad \boldsymbol{\Gamma}\left(n_\eta \times n_\xi\right), \tag{11}$$

$$\boldsymbol{\Lambda}_1\left(n_{x_1} \times n_\eta\right), \qquad \boldsymbol{\Lambda}_2\left(n_{x_2} \times n_\xi\right). \tag{12}$$

The model framework additionally include the following covariance matrices

$$\mathsf{E}\left\{\boldsymbol{\xi}\boldsymbol{\xi}^T\right\} = \boldsymbol{\Phi}, \qquad \mathsf{E}\left\{\boldsymbol{\delta}\boldsymbol{\delta}^T\right\} = \boldsymbol{\Psi}_\delta, \tag{13}$$

$$\mathsf{E}\left\{\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^T\right\} = \boldsymbol{\Psi}_{\epsilon_1}, \qquad \mathsf{E}\left\{\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_2^T\right\} = \boldsymbol{\Psi}_{\epsilon_2}, \tag{14}$$

where E is the expectation operator. The dimensions of the matrices in (13) and (14) follow immediately from the dimensions of the involved vectors. Depending on the noise structure, $\boldsymbol{\Psi}_\delta$, $\boldsymbol{\Psi}_{\epsilon_1}$ and $\boldsymbol{\Psi}_{\epsilon_2}$ may or may not be diagonal.

The elements of $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}_\delta$, $\boldsymbol{\Psi}_{\epsilon_1}$ and $\boldsymbol{\Psi}_{\epsilon_2}$ are either free or constrained. An element is said to be constrained if it is assigned a specific value, or if it is a function (linear or nonlinear) of other elements. In SEM, it is common to constrain a large number of elements to zero. An example is when any or all of the

matrices $\mathbf{\Psi}_\delta$, $\mathbf{\Psi}_{\epsilon_1}$ and $\mathbf{\Psi}_{\epsilon_2}$ are restricted to be diagonal. Another form of constraint that is commonly imposed is to restrict two or more elements to be equal. Such constraints are referred to as *equality constraints*. By appropriately constraining the elements of the parameter matrices, various statistical problems can be formulated. An essential task in many applications of SEM is to assess the validity of the imposed constraints.

A parameter vector $\boldsymbol{\vartheta}$ comprises the unique elements of $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\mathbf{\Phi}$, $\mathbf{\Psi}_\delta$, $\mathbf{\Psi}_{\epsilon_1}$ and $\mathbf{\Psi}_{\epsilon_2}$ for which no pre-assigned values have been specified. The entries of this vector are referred to as the *free parameters*.

3.2. Estimation

Suppose that a set of data points on $\mathbf{x}_1$ and $\mathbf{x}_2$ are observed. Given the data, one would like to estimate the parameter vector $\boldsymbol{\vartheta}_0$. From the data, a sample covariance matrix is obtained by

$$\begin{aligned}\hat{\mathbf{R}} &= \begin{pmatrix} \hat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1} & \hat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_2} \\ \hat{\mathbf{R}}_{\mathbf{x}_2\mathbf{x}_1} & \hat{\mathbf{R}}_{\mathbf{x}_2\mathbf{x}_2} \end{pmatrix} \\ &= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1i}^T & \mathbf{x}_{2i}^T \end{pmatrix}. \end{aligned} \tag{15}$$

Another covariance matrix is derived from the model

$$\begin{aligned}\mathbf{R}(\boldsymbol{\vartheta}) &= \begin{pmatrix} \mathsf{E}\left\{\mathbf{x}_1\mathbf{x}_1^T\right\} & \mathsf{E}\left\{\mathbf{x}_1\mathbf{x}_2^T\right\} \\ \mathsf{E}\left\{\mathbf{x}_2\mathbf{x}_1^T\right\} & \mathsf{E}\left\{\mathbf{x}_2\mathbf{x}_2^T\right\} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{\Lambda}_1\mathbf{A}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}^T + \mathbf{\Psi}_\delta)\mathbf{A}^T\mathbf{\Lambda}_1^T + \mathbf{\Psi}_{\epsilon_1} & \mathbf{\Lambda}_1\mathbf{A}\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Lambda}_2^T \\ \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Gamma}^T\mathbf{A}^T\mathbf{\Lambda}_1^T & \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_2^T + \mathbf{\Psi}_{\epsilon_2} \end{pmatrix}, \end{aligned} \tag{16}$$

where $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$. The matrix in (16) is known as the *model implied covariance matrix*. The estimation problem is to determine the vector $\boldsymbol{\vartheta}$ that is compatible with the observations in the sense that

$$\hat{\mathbf{R}} \approx \mathbf{R}(\boldsymbol{\vartheta}). \tag{17}$$

One has to consider that both sides of (17) are symmetric matrices, and hence the effective number of covariance elements to approximate is (at most)

$$n^* = \frac{1}{2}n(n+1), \tag{18}$$

where

$$n = n_{x_1} + n_{x_2} \tag{19}$$

is the total number of observed quantities. Depending on the problem, there are cases in which the number of unique elements in $\mathbf{R}(\boldsymbol{\vartheta})$ is indeed smaller than n^*. For instance, if $\mathbf{R}(\boldsymbol{\vartheta})$ is fully or partly Toeplitz. For the estimation problem to be feasible, it is required that n^* is at least as large as the number of elements in $\boldsymbol{\vartheta}$.

Assuming the estimation problem is feasible, one defines an estimation of $\boldsymbol{\vartheta}_0$ as the minimizing value of some criterion function, expressing how much $\hat{\mathbf{R}}$ differs from $\mathbf{R}(\boldsymbol{\vartheta})$. Generally, one writes such a criterion function (typically called a *fit function* in the SEM literature) as

$$V(\boldsymbol{\vartheta}) = f\big(\hat{\mathbf{R}}, \mathbf{R}(\boldsymbol{\vartheta})\big). \tag{20}$$

There are several ways for how to formulate $V(\boldsymbol{\vartheta})$. In SEM, the most prominent criterion is

$$V_1(\boldsymbol{\vartheta}) = \log\left\{\det\big(\mathbf{R}(\boldsymbol{\vartheta})\big)\right\} + \operatorname{tr}\left\{\hat{\mathbf{R}}\mathbf{R}^{-1}(\boldsymbol{\vartheta})\right\}, \tag{21}$$

where det denotes the determinant of a matrix and tr denotes the trace. If the data are *Gaussian and independently distributed*, minimizing $V_1(\boldsymbol{\vartheta})$ gives the maximum likelihood (ML) estimator (based on the information in $\hat{\mathbf{R}}$); see Jöreskog (1967, 1970). Note that $V_1(\boldsymbol{\vartheta})$ may also be used for other distributions of the data, even if it no longer leads to the ML estimator. Some caution has to be exercised when applying $V_1(\boldsymbol{\vartheta})$. It is necessary to constrain the elements of $\boldsymbol{\vartheta}$ during the numerical minimization such that the matrix $\mathbf{R}(\boldsymbol{\vartheta})$ remain positive definite. If this is not considered, the numerical search may easily diverge and not lead to an appropriate estimate.

A second possibility is to consider the criterion

$$V_2(\boldsymbol{\vartheta}) = \operatorname{tr}\left\{\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big)\mathbf{Q}_1\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big)\mathbf{Q}_2\right\}, \tag{22}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are symmetric (user-chosen) positive definite weighting matrices that are either fixed or random. By suitable choices of $\mathbf{Q}_1$ and $\mathbf{Q}_2$, a few well-known estimators can be derived. For instance, when $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{I}$, the resulting estimator becomes *Unweighted Least Squares* (ULS), whereas the choice $\mathbf{Q}_1 = \mathbf{Q}_2 = \hat{\mathbf{R}}^{-1}$ leads to an estimator known as *Generalized Least Squares* (GLS). An early treatment of the GLS estimator is provided in Jöreskog and Goldberger (1972). Yet, another estimator is obtained by letting $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{R}^{-1}(\boldsymbol{\vartheta})$, in which case $V_2(\boldsymbol{\vartheta})$ forms a *Re-weighted Least Squares* (RLS) estimator. The term 're-weighted' reflects that the weighting matrix is iteratively updated during the numerical search.

Finally, a third alternative is to vectorize the difference between the covariance matrices. This development is due to Browne (1974), where it is shown that (22) is a special case of a more generic criterion formed by

$$V_3(\boldsymbol{\vartheta}) = \tilde{\mathbf{r}}^T(\boldsymbol{\vartheta})\mathbf{W}\tilde{\mathbf{r}}(\boldsymbol{\vartheta}). \tag{23}$$

In this expression,

$$\begin{aligned}\tilde{\mathbf{r}}(\boldsymbol{\vartheta}) &= \operatorname{vech}\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big) \\ &= \mathbf{K}_n^T\operatorname{vec}\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big), \end{aligned} \tag{24}$$

and $\mathbf{W}$ is a positive definite weighting matrix of appropriate dimension. In (24), vech denotes the operation of vectorizing the lower triangular part (including the diagonal) of a square matrix, whereas vec denotes the operation of vectorizing the full matrix. Moreover, $\mathbf{K}_n$ is a matrix of dimension $n^2 \times n^*$ with rank n^*. Due to its functional form, $V_3(\boldsymbol{\vartheta})$ is often labeled *Weighted Least Squares* (WLS).

It follows from (24) that the construction of $\mathbf{K}_n$ is such that $\tilde{\mathbf{r}}(\boldsymbol{\vartheta})$ only contains the n^* non-repeated elements of $\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})$. By this, the redundancy owing to the symmetry of $\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})$ is effectively removed. Since $\mathbf{K}_n$ has full column rank, a left inverse can be obtained by $\mathbf{K}_n^- = \big(\mathbf{K}_n^T\mathbf{K}_n\big)^{-1}\mathbf{K}_n^T$. A desirable property of $\mathbf{K}_n$ is

$$\operatorname{vec}\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big) = \mathbf{K}_n^{-T}\operatorname{vech}\big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\vartheta})\big). \tag{25}$$

Remark 3.1. If the considered matrix is symmetric and no other restrictions have been placed on its elements, the result of applying vech operation is an n^*-dimensional vector consisting of the non-repeated elements. If the considered matrix is also Toeplitz (or partly Toeplitz), it is characterized by fewer unique elements. For later purposes, it will be useful to switch to a more general notation represented by vecnr. We define vecnr as the operation of vectorizing only the non-repeated elements. It follows that $\mathbf{K}_n$ may have to be modified. In Appendix A, we derive expressions for $\mathbf{K}_n$ for various forms of the considered matrix.

The following example demonstrates the use and properties of $\mathbf{K}_n$ when the considered matrix is symmetric and no other restrictions have been placed on its elements.

Example 3.1. Let $\mathbf{R}$ be a matrix of dimension $n = 3$. Vectorizing the non-repeated elements of $\mathbf{R}$ gives

$$\begin{aligned}\operatorname{vecnr}(\mathbf{R}) &= \mathbf{K}_3^T \operatorname{vec}(\mathbf{R}) \\ &= \begin{pmatrix} r_{11} & r_{21} & r_{31} & r_{22} & r_{32} & r_{33} \end{pmatrix}^T. \end{aligned} \tag{26}$$

Let $\mathbf{K}_3$ be written in the rather general form

$$\mathbf{K}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_2 & 0 & 0 & 0 \\ 0 & 1-\alpha_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 1-\alpha_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-\alpha_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{27}$$

Note that any value of α_1, α_2 and α_3 will produce (26), but only $\alpha_1 = \alpha_2 = \alpha_3 = 0.5$ will lead to the property described in (25). It also holds that

$$\mathbf{K}_3^{-T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{28}$$

The generality of $V_3(\boldsymbol{\vartheta})$ is largely due to the various forms the weighting matrix $\mathbf{W}$ can take. We now demonstrate how $V_3(\boldsymbol{\vartheta})$ and $V_2(\boldsymbol{\vartheta})$ are related through the choice of $\mathbf{W}$. Before doing so, it is be useful to introduce the following lemma:

Lemma 3.1. *Let* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *and* $\mathbf{D}$ *be matrices such that* $\mathbf{ABCD}$ *is well defined and square. Then it holds that*

$$\operatorname{tr}(\mathbf{ABCD}) = \operatorname{vec}^T(\mathbf{D})\left(\mathbf{A} \otimes \mathbf{C}^T\right)\operatorname{vec}\left(\mathbf{B}^T\right). \tag{29}$$

See Bernstein (2009).

Consider the following expression for the weighting matrix

$$\mathbf{W} = \mathbf{K}_n^{-}(\mathbf{Q}_1 \otimes \mathbf{Q}_2)\mathbf{K}_n^{-T}. \tag{30}$$

By applying the lemma stated in (29), it can be shown that $V_3(\boldsymbol{\vartheta})$ coincides with $V_2(\boldsymbol{\vartheta})$. It immediately follows that ULS, GLS and RLS are all special cases of WLS, obtained from specific choices of $\mathbf{W}$.

Remark 3.2. In terms of implementation, the quadratic form of the WLS criteria allows it to be formulated as a *Separable Least Squares* problem; see Golub and Pereyra (1973, 2003). In Appendix B, it is demonstrated how this can be done when the considered model takes the form of a CFA model.

Below, we briefly summarize some important statistical properties of the listed criteria. If the covariance structure is correctly specified and the conditions for identifiability are satisfied, the following general results are reported to hold; see also Bartholomew et al. (2011) and Bollen (1989):

- Parameters are consistently estimated.
- Asymptotic efficiency is achieved when the data are Gaussian and independently distributed.
- The WLS estimator is asymptotically efficient without the Gaussian assumption, provided that the weighting matrix takes the form

$$\mathbf{W} = \left(\operatorname{cov}(\tilde{\mathbf{r}})\right)^{-1}. \tag{31}$$

The WLS estimator in the non-Gaussian case is treated in Browne (1984).

3.3. Estimation algorithms

Different non-linear optimization schemes may be considered for minimizing the various criterion functions. Numerical procedures such as *Quasi-Newton* methods are frequently used. An introduction to the algorithmic aspects of SEM-based estimation with examples can be found in Lee (2007). A number of commercial and free software solutions that apply these algorithms are available.

4. Applying structural equation modeling for EIV system identification

Although SEM is typically applied to static problems, frequent attempts have been made to fit dynamic models into SEM. For instance, in van Buuren (1997) and Hamaker, Dolan, and Molenaar (2002), it is demonstrated how ARMA processes can be formulated as SEMs. An extension of this work to the multivariate case is given in du Toit and Browne (2007). Another development, which has been growing in interest in recent years, is represented by the *Dynamic Factor Analysis* (DFA) model. This development is mainly due to Browne and Nesselroade (2005), Browne and Zhang (2007), Molenaar (1985) and Molenaar and Nesselroade (1998). The DFA model allows for time dependent latent factors, a feature shared with the type of models considered here. Also, in Geweke and Singleton (1981), a frequency domain approach to CFA modeling is presented.

The usual implementation of dynamic models in SEM involves using a sample covariance matrix $\hat{\mathbf{R}}$ which is fully or partly Toeplitz. Models are then fitted into SEM by appropriately constraining the elements of the parameter matrices. In what follows, we consider two different formulations of the EIV SISO system within SEM.

4.1. First formulation

In the first formulation, it suffices to work with only part of the SEM framework. The parameterization represented by this formulation is closely related to the covariance matching approach (Söderström et al., 2009).

For later use, let the covariance function for the general random process $x(t)$ be defined as

$$r_x(\tau) = \mathsf{E}\left\{x(t+\tau)x(t)\right\}. \tag{32}$$

To simplify notation, it is useful to introduce the random vectors

$$\mathbf{y}(t) = \begin{pmatrix} y(t) & \cdots & y(t-n_y+1) \end{pmatrix}^T, \tag{33}$$

$$\mathbf{u}(t) = \begin{pmatrix} u(t-1) & \cdots & u(t-n_u) \end{pmatrix}^T, \tag{34}$$

wherein (33), n_y denotes the number of y-elements starting at lag 0, and in (34), n_u denotes the number of u-elements starting at lag 1. Note that n_y and n_u are values to be chosen by the user. Vectors of the undisturbed signals, denoted $\mathbf{y}_0(t)$ and $\mathbf{u}_0(t)$, and vectors of the measurement noises, denoted $\tilde{\mathbf{y}}(t)$ and $\tilde{\mathbf{u}}(t)$, are formed by expressions similar to (33) and (34).

Now, define the auxiliary process

$$z_0(t) = \frac{1}{A(q^{-1})}u_0(t). \tag{35}$$

Using (35), $y(t)$ and $u(t)$, as given in (4) and (5), can be expressed as

$$y(t) = B(q^{-1})z_0(t) + \tilde{y}(t), \tag{36}$$

$$u(t) = A(q^{-1})z_0(t) + \tilde{u}(t). \tag{37}$$

To facilitate the SEM representation of the system, use (36) and (37) to obtain the following system of equations

$$\begin{aligned} y(t) &= B(q^{-1})z_0(t) + \tilde{y}(t) \\ &\vdots \\ y(t-n_y+1) &= B(q^{-1})z_0(t-n_y+1) + \tilde{y}(t-n_y+1) \\ u(t-1) &= A(q^{-1})z_0(t-1) + \tilde{u}(t-1) \\ &\vdots \\ u(t-n_u) &= A(q^{-1})z_0(t-n_u) + \tilde{u}(t-n_u). \end{aligned} \tag{38}$$

The system in (38) can be written as a CFA model in the form of (10) (here given without subscripts, but with time indexing on the random vectors)

$$\mathbf{x}(t) = \boldsymbol{\Lambda}\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}(t). \tag{39}$$

To do so, let the vector of the observed quantities be given by

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{y}^T(t) & \mathbf{u}^T(t) \end{pmatrix}^T, \tag{40}$$

where the dimension of $\mathbf{x}(t)$ equals

$$n = n_y + n_u. \tag{41}$$

Moreover, let the vectors of the unobserved quantities be

$$\boldsymbol{\xi}(t) = \begin{pmatrix} z_0(t-1) & \cdots & z_0(t-k-1) \end{pmatrix}^T, \tag{42}$$

$$\boldsymbol{\epsilon}(t) = \begin{pmatrix} \tilde{\mathbf{y}}^T(t) & \tilde{\mathbf{u}}^T(t) \end{pmatrix}^T, \tag{43}$$

wherein (42), k is determined by

$$k = \max\left(n_y + n_b - 1, n_u + n_a\right) - 1. \tag{44}$$

The parameter matrix relating $\boldsymbol{\xi}(t)$ to $\mathbf{x}(t)$ takes the form of a Sylvester matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{M}_b \\ \mathbf{M}_a \end{pmatrix}, \tag{45}$$

where the partitions $\mathbf{M}_b$ and $\mathbf{M}_a$ depend on the relative magnitude of $n_y + n_b - 1$ and $n_u + n_a$. To further describe $\boldsymbol{\Lambda}$, let

$$v = n_y - n_u - 1, \tag{46}$$

such that the difference between $n_y + n_b - 1$ and $n_u + n_a$ can be expressed as

$$v^* = v + n_b - n_a. \tag{47}$$

Two cases of (45) are relevant to consider.

Case 1: When $v^* \geq 0$, $\boldsymbol{\Lambda}$ takes the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} b_1 & \cdots & \cdots & b_{n_b} & 0 & \cdots & 0 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & b_1 & \cdots & \cdots & b_{n_b} \\ 1 & \cdots & a_{n_a} & 0 & \cdots & 0 & \mathbf{0} \\ 0 & \ddots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & a_{n_a} & \mathbf{0} \end{pmatrix}, \tag{48}$$

where the last column of the lower partition $\mathbf{M}_a$ consists of v^*-dimensional row vectors of zeros.

Case 2: In the reversed case when $v^* \leq 0$, $\boldsymbol{\Lambda}$ becomes

$$\boldsymbol{\Lambda} = \begin{pmatrix} b_1 & \cdots & b_{n_b} & 0 & \cdots & 0 & \mathbf{0} \\ 0 & \ddots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & b_1 & \cdots & b_{n_b} & \mathbf{0} \\ 1 & \cdots & \cdots & a_{n_a} & 0 & \cdots & 0 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & \cdots & a_{n_a} \end{pmatrix}, \tag{49}$$

where the last column of the upper partition $\mathbf{M}_b$ consists of row vectors of zeros with dimension $-v^*$. Obviously, in the special case when $v^* = 0$, (48) and (49) coincides to form the same matrix. Irrespective of form, the dimension of $\boldsymbol{\Lambda}$ is $n \times (k+1)$.

To fully describe the system, we include the following covariance matrices

$$\boldsymbol{\Phi} = \begin{pmatrix} r_{z_0}(0) & \cdots & r_{z_0}(k) \\ \vdots & \ddots & \vdots \\ r_{z_0}(k) & \cdots & r_{z_0}(0) \end{pmatrix}, \tag{50}$$

$$\boldsymbol{\Psi}_\epsilon = \begin{pmatrix} \psi_{\tilde{y}}\mathbf{I}_{n_y} & \mathbf{0} \\ \mathbf{0} & \psi_{\tilde{u}}\mathbf{I}_{n_u} \end{pmatrix}, \tag{51}$$

where the dimensions immediately follow. Since $\boldsymbol{\Phi}$ is Toeplitz, a vector of the non-repeated elements is given by

$$\mathbf{r}_\xi = \begin{pmatrix} r_{z_0}(0) & \cdots & r_{z_0}(k) \end{pmatrix}^T. \tag{52}$$

The parameter vector to be estimated from the data becomes

$$\boldsymbol{\vartheta}_0 = \begin{pmatrix} \boldsymbol{\theta}_0^T & \mathbf{r}_{\xi,0}^T & \boldsymbol{\psi}_0^T \end{pmatrix}^T, \tag{53}$$

and the system implied covariance matrix from (16) simplifies to

$$\mathbf{R}(\boldsymbol{\vartheta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}_\epsilon. \tag{54}$$

We finalize the description by adding a few comments:

- Due to the block Toeplitz form, $\mathbf{R}(\boldsymbol{\vartheta})$ contains $2n-1$ non-repeated elements.
- When estimation is performed using $V_3(\boldsymbol{\vartheta})$, the redundancy originating from the block Toeplitz form of $\mathbf{R}(\boldsymbol{\vartheta})$ can be eliminated by appropriately modifying $\mathbf{K}_n$ (see Appendix A for details).
- With reference to the previous point, to have at least as many covariance elements to approximate as there are free parameters to estimate, n_y and n_u must be chosen such that $2n - 1 \geq n_\vartheta = n_\theta + k + 3$.

4.2. Second formulation

We now consider a formulation in which the noise-free output and input signals $y_0(t)$ and $u_0(t)$ comprise the latent quantities in the specification. The second formulation is first derived using the complete SEM. It is then shown how the resulting expressions can be simplified to admit a CFA representation. As before, let n_y and n_u be parameters chosen by the user.

First, we describe the structural equation. Expand (1) to obtain a system of equations

$$\begin{aligned} A(q^{-1})y_0(t) &= B(q^{-1})u_0(t) \\ &\vdots \\ A(q^{-1})y_0(t-n_y+1) &= B(q^{-1})u_0(t-n_y+1). \end{aligned} \tag{55}$$

The system in (55) is now written as a structural equation in the form of (8). To this aim, let $\boldsymbol{\delta}(t)$ be an n_y-dimensional vector of zeros so that (8) becomes

$$\boldsymbol{\eta}(t) = \mathbf{B}\boldsymbol{\eta}(t) + \boldsymbol{\Gamma}\boldsymbol{\xi}(t), \tag{56}$$

where

$$\boldsymbol{\eta}(t) = \mathbf{y}_0(t), \qquad \boldsymbol{\xi}(t) = \begin{pmatrix}\mathbf{u}_0^T(t) & \boldsymbol{\nu}^T(t)\end{pmatrix}^T. \tag{57}$$

In the latter expression of (57),

$$\boldsymbol{\nu}(t) = \begin{pmatrix}\nu_1(t) & \cdots & \nu_{n_\nu}(t)\end{pmatrix}^T \tag{58}$$

denotes a sequence of auxiliary processes which capture the dynamics of the undisturbed output and input signals prior to $t - n_y + 1$ and $t - n_u$, respectively. The appropriate number of auxiliary processes to include is

$$\begin{aligned} n_\nu &= \max(n_y - n_u - 1 + n_b, n_a) \\ &= \max(\nu^*, 0) + n_a. \end{aligned} \tag{59}$$

Recall that ν^* is determined by (47). The auxiliary processes must take a form such that the structural equation becomes consistent with the system in (55). Introduce

$$\kappa_j = n_y - n_\nu + j - 1, \quad j = 1, \ldots, n_\nu. \tag{60}$$

The auxiliary processes are then described by

$$\begin{aligned} \nu_j(t) = &- \sum_{i=1}^{n_a} \mathbb{1}_{\{\kappa_j + i \geq n_y\}} a_i y_0(t - \kappa_j - i) \\ &+ \sum_{i=1}^{n_b} \mathbb{1}_{\{\kappa_j + i \geq n_u + 1\}} b_i u_0(t - \kappa_j - i), \end{aligned} \tag{61}$$

where $\mathbb{1}_{\{A\}}$ is the indicator function of the set A, which takes the value 1 if A is true and 0 otherwise.

The specification of the structural equation additionally involves describing the parameter matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$. The n_y-dimensional square matrix relating $\boldsymbol{\eta}(t)$ to itself is given by

$$\mathbf{B} = \begin{pmatrix} 0 & -a_1 & \cdots & -a_{n_a} & 0 & \cdots & 0 \\ & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & 0 \\ & & & & \ddots & & -a_{n_a} \\ \vdots & & & & \ddots & \ddots & \vdots \\ & & & & & & -a_1 \\ 0 & & \cdots & & \cdots & & 0 \end{pmatrix}. \tag{62}$$

A general expression for the matrix relating $\boldsymbol{\xi}(t)$ to $\boldsymbol{\eta}(t)$ is

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_b & \underset{(n_y - n_\nu \times n_\nu)}{\mathbf{0}} \\ & \underset{n_\nu}{\mathbf{I}} \end{pmatrix}, \tag{63}$$

where the partition $\boldsymbol{\Gamma}_b$ is a matrix of dimension $n_y \times n_u$ whose form depends on the difference between n_y and n_u. It follows that the dimension of $\boldsymbol{\Gamma}$ is $n_y \times (n_u + n_\nu)$.

Two cases of $\boldsymbol{\Gamma}_b$ are relevant to consider.

Case 1: When $n_y - n_u \geq 0$, $\boldsymbol{\Gamma}_b$ is formed by

$$\boldsymbol{\Gamma}_b = \begin{pmatrix} b_1 & \cdots & b_{n_b} & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ & & \ddots & & \ddots & 0 \\ \vdots & & \ddots & \ddots & & b_{n_b} \\ & & & & \ddots & \vdots \\ 0 & & \cdots & & 0 & b_1 \\ \mathbf{0} & & \cdots & \cdots & & \mathbf{0} \end{pmatrix}, \tag{64}$$

where the last row consists of column vectors of zeros with dimension $n_y - n_u$.

Case 2: In the reversed case when $n_y - n_u \leq 0$, $\boldsymbol{\Gamma}_b$ becomes

$$\boldsymbol{\Gamma}_b = \begin{pmatrix} b_1 & \cdots & b_{n_b} & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ & \ddots & & & \ddots & 0 \\ \vdots & \ddots & \ddots & & & b_{n_b} \\ & & & \ddots & & \vdots \\ 0 & \cdots & & 0 & b_1 & b_{-\upsilon} \end{pmatrix}, \tag{65}$$

where υ follows from (46). In the special case when $n_y = n_u$, the expressions in (64) and (65) coincides to form the same matrix.

Second, to obtain the measurement equations, expand (4) and (5) into the following two equation systems

$$\begin{aligned} y(t) &= y_0(t) + \tilde{y}(t) \\ &\vdots \\ y(t - n_y + 1) &= y_0(t - n_y + 1) + \tilde{y}(t - n_y + 1), \\ u(t - 1) &= u_0(t - 1) + \tilde{u}(t - 1) \\ &\vdots \\ u(t - n_u) &= u_0(t - n_u) + \tilde{u}(t - n_u). \end{aligned} \tag{66, 67}$$

The equation systems in (66) and (67) are now written in the form of (9) and (10)

$$\mathbf{x}_1(t) = \boldsymbol{\Lambda}_1\boldsymbol{\eta}(t) + \boldsymbol{\epsilon}_1(t), \tag{68}$$

$$\mathbf{x}_2(t) = \boldsymbol{\Lambda}_2\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}_2(t), \tag{69}$$

where the observed quantities are given by

$$\mathbf{x}_1(t) = \mathbf{y}(t), \qquad \mathbf{x}_2(t) = \mathbf{u}(t), \tag{70}$$

and the noise vectors are

$$\boldsymbol{\epsilon}_1(t) = \tilde{\mathbf{y}}(t), \qquad \boldsymbol{\epsilon}_2(t) = \tilde{\mathbf{u}}(t). \tag{71}$$

As before, the total number of observed quantities is determined by (41). The parameter matrices relating the latent and observed quantities are simply

$$\boldsymbol{\Lambda}_1 = \mathbf{I}_{n_y}, \qquad \boldsymbol{\Lambda}_2 = \begin{pmatrix}\mathbf{I}_{n_u} & \underset{(n_u \times n_\nu)}{\mathbf{0}}\end{pmatrix}. \tag{72}$$

The description includes the following covariance matrices. Let $\mathbf{R}_{\mathbf{u}_0}$ denote the covariance matrix of the latent input signal sequence $\mathbf{u}_0(t)$, and let $\mathbf{R}_{\boldsymbol{\nu}}$ denote the covariance matrix of the auxiliary sequence $\boldsymbol{\nu}(t)$. Then,

$$\boldsymbol{\Phi} = \begin{pmatrix} \underset{(n_u \times n_u)}{\mathbf{R}_{\mathbf{u}_0}} & \underset{(n_u \times n_\nu)}{\mathbf{R}_{\mathbf{u}_0\boldsymbol{\nu}}} \\ \underset{(n_\nu \times n_u)}{\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}} & \underset{(n_\nu \times n_\nu)}{\mathbf{R}_{\boldsymbol{\nu}}} \end{pmatrix}, \tag{73}$$

$$\boldsymbol{\Psi}_{\epsilon_1} = \psi_{\tilde{y}}\mathbf{I}_{n_y}, \qquad \boldsymbol{\Psi}_{\epsilon_2} = \psi_{\tilde{u}}\mathbf{I}_{n_u}, \tag{74}$$

where the dimensions immediately follow. Let $\mathbf{r}_\xi$ be a column vector composed of the non-repeated elements of $\boldsymbol{\Phi}$. If the Toeplitz form of $\mathbf{R}_{\mathbf{u}_0}$ is the only restriction placed on $\boldsymbol{\Phi}$, the dimension of $\mathbf{r}_\xi$ is

$$\begin{aligned} n_{\mathbf{r}_\xi} &= n_u + n_u n_\nu + \frac{1}{2} n_\nu (n_\nu + 1) \\ &= \left(n_u + \frac{1}{2} n_\nu\right)(n_\nu + 1). \end{aligned} \tag{75}$$

The parameter vector to be estimated from the data becomes

$$\boldsymbol{\vartheta}_0 = \begin{pmatrix} \boldsymbol{\theta}_0^T & \mathbf{r}_{\xi,0}^T & \boldsymbol{\psi}_0^T \end{pmatrix}^T. \tag{76}$$

Since the complete SEM framework is used, the system implied covariance matrix $\mathbf{R}(\boldsymbol{\vartheta})$ takes the general form of (16), but with the slight simplification that $\boldsymbol{\Psi}_\delta = \mathbf{0}$.

It may be of interest to present the second formulation in a simpler form. Consider the reduced form of (56)

$$\boldsymbol{\eta}(t) = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi}(t). \tag{77}$$

Substituting (77) into (68) gives

$$\mathbf{x}_1(t) = \boldsymbol{\Lambda}_1(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}_1(t). \tag{78}$$

Stacking $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ allows us to write

$$\begin{pmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{pmatrix} = \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} \\ \boldsymbol{\Lambda}_2 \end{pmatrix} \boldsymbol{\xi}(t) + \begin{pmatrix} \boldsymbol{\epsilon}_1(t) \\ \boldsymbol{\epsilon}_2(t) \end{pmatrix}, \tag{79}$$

from which it follows that the second formulation can be written as a CFA model

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{pmatrix} = \boldsymbol{\Lambda}\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}(t). \tag{80}$$

Obviously, when the system is described by (80), the use of (54) provides the same system implied covariance matrix as when the complete SEM and (16) is used.

As previously, we finalize the description by adding a few comments:

- Estimation using $V_3(\boldsymbol{\vartheta})$ involves $n^* > 2n - 1$ covariance elements. Thus, the redundancy due to the block Toeplitz form of $\mathbf{R}(\boldsymbol{\vartheta})$ is preserved.
- To have at least as many covariance elements to approximate as there are free parameters to estimate, n_y and n_u must be chosen such that $n^* \geq n_\vartheta = n_\theta + n_{r_\xi} + 2$.
- The choice of n_y and n_u should additionally satisfy

$$n_y - n_a - 1 \geq 0, \qquad n_u - n_b \geq 0, \tag{81}$$

$$n_y - 1 \geq n_u - n_b. \tag{82}$$

  The conditions in (81) ensure that all parameters are present in the specification. The condition in (82) ensures that all u_0-elements in $\boldsymbol{\xi}(t)$ are related to at least one y_0-element in $\boldsymbol{\eta}(t)$.
- Improved parsimony can be achieved by additionally constraining the elements of $\boldsymbol{\Phi}$. In Appendix C, it is demonstrated how such constraints can be imposed.

4.3. The relation between the two formulations

The difficulty of applying SEM to the EIV problem is to account for the transient effects of the undisturbed output and input signals prior to $t - n_y + 1$ and $t - n_u$, respectively. As shown, this difficulty can be handled by introducing latent auxiliary processes into the specification. The proposed formulations are distinguished by the way these auxiliary processes are specified. In the first formulation, the auxiliary process of (35) imposes additional structure on the parameterization such that $\boldsymbol{\Phi}$ becomes Toeplitz. This is in contrast to the second formulation where the auxiliary processes are allowed to freely correlate with any other process in $\boldsymbol{\xi}(t)$, leading to a larger number of free elements in $\boldsymbol{\Phi}$. Despite these differences, given the same choice of n_y and n_u, the auxiliary processes are such that the two formulations are covariance equivalent. That is, the covariance functions comprising the elements of $\mathbf{R}(\boldsymbol{\vartheta})$ map the same system covariance matrix in both cases. If $\boldsymbol{\vartheta}_0$ is uniquely identifiable, it follows that the two formulations are asymptotically equivalent.

Imposing similar structure on the parameterization of the second formulation as on the first one, it is possible to show their connection. To show this, let $\boldsymbol{\xi}^{(1)}(t)$ and $\boldsymbol{\xi}^{(2)}(t)$ denote $\boldsymbol{\xi}(t)$ of the first and second formulation, respectively. Then for a given choice of n_y, n_u, and using

$$y_0(t) = B(q^{-1})z_0(t), \qquad u_0(t) = A(q^{-1})z_0(t), \tag{83}$$

it follows that

$$\boldsymbol{\eta}(t) = \mathbf{M}_b\boldsymbol{\xi}^{(1)}(t), \qquad \mathbf{u}_0(t) = \mathbf{M}_a\boldsymbol{\xi}^{(1)}(t). \tag{84}$$

The measurement equations given in (68) and (69) become

$$\begin{aligned} \mathbf{x}_1(t) &= \boldsymbol{\Lambda}_1\mathbf{M}_b\boldsymbol{\xi}^{(1)}(t) + \boldsymbol{\epsilon}_1(t) \\ &= \mathbf{M}_b\boldsymbol{\xi}^{(1)}(t) + \boldsymbol{\epsilon}_1(t), \end{aligned} \tag{85}$$

$$\begin{aligned} \mathbf{x}_2(t) &= \boldsymbol{\Lambda}_2 \begin{pmatrix} \mathbf{M}_a\boldsymbol{\xi}^{(1)}(t) \\ \boldsymbol{\nu}(t) \end{pmatrix} + \boldsymbol{\epsilon}_2(t) \\ &= \mathbf{M}_a\boldsymbol{\xi}^{(1)}(t) + \boldsymbol{\epsilon}_2(t), \end{aligned} \tag{86}$$

and it is clear that

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{pmatrix} = \boldsymbol{\Lambda}\boldsymbol{\xi}^{(1)}(t) + \boldsymbol{\epsilon}(t), \tag{87}$$

where $\boldsymbol{\Lambda}$ takes the form of (45).

It may additionally be of interest to see how the elements of $\boldsymbol{\xi}^{(2)}(t)$ can be expressed in terms of the elements of $\boldsymbol{\xi}^{(1)}(t)$. Before presenting the final result, it is useful to introduce some intermediate results.

First, note that $\boldsymbol{\xi}^{(1)}(t)$ and $\boldsymbol{\xi}^{(2)}(t)$ are vectors of equal dimension. This is verified by considering (44) and (59) in the following way

$$\begin{aligned} n_{\xi^{(1)}} &= k + 1 \\ &= \max(n_y + n_b - 1, n_u + n_a) \\ &= \max(v^*, 0) + n_a + n_u \\ &= n_{\xi^{(2)}}. \end{aligned} \tag{88}$$

Second, as an alternative to (61), the auxiliary vector $\boldsymbol{\nu}(t)$ can be obtained from the structural equation

$$\begin{aligned} \boldsymbol{\eta}(t) &= \mathbf{B}\boldsymbol{\eta}(t) + \boldsymbol{\Gamma}\boldsymbol{\xi}^{(2)}(t) \\ &= \mathbf{B}\boldsymbol{\eta}(t) + \boldsymbol{\Gamma}_b\mathbf{u}_0(t) + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}(t) \end{pmatrix}, \end{aligned} \tag{89}$$

which is re-arranged to give

$$\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}(t) \end{pmatrix} = (\mathbf{I} - \mathbf{B})\boldsymbol{\eta}(t) - \boldsymbol{\Gamma}_b\mathbf{u}_0(t). \tag{90}$$

Let $(\mathbf{I} - \mathbf{B})^*$ and $\boldsymbol{\Gamma}_b^*$ denote two matrices composed of the last n_ν rows of $\mathbf{I} - \mathbf{B}$ and $\boldsymbol{\Gamma}_b$, respectively. Then from (90), one can write

$$\boldsymbol{\nu}(t) = (\mathbf{I} - \mathbf{B})^*\boldsymbol{\eta}(t) - \boldsymbol{\Gamma}_b^*\mathbf{u}_0(t). \tag{91}$$

It is straightforward to verify the correspondence between (61) and (91).

We are now in a position to establish the direct relationship between $\boldsymbol{\xi}^{(1)}(t)$ and $\boldsymbol{\xi}^{(2)}(t)$. Define

$$\mathbf{J}_1 = \begin{pmatrix} \mathbf{0}_{(n_u \times n_y)} \\ (\mathbf{I} - \mathbf{B})^* \end{pmatrix}, \qquad \mathbf{J}_2 = \begin{pmatrix} \mathbf{I}_{n_u} \\ -\boldsymbol{\Gamma}_b^* \end{pmatrix}. \tag{92}$$

Then,

$$\begin{aligned} \boldsymbol{\xi}^{(2)}(t) &= \mathbf{J}_1\boldsymbol{\eta}(t) + \mathbf{J}_2\mathbf{u}_0(t) \\ &= (\mathbf{J}_1\mathbf{M}_b + \mathbf{J}_2\mathbf{M}_a)\,\boldsymbol{\xi}^{(1)}(t) \\ &\triangleq \mathbf{P}\boldsymbol{\xi}^{(1)}(t), \end{aligned} \tag{93}$$

where the second equality follows from the use of (84). The dimensions of the involved matrices are such that $\mathbf{P}$ has dimension $n_\xi \times n_\xi$.

5. Simulation examples

By simulation examples, we evaluate the performance of the different SEM-based estimators discussed in Section 3 for the two formulations presented in Section 4. As a reference, we include the CM estimator (Söderström et al., 2009). The SEM-based estimators are computed using the commercial software LISREL (Jöreskog & Sörbom, 1996), whereas the CM estimator is programmed in Matlab. Some algorithmic aspects of the implementation will also be considered.

5.1. Simulation setup

The default minimization routine in LISREL is the Davidon–Fletcher–Powell (DFP) algorithm. The implementation of the algorithm allows the user to optimize performance by adjusting the default settings; see Jöreskog and Sörbom (1996) for additional details. Minimization in Matlab is performed using the Nelder–Mead simplex algorithm (Mathworks, 2014). Initial values are derived by first applying the Frisch estimator to obtain $\boldsymbol{\theta}^{(0)}$ (Beghelli et al., 1990). Then in a second step, the remaining elements of $\boldsymbol{\vartheta}^{(0)}$ are found by the procedures outlined in Appendix B.

The LISREL software comes with several built-in features to ensure the stability (or admissibility) of the final solution. These features may cause the program to issue warning statements, indicating either non-convergence (NC) or problems of identifying specific system parameters. Warnings may also be related to the definiteness of the model covariance matrices. In the simulation setup, warning statements are handled according to the rules: if convergence is not reached within a fixed number of iterations, or if $\boldsymbol{\Psi}_\epsilon$ (alternatively, $\boldsymbol{\Psi}_{\epsilon_1}$ and $\boldsymbol{\Psi}_{\epsilon_2}$) is not positive definite, the solution is deemed *improper* and is discarded. We treat the elements of $\boldsymbol{\Phi}$ as auxiliary parameters and ignore any warnings related to the definiteness of this matrix. Note that even if the estimate of $\boldsymbol{\Phi}$ is not positive definite, the remaining parameter estimates may still be of sufficient quality. The Matlab implementation of the CM approach is equipped with similar features.

The considered system is second order and takes the form

$$y_0(t) = \frac{1.0q^{-1} + 0.5q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}} u_0(t). \tag{94}$$

The input signal is described by the ARMA process

$$u_0(t) = \frac{1 + 0.7q^{-1}}{1 - 0.5q^{-1}} e(t), \tag{95}$$

where $e(t)$ is a zero mean white noise process with unit variance. The variance of the white measurement noises are

$$\psi_{\tilde{y}} = 10, \qquad \psi_{\tilde{u}} = 1, \tag{96}$$

and the user choices are

$$n_y = 6, \qquad n_u = 5. \tag{97}$$

It follows from (46) and (47) that $v^* = v = 0$. The user choices for the CM approach are

$$p_y = p_2 = 5, \qquad p_u = -p_1 = 4; \tag{98}$$

see Söderström et al. (2009) for more details. This choice ensures that the same covariance elements are used across the two approaches.

The estimators considered in the examples are summarized as follows:

Table 1
Simulated means and standard deviations for the 1st formulation.

Method		V_1	GLS	ULS	CM
Param.	True value				
a_1	−1.5	−1.5006 ±0.0228	−1.5003 ±0.0230	−1.5006 ±0.0219	−1.5004 ±0.0219
a_2	0.7	0.7005 ±0.0179	0.7003 ±0.0180	0.7005 ±0.0173	0.7004 ±0.0173
b_1	1.0	0.9933 ±0.0981	0.9935 ±0.0985	0.9968 ±0.0954	0.9962 ±0.0977
b_2	0.5	0.5023 ±0.1427	0.5038 ±0.1437	0.4989 ±0.1363	0.5001 ±0.1382
$\psi_{\tilde{y}}$	10.0	9.9974 ±0.5837	9.8640 ±0.5803	9.9961 ±0.5813	9.9963 ±0.5779
$\psi_{\tilde{u}}$	1.0	0.9758 ±0.1324	0.9722 ±0.1319	0.9852 ±0.1650	0.9894 ±0.1752
(a) Main parameters					
$r_{z_0}(0)$	73.766	74.431 ±9.853	74.265 ±9.859	74.374 ±9.592	74.303 ±9.526
$r_{z_0}(1)$	68.077	68.680 ±9.233	68.524 ±9.239	68.629 ±8.990	68.563 ±8.923
$r_{z_0}(2)$	53.021	53.461 ±7.599	53.331 ±7.604	53.419 ±7.402	53.367 ±7.341
$r_{z_0}(3)$	33.148	33.362 ±5.572	33.271 ±5.575	33.332 ±5.442	33.297 ±5.402
$r_{z_0}(4)$	13.242	13.209 ±4.216	13.163 ±4.214	13.192 ±4.173	13.176 ±4.162
$r_{z_0}(5)$	−3.022	−3.294 ±4.393	−3.294 ±4.385	−3.291 ±4.404	−3.288 ±4.406
$r_{z_0}(6)$	−13.644	−14.115 ±5.306	−14.079 ±5.292	−14.085 ±5.296	−14.069 ±5.296
(b) Auxiliary parameters					

- The minimizer of the criterion $V_1(\boldsymbol{\vartheta})$ from (21), implemented in LISREL. This estimator is labeled V_1 (note that ML is not an appropriate label since the data are not independent).
- The minimizer of the criterion $V_3(\boldsymbol{\vartheta})$ from (23) using $\mathbf{Q}_1 = \mathbf{Q}_2 = \hat{\mathbf{R}}^{-1}$ in (30), implemented in LISREL. This estimator is labeled GLS.
- The minimizer of the criterion $V_3(\boldsymbol{\vartheta})$ from (23) using $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{I}$ in (30), implemented in LISREL. This estimator is labeled ULS.
- The CM estimator, implemented in Matlab. For simplicity no weighting is applied. This estimator is labeled CM.

To obtain the empirical means and standard deviations of the parameter estimates, the system is simulated $n_{Rep} = 1000$ times with sample length $N = 1000$. All estimators are subject to the same data records.

5.2. Examples

Example 5.1. In the first example, the implementation of the first formulation is considered. For this case, $k = 6$ and hence the number of auxiliary processes in $\boldsymbol{\xi}(t)$ is 7.

The simulation results are summarized in Table 1. From the table, it is evident that the different estimation procedures behave rather similarly. In all cases, the mean parameter estimates are close to the true parameter values, and the biases are considerably smaller than the standard deviations. Although the differences are marginal, it is observed that the realized accuracy of the estimates, as measured by the standard deviation, tend to be better for ULS and CM. For any of the estimation procedures, no negative estimates of the measurement noise variances were recorded.

We briefly comment on the numerical performance of the optimization algorithms applied in this study. No occurrences of

Table 2
Simulated means and standard deviations for the 2nd formulation.

Method		V_1	GLS	ULS
Param.	True value			
a_1	-1.5	-1.5004 ± 0.0229	-1.4999 ± 0.0231	-1.5006 ± 0.0218
a_2	0.7	0.7004 ± 0.0180	0.7001 ± 0.0181	0.7005 ± 0.0173
b_1	1.0	0.9930 ± 0.0998	0.9935 ± 0.1002	0.9966 ± 0.0960
b_2	0.5	0.5033 ± 0.1446	0.5046 ± 0.1458	0.4994 ± 0.1367
$\psi_{\tilde{y}}$	10.0	9.9956 ± 0.5836	9.8637 ± 0.5803	9.9960 ± 0.5832
$\psi_{\tilde{u}}$	1.0	0.9765 ± 0.1321	0.9725 ± 0.1318	0.9848 ± 0.1651
(a) Main parameters				
$[\mathbf{R}_{\mathbf{u}_0}]_{11}$	2.920	2.942 ± 0.258	2.935 ± 0.258	2.933 ± 0.277
$[\mathbf{R}_{\mathbf{u}_0}]_{21}$	2.160	2.158 ± 0.203	2.159 ± 0.203	2.159 ± 0.206
$[\mathbf{R}_{\mathbf{u}_0}]_{31}$	1.080	1.074 ± 0.181	1.073 ± 0.181	1.075 ± 0.188
$[\mathbf{R}_{\mathbf{u}_0}]_{41}$	0.540	0.531 ± 0.170	0.529 ± 0.170	0.533 ± 0.176
$[\mathbf{R}_{\mathbf{u}_0}]_{51}$	0.270	0.263 ± 0.169	0.261 ± 0.169	0.262 ± 0.177
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{11}$	-0.071	-0.016 ± 0.882	-0.012 ± 0.883	-0.029 ± 0.879
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{12}$	-0.143	-0.088 ± 0.884	-0.092 ± 0.883	-0.098 ± 0.870
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{13}$	-0.286	-0.236 ± 0.884	-0.244 ± 0.884	-0.244 ± 0.868
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{14}$	-0.572	-0.526 ± 0.889	-0.526 ± 0.889	-0.530 ± 0.879
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{15}$	-1.144	-1.098 ± 0.938	-1.087 ± 0.940	-1.101 ± 0.930
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{21}$	0.397	0.318 ± 1.266	0.317 ± 1.267	0.325 ± 1.265
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{22}$	0.794	0.721 ± 1.267	0.725 ± 1.267	0.723 ± 1.268
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{23}$	1.588	1.521 ± 1.275	1.524 ± 1.274	1.520 ± 1.274
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{24}$	3.177	3.111 ± 1.304	3.107 ± 1.303	3.110 ± 1.302
$[\mathbf{R}_{\boldsymbol{\nu}\mathbf{u}_0}]_{25}$	6.353	6.283 ± 1.361	6.281 ± 1.361	6.286 ± 1.365
$[\mathbf{R}_{\boldsymbol{\nu}}]_{11}$	74.822	74.879 ± 8.449	74.849 ± 8.465	74.903 ± 8.360
$[\mathbf{R}_{\boldsymbol{\nu}}]_{21}$	-98.290	-98.058 ± 10.480	-98.035 ± 10.492	-98.118 ± 10.435
$[\mathbf{R}_{\boldsymbol{\nu}}]_{22}$	160.285	159.894 ± 15.791	159.951 ± 15.790	159.897 ± 15.790
(b) Auxiliary parameters				

NC for any of the estimators were detected. However, some initial trials for the SEM-based estimators showed that in a few instances the algorithm failed to satisfy the convergence criteria. Note that the DFP algorithm involves obtaining a step-length parameter between iterations. The problem of NC was solved by changing the default settings, allowing for more accurate computations of the step-length parameter. It was additionally noticed that the numerical performance somewhat depends on choice of $\boldsymbol{\vartheta}^{(0)}$.

Example 5.2. In the second example, the implementation of the second formulation is considered. For this case, $n_\nu = n_a = 2$.

Table 3
Standard deviations of the difference between parameter estimates obtained from the two formulations.

Method	V_1	GLS	ULS
Param.			
$s_{\hat{a}_1^{(1)}-\hat{a}_1^{(2)}}$	$\pm 6.082 \cdot 10^{-4}$	$\pm 7.770 \cdot 10^{-4}$	$\pm 1.849 \cdot 10^{-3}$
$s_{\hat{a}_2^{(1)}-\hat{a}_2^{(2)}}$	$\pm 4.325 \cdot 10^{-4}$	$\pm 5.919 \cdot 10^{-4}$	$\pm 1.190 \cdot 10^{-3}$
$s_{\hat{b}_1^{(1)}-\hat{b}_1^{(2)}}$	$\pm 5.645 \cdot 10^{-3}$	$\pm 5.773 \cdot 10^{-3}$	$\pm 1.109 \cdot 10^{-2}$
$s_{\hat{b}_2^{(1)}-\hat{b}_2^{(2)}}$	$\pm 6.505 \cdot 10^{-3}$	$\pm 6.822 \cdot 10^{-3}$	$\pm 1.652 \cdot 10^{-2}$
$s_{\hat{\psi}_{\tilde{y}}^{(1)}-\hat{\psi}_{\tilde{y}}^{(2)}}$	$\pm 4.704 \cdot 10^{-3}$	$\pm 4.127 \cdot 10^{-3}$	$\pm 1.557 \cdot 10^{-2}$
$s_{\hat{\psi}_{\tilde{u}}^{(1)}-\hat{\psi}_{\tilde{u}}^{(2)}}$	$\pm 2.560 \cdot 10^{-3}$	$\pm 2.552 \cdot 10^{-3}$	$\pm 5.735 \cdot 10^{-3}$

Using (61), with $\kappa_j = 3 + j$ for $j = 1, 2$, the auxiliary processes become

$$\nu_1(t) = -a_2 y_0(t-6) + b_2 u_0(t-6), \tag{99}$$

$$\nu_2(t) = -a_1 y_0(t-6) - a_2 y_0(t-7) + b_1 u_0(t-6) + b_2 u_0(t-7). \tag{100}$$

To ensure convergence in all cases, we applied the same settings to the optimization algorithm as in the previous example.

The simulation results are summarized in Table 2. As before, the different estimation procedures tend to behave similarly. From Table 2(a), it is seen that the results for the main parameters are close to those obtained in the previous example. Clearly, the estimates of these parameters appear to be of the same quality across the two formulations. Considering the auxiliary parameters, the results in Table 2(b) show some tendency to underestimate the covariation among the latent processes. Again, no occurrences of the estimated measurement noise variances being negative were recorded.

In terms of numerical performance, the second formulation showed to be less sensitive to the choice of $\boldsymbol{\vartheta}^{(0)}$, and therefore had a smoother implementation.

Example 5.3. In the following example, we investigate to what extent the main parameter estimates obtained from the two formulations deviate. To do so, it is useful to introduce

$$\boldsymbol{\rho} = \begin{pmatrix} \boldsymbol{\theta}^T & \boldsymbol{\psi}^T \end{pmatrix}^T. \tag{101}$$

Let $s_{\hat{\rho}_h}$ denote the empirical standard deviation of the hth element of $\hat{\boldsymbol{\rho}}$ for $h = 1, \ldots, 6$. For small deviations, we expect that

$$s_{\hat{\rho}_h^{(1)}-\hat{\rho}_h^{(2)}} \ll s_{\hat{\rho}_h^{(1)}} \approx s_{\hat{\rho}_h^{(2)}}. \tag{102}$$

Note that our approach is rather conservative in the sense that (102) is expressed in terms of standard deviations instead of variances. The computations of $s_{\hat{\rho}_h^{(1)}-\hat{\rho}_h^{(2)}}$ are presented in Table 3, while $s_{\hat{\rho}_h^{(1)}}$ and $s_{\hat{\rho}_h^{(2)}}$ are found in Tables 1(a) and Table 2(a), respectively.

As can be observed from the tables, the general pattern is that $s_{\hat{\rho}_h^{(1)}-\hat{\rho}_h^{(2)}}$ is quite a bit smaller than $s_{\hat{\rho}_h^{(1)}}$ and $s_{\hat{\rho}_h^{(2)}}$. This observation supports the conjecture that the estimates obtained from the two formulations are in close agreement. It is additionally noticed that among the three estimators, the agreement is closer for V_1 and GLS than for ULS.

Example 5.4. In the final example, the empirical variance of the transfer function for the various SEM-based estimators is considered. Let the empirical variance be given by

$$s^2_{\hat{G}}(\omega) = \frac{1}{n_{Rep}} \sum_{i=1}^{n_{Rep}} \left(|\hat{G}_i| - |G| \right)^2, \quad 0 \le \omega \le \pi, \tag{103}$$

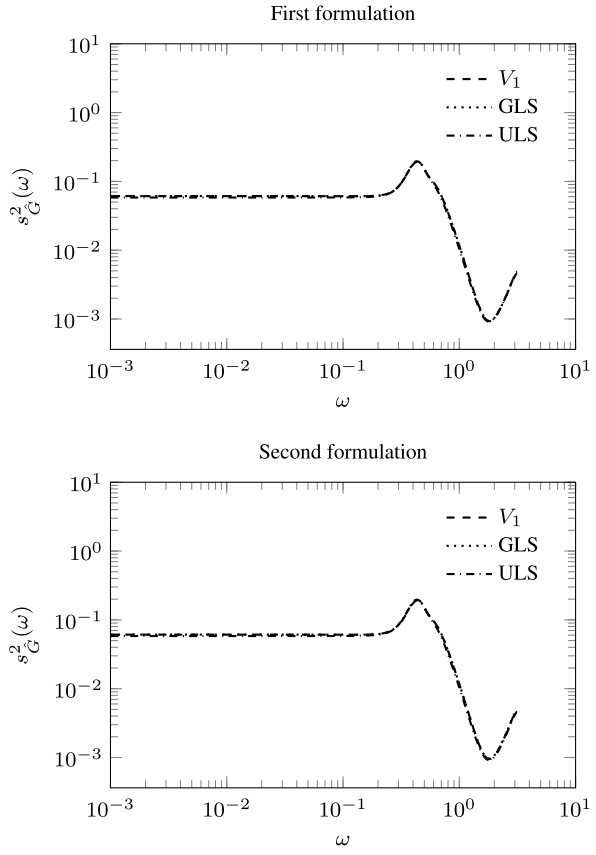

Fig. 2. Empirical variance of the transfer function.

where $G = G(e^{i\omega})$ follows from (6) with $n_a = n_b = 2$. Evaluating the estimation performance using (103) is useful since $s^2_{\hat{G}}(\omega)$ is influenced not only by the variance of the individual estimates in $\hat{\boldsymbol{\theta}}$ but also their covariances. The results of the computations are summarized in Fig. 2.

As clearly seen from the figure, the variance is nearly identical across the various estimation procedures. Moreover, no noticeable differences between the two formulations are evident. Thus, the results are shown to be well in line with those reported in Example 5.3.

6. Concluding remarks

EIV identification using SEM has been examined. More specifically, two formulations of the EIV SISO system leading to the same covariance structure were presented. The proposed formulations allow for quick implementation using standard statistical software. To make the SEM framework more specific for the EIV problem, several extensions were proposed. First, for the sake of parsimony, it was shown how to eliminate the redundancy originating from the block Toeplitz form of the considered covariance matrix. Second, when the system admits a CFA representation, the estimation problem can be formulated as a separate least squares problem.

Statistical performance was evaluated using simulated data. The main conclusion from the simulation examples is that SEM-based estimation of dynamic EIV systems works well. Performance across the two formulations showed to be nearly identical in terms of bias and accuracy. As compared to existing procedures, here represented by the CM approach, SEM-based estimators provide estimates of similar quality. Algorithmic aspects were briefly considered. Although the optimization algorithm in LISREL performed well in the simulations, one cannot exclude the possibility that other algorithms may be better suited for the EIV problem.

It is clear that more work is needed to fully understand the potential of applying SEM to the EIV problem. Future research should include a rigorous investigation of the accuracy of the SEM-based estimators. It may also be of interest to extend the present work to include formulations of Multiple-Input Multiple-Output (MIMO) systems.

Acknowledgments

The authors would like to thank Professor Johan Lyhagen, Associate professor Gillian Warner-Søderholm and the reviewers for their valuable comments and suggestions.

Appendix A. Deriving K

We now provide a treatment of how to obtain the matrix $\mathbf{K}_n$ for various forms of $\mathbf{R}$, where $\mathbf{R}$ denotes an n-dimensional symmetric matrix. Let $\mathbf{L}_n$ denote a *selection* matrix of dimension $n^2 \times n^*$ consisting of ones and zeros such that

$$\operatorname{vec}(\mathbf{R}) = \mathbf{L}_n \operatorname{vecnr}(\mathbf{R}), \tag{A.1}$$

where

$$\operatorname{vec}(\mathbf{R}) = \begin{pmatrix} r_{11} & \cdots & r_{n1} & r_{12} & \cdots & r_{n2}, & \cdots & , r_{1n} & \cdots & r_{nn} \end{pmatrix}^T, \tag{A.2}$$

and n^* is the number of non-repeated elements in $\mathbf{R}$. The matrix $\mathbf{K}_n$ is then chosen as

$$\mathbf{K}_n = \mathbf{L}_n \left(\mathbf{L}_n^T \mathbf{L}_n\right)^{-1}. \tag{A.3}$$

In this form, $\mathbf{K}_n$ possesses the property described in (25). The matrices in (A.3) are typically derived using *double subscript* notation; see Browne (1974) and Fuller (2006). It is now demonstrated how to apply this type of notation to obtain $\mathbf{L}_n$ for various forms of $\mathbf{R}$.

Case 1: First, consider the case when $\mathbf{R}$ is symmetric and no other restrictions are placed on its elements. A vector of the non-repeated elements is given by

$$\operatorname{vecnr}(\mathbf{R}) = \begin{pmatrix} r_{11} & \cdots & r_{n1} & r_{22} & \cdots & r_{n2}, & \cdots & , r_{nn} \end{pmatrix}^T, \tag{A.4}$$

where the number of elements follows from (18). Thus, the dimension of $\mathbf{L}_n$ is $n^2 \times 2^{-1} n(n+1)$.

Applying double subscript notation involves indexing the elements of $\mathbf{L}_n$ using the same subscripts as for the elements of $\operatorname{vec}(\mathbf{R})$ and $\operatorname{vecnr}(\mathbf{R})$. The indexing scheme is illustrated by

$$\mathbf{L}_n = \begin{pmatrix} l_{11,11} & \cdots & l_{11,n1} & l_{11,22} & \cdots & l_{11,n2} & \cdots & \cdots & l_{11,nn} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ l_{n1,11} & \cdots & l_{n1,n1} & l_{n1,22} & \cdots & l_{n1,n2} & \cdots & \cdots & l_{n1,nn} \\ l_{12,11} & \cdots & l_{12,n1} & l_{12,22} & \cdots & l_{12,n2} & \cdots & \cdots & l_{12,nn} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ l_{n2,11} & \cdots & l_{n2,n1} & l_{n2,22} & \cdots & l_{n2,n2} & \cdots & \cdots & l_{n2,nn} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ l_{1n,11} & \cdots & l_{1n,n1} & l_{1n,22} & \cdots & l_{1n,n2} & \cdots & \cdots & l_{1n,nn} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ l_{nn,11} & \cdots & l_{nn,n1} & l_{nn,22} & \cdots & l_{nn,n2} & \cdots & \cdots & l_{nn,nn} \end{pmatrix}. \tag{A.5}$$

In order to describe the individual elements of (A.5), let $i, j = 1, \ldots, n$ and $u \geq v = 1, \ldots, n$. The elements are then determined by

$$l_{ij,uv} = \begin{cases} 1 & \text{if } (i,j) = (u,v) \\ 1 & \text{if } (i,j) = (v,u) \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

Example 3.1 illustrates the case when $n = 3$, in which $\mathbf{L}_3$ takes the form of (28). Using (A.3), we obtain (27) with $\alpha_1 = \alpha_2 = \alpha_3 = 0.5$.

Case 2: In the second case, $\mathbf{R}$ is a matrix that is both symmetric and Toeplitz. The vector of non-repeated elements simplifies to the first column of $\mathbf{R}$

$$\text{vecnr}(\mathbf{R}) = \begin{pmatrix} r_{11} & \cdots & r_{n1} \end{pmatrix}^T, \tag{A.7}$$

and the dimension of $\mathbf{L}_n$ becomes $n^2 \times n$. Before describing the elements of this matrix, it is useful to introduce the following general notation. Let $H_u = \{0, \ldots, \imath_u\}$ denote an integer set. Now, define

$$\{(u+h, v+h)\}_{h \in H_u} = \{(u,v), \ldots, (u+\imath_u, v+\imath_u)\}. \tag{A.8}$$

If $i, j, u = 1, \ldots, n$, $v = 1$ and $\imath_u = n - u$, then

$$l_{ij,uv} = \begin{cases} 1 & \text{if } (i,j) \in \{(u+h, v+h)\}_{h \in H_u} \\ 1 & \text{if } (i,j) \in \{(v+h, u+h)\}_{h \in H_u} \\ 0 & \text{otherwise.} \end{cases} \tag{A.9}$$

The set $\{(u+h, v+h)\}_{h \in H_u}$ additionally depends on u through H_u. Thus, columns further to the right in $\mathbf{L}_n$ are characterized by fewer non-zero elements. Note that the first column is a special case in which the first two conditions in (A.9) coincide.

Using (A.6) and (A.9), either separately or in combination, one can handle a variety of matrices. This is further demonstrated in the remaining two cases.

Case 3: In the third and more involved case, $\mathbf{R}$ is a matrix with a block Toeplitz structure

$$\mathbf{R} = \begin{pmatrix} \underset{(n_1 \times n_1)}{\mathbf{R}_{11}} & \underset{(n_1 \times n_2)}{\mathbf{R}_{12}} \\ \underset{(n_2 \times n_1)}{\mathbf{R}_{21}} & \underset{(n_2 \times n_2)}{\mathbf{R}_{22}} \end{pmatrix}, \tag{A.10}$$

where $n = n_1 + n_2$. Let $[\mathbf{R}_{kl}]_{fg}$ denote the fgth element of the klth block of $\mathbf{R}$. The vector of non-repeated elements is

$$\begin{aligned} \text{vecnr}(\mathbf{R}) = \big(& [\mathbf{R}_{11}]_{11} \cdots [\mathbf{R}_{11}]_{n_1 1} \quad [\mathbf{R}_{21}]_{11} \cdots [\mathbf{R}_{21}]_{n_2 1} \\ & [\mathbf{R}_{12}]_{21} \cdots [\mathbf{R}_{12}]_{n_1 1} \quad [\mathbf{R}_{22}]_{11} \cdots [\mathbf{R}_{22}]_{n_2 1} \big)^T. \end{aligned} \tag{A.11}$$

It will be convenient to let $\mathbf{L}_n$ be a matrix composed of four partitions, where each partition corresponds to a block in $\mathbf{R}$. The selection matrix then takes the general form

$$\mathbf{L}_n = \begin{pmatrix} \underset{(n^2 \times n_1)}{\mathbf{L}_{11}} & \underset{(n^2 \times n_2)}{\mathbf{L}_{12}} & \underset{(n^2 \times (n_1 - 1))}{\mathbf{L}_{13}} & \underset{(n^2 \times n_2)}{\mathbf{L}_{14}} \end{pmatrix}, \tag{A.12}$$

with overall dimension $n^2 \times (2n-1)$. As before, let $i, j = 1, \ldots, n$. By appropriately specifying the range of u and v, one can obtain the individual elements of $\mathbf{L}_n$ by its partitions. This is accomplished by

$\mathbf{L}_{11}$: (A.9), using $u = 1, \ldots, n_1$, $v = 1$ and $\imath_u = n_1 - u$.
$\mathbf{L}_{12}$: (A.9), using $u = n_1 + 1, \ldots, n$, $v = 1$ and $\imath_u = \min(n_1 - 1, n - u)$.
$\mathbf{L}_{13}$: (A.9), using $u = 2, \ldots, n_1$, $v = n_1 + 1$ and $\imath_u = \min(n_2 - 1, n_1 - u)$.
$\mathbf{L}_{14}$: (A.9), using $u = n_1 + 1, \ldots, n$, $v = n_1 + 1$ and $\imath_u = n - u$.

Case 4: In the final case, $\mathbf{R}$ is written in the form of (A.10), but only $\mathbf{R}_{11}$ has a Toeplitz structure. An example of such a matrix is given in (73). The vector of non-repeated elements is in this case

$$\begin{aligned} \text{vecnr}(\mathbf{R}) = \big(& [\mathbf{R}_{11}]_{11} \cdots [\mathbf{R}_{11}]_{n_1 1} \quad [\mathbf{R}_{21}]_{11} \cdots [\mathbf{R}_{21}]_{n_2 1}, \ldots, \\ & [\mathbf{R}_{21}]_{1 n_1} \cdots [\mathbf{R}_{21}]_{n_2 n_1} \quad [\mathbf{R}_{22}]_{11} \cdots [\mathbf{R}_{22}]_{n_2 1} \\ & [\mathbf{R}_{22}]_{22} \cdots [\mathbf{R}_{22}]_{n_2 2}, \cdots, [\mathbf{R}_{22}]_{n_2 n_2} \big)^T, \end{aligned} \tag{A.13}$$

and the selection matrix is written as

$$\mathbf{L}_n = \begin{pmatrix} \underset{(n^2 \times n_1)}{\mathbf{L}_{11}} & \underset{(n^2 \times (n_2 n_1 + n_2^*))}{\mathbf{L}_{12}} \end{pmatrix}. \tag{A.14}$$

The overall dimension of $\mathbf{L}_n$ is $n^2 \times (n_1 + n_2 n_1 + n_2^*)$, where $n_2^* = 2^{-1} n_2 (n_2 + 1)$. Let $i, j = 1, \ldots, n$. The individual elements of $\mathbf{L}_n$ are then obtained by

$\mathbf{L}_{11}$: (A.9), using $u = 1, \ldots, n_1$, $v = 1$ and $\imath_u = n_1 - u$.
$\mathbf{L}_{12}$: (A.6), using $u = n_1 + 1, \ldots, n$ and $u \geq v = 1, \ldots, n$.

Appendix B. WLS as a separable least squares problem

The functional form of the WLS criteria allows it to be formulated as separate least squares problem. Applying the WLS estimator involves

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} V_3(\boldsymbol{\vartheta}), \tag{B.1}$$

where

$$V_3(\boldsymbol{\vartheta}) = \tilde{\mathbf{r}}^T(\boldsymbol{\vartheta}) \mathbf{W} \tilde{\mathbf{r}}(\boldsymbol{\vartheta}) \tag{B.2}$$

is the criterion function presented in (23). From (24), one can write

$$\begin{aligned} \tilde{\mathbf{r}}(\boldsymbol{\vartheta}) &= \hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\vartheta}) \\ &= \hat{\mathbf{r}} - \mathbf{F}(\boldsymbol{\theta}) \mathbf{r}_\zeta, \end{aligned} \tag{B.3}$$

where

$$\mathbf{r}_\zeta = \begin{pmatrix} \mathbf{r}_\xi^T & \boldsymbol{\psi}^T \end{pmatrix}^T. \tag{B.4}$$

In what follows, it is assumed that $\mathbf{F}(\boldsymbol{\theta})$ is a matrix of full column rank. Using (B.3), the WLS criteria can be written in the form

$$V_3(\boldsymbol{\theta}, \mathbf{r}_\zeta) = \| \hat{\mathbf{r}} - \mathbf{F}(\boldsymbol{\theta}) \mathbf{r}_\zeta \|_{\mathbf{W}}^2. \tag{B.5}$$

For a given $\boldsymbol{\theta}$, and using that $\mathbf{r}_\zeta$ enters linearly into (B.5), the minimization problem with respect to $\mathbf{r}_\zeta$ is solved by linear least squares

$$\hat{\mathbf{r}}_\zeta = \big(\mathbf{F}^T(\boldsymbol{\theta}) \mathbf{W} \mathbf{F}(\boldsymbol{\theta}) \big)^{-1} \mathbf{F}^T(\boldsymbol{\theta}) \mathbf{W} \hat{\mathbf{r}}, \tag{B.6}$$

which in turn leads to the modified criterion function

$$\widetilde{V}_3(\boldsymbol{\theta}) = \hat{\mathbf{r}}^T \mathbf{W} \hat{\mathbf{r}} - \hat{\mathbf{r}}^T \mathbf{W} \mathbf{F}(\boldsymbol{\theta}) \big(\mathbf{F}^T(\boldsymbol{\theta}) \mathbf{W} \mathbf{F}(\boldsymbol{\theta}) \big)^{-1} \mathbf{F}^T(\boldsymbol{\theta}) \mathbf{W} \hat{\mathbf{r}}. \tag{B.7}$$

The minimization of (B.7) describes a variable projection problem that can be solved using standard optimization routines.

To make (B.6) and (B.7) applicable, one must derive an expression for $\mathbf{F}(\boldsymbol{\theta})$. Recall that both formulations described in Section 4 have a CFA representation. The simple form of the system implied covariance matrix, as given in (54), simplifies the problem of determining $\mathbf{F}(\boldsymbol{\theta})$. It is useful to introduce the following lemma:

Lemma B.1. *Let* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *be matrices such that the product* $\mathbf{ABC}$ *is well defined. Then it holds that*

$$\text{vec}(\mathbf{ABC}) = \big(\mathbf{C}^T \otimes \mathbf{A} \big) \text{vec}(\mathbf{B}). \tag{B.8}$$

Proof. See Bernstein (2009). □

Applying the lemma to the covariance matrix in (54) allows us to write

$$\operatorname{vec}\big(\mathbf{R}(\vartheta)\big) = \big(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\big)\operatorname{vec}(\mathbf{\Phi}) + \operatorname{vec}(\mathbf{\Psi}_\epsilon). \tag{B.9}$$

Use that

$$\operatorname{vec}\big(\mathbf{R}(\vartheta)\big) = \mathbf{K}_n^{-T}\mathbf{r}(\vartheta), \qquad \operatorname{vec}(\mathbf{\Phi}) = \mathbf{K}_{n_\xi}^{-T}\mathbf{r}_\xi, \tag{B.10}$$

such that (B.9) can be written as

$$\mathbf{r}(\vartheta) = \mathbf{K}_n^T\left((\mathbf{\Lambda} \otimes \mathbf{\Lambda})\mathbf{K}_{n_\xi}^{-T}\mathbf{r}_\xi + \operatorname{vec}(\mathbf{\Psi}_\epsilon)\right). \tag{B.11}$$

Moreover, note that

$$\operatorname{vec}(\mathbf{\Psi}_\epsilon) = (\boldsymbol{\iota}_y\ \boldsymbol{\iota}_u)\boldsymbol{\psi}, \tag{B.12}$$

where $\boldsymbol{\iota}_y$ and $\boldsymbol{\iota}_u$ are vectors obtained from

$$\boldsymbol{\iota}_y = \operatorname{vec}\begin{pmatrix} \mathbf{I}_{n_y} & \mathbf{0}_{(n_y\times n_u)} \\ \mathbf{0}_{(n_u\times n_y)} & \mathbf{0}_{(n_u\times n_u)} \end{pmatrix}, \tag{B.13}$$

$$\boldsymbol{\iota}_u = \operatorname{vec}\begin{pmatrix} \mathbf{0}_{(n_y\times n_y)} & \mathbf{0}_{(n_y\times n_u)} \\ \mathbf{0}_{(n_u\times n_y)} & \mathbf{I}_{n_u} \end{pmatrix}. \tag{B.14}$$

It now follows that

$$\begin{aligned}\mathbf{r}(\vartheta) &= \mathbf{K}_n^T\left((\mathbf{\Lambda} \otimes \mathbf{\Lambda})\mathbf{K}_{n_\xi}^{-T}\ \ \boldsymbol{\iota}_y\ \ \boldsymbol{\iota}_u\right)\begin{pmatrix}\mathbf{r}_\xi \\ \boldsymbol{\psi}\end{pmatrix} \\ &\triangleq \mathbf{F}(\boldsymbol{\theta})\,\mathbf{r}_\zeta.\end{aligned} \tag{B.15}$$

The exact form and dimension of $\mathbf{F}(\boldsymbol{\theta})$ depends on $\mathbf{K}_n$ and $\mathbf{K}_{n_\xi}$. The relevant matrices are found by the procedures outlined in Appendix A. Considering the first formulation, we obtain

$\mathbf{K}_{n_\xi}^{(1)}$: by **Case 3** using $n_1, n_2 = n_y, n_u$.
$\mathbf{K}_{n_\xi}^{(1)}$: by **Case 2** using $n_\xi = k + 1$.

The dimension of $\mathbf{F}(\boldsymbol{\theta})$ is then $(2n-1)\times(k+3)$. For the second formulation, we obtain

$\mathbf{K}_{n_\xi}^{(2)}$: by **Case 1** using $n = n_y + n_u$.
$\mathbf{K}_{n_\xi}^{(2)}$: by **Case 4** using $n_1, n_2 = n_u, n_\nu$.

The dimension of $\mathbf{F}(\boldsymbol{\theta})$ is in this case $n^*\times(n_{r_\xi}+2)$, where n^* and n_{r_ξ} are determined by (18) and (75), respectively.

It follows from the preceding analysis that if estimates of $\mathbf{r}(\vartheta_0)$ and $\boldsymbol{\theta}_0$ are available, then estimates of $\mathbf{r}_{\xi,0}$ and $\boldsymbol{\psi}_0$ can be obtained from the use of (B.6).

Appendix C. Imposing additional constraints on the elements of $\mathbf{\Phi}$

It is possible to achieve a more parsimonious implementation of the second formulation by further constraining the elements of $\mathbf{\Phi}$.

As defined in (32), let $r_{u_0}(\tau)$ denote the covariance function of the undisturbed input signal $u_0(t)$. Further, define

$$r_{y_0u_0}(\tau) = \mathsf{E}\left\{y_0(t+\tau)u_0(t)\right\}, \tag{C.1}$$

$$r_{v_ju_0}(\tau) = \mathsf{E}\left\{v_j(t+\tau)u_0(t)\right\}, \tag{C.2}$$

where $v_j(t)$ is the auxiliary process obtained from (61). For $j = 2,\dots,n_\nu$, one can modify (61) so that

$$\begin{aligned}v_j(t) = &-\sum_{i=1}^{n_a}\mathbb{1}_{\{\kappa_j+i=n_y\}}a_iy_0(t-n_y) \\ &+\sum_{i=1}^{n_b}\mathbb{1}_{\{\kappa_j+i=n_u+1\}}b_iu_0(t-n_u-1)+v_{j-1}(t-1).\end{aligned} \tag{C.3}$$

Note that (C.3) becomes applicable for $j = 1$ by canceling the last term.

By the use of (C.3), and letting $h = 2,\dots,n_u$, one can express the elements of $\mathbf{R}_{\nu\mathbf{u}_0}$ as

$$\begin{aligned}\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{jh} =& r_{v_ju_0}(h) \\ =& -\sum_{i=1}^{n_a}\mathbb{1}_{\{\kappa_j+i=n_y\}}a_ir_{y_0u_0}(h-n_y) \\ &+\sum_{i=1}^{n_b}\mathbb{1}_{\{\kappa_j+i=n_u+1\}}b_ir_{u_0}(h-n_u-1) \\ &+r_{v_{j-1}u_0}(h-1),\end{aligned} \tag{C.4}$$

where $\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{jh}$ denotes the jhth element of $\mathbf{R}_{\nu\mathbf{u}_0}$. As before, canceling the last term makes (C.4) applicable for $j = 1$. To make use of (C.4), one must obtain an expression for $r_{y_0u_0}(h-n_y)$. To do so, consider

$$\begin{aligned}y_0(t-n_y+1) =& \sum_{i=1}^{n_b}\mathbb{1}_{\{\kappa_{n_\nu}+i<n_u+1\}}b_iu_0(t-\kappa_{n_\nu}-i) \\ &+v_{n_\nu}(t).\end{aligned} \tag{C.5}$$

Multiplying both sides of (C.5) by $u_0(t-h+1)$ and taking expectations, while using that $\kappa_{n_\nu} = n_y - 1$, gives

$$\begin{aligned}r_{y_0u_0}(h-n_y) =& \sum_{i=1}^{n_b}\mathbb{1}_{\{\kappa_{n_\nu}+i<n_u+1\}}b_ir_{u_0}(h-n_y-i) \\ &+r_{v_{n_\nu}u_0}(h-1).\end{aligned} \tag{C.6}$$

In (C.6), when the condition in the indicator function is satisfied, $r_{u_0}(h-n_y-i)$ for $h = 2,\dots,n_u$ are guaranteed to be contained in $\mathbf{R}_{\mathbf{u}_0}$. It follows that $r_{y_0u_0}(h-n_y)$ can be obtained from the elements of $\boldsymbol{\theta}$, $\mathbf{R}_{\mathbf{u}_0}$ and $\mathbf{R}_{\nu\mathbf{u}_0}$. The expressions in (C.4) and (C.6) implies that the complete set of constraints can be obtained by recursive substitution.

After imposing the constraints, the dimension of $\mathbf{r}_\xi$ becomes

$$\begin{aligned}n_{r_\xi} &= n_u + n_\nu + \frac{1}{2}n_\nu(n_\nu+1) \\ &= n_u + \frac{1}{2}n_\nu(n_\nu+3).\end{aligned} \tag{C.7}$$

As an illustration, consider the system used in Example 5.2. Using (C.3), the auxiliary processes are given by

$$v_1(t) = -a_2y_0(t-6) + b_2u_0(t-6) \tag{C.8}$$

$$v_2(t) = -a_1y_0(t-6) + b_1u_0(t-6) + v_1(t-1). \tag{C.9}$$

The choice of n_y and n_u are such that (C.5) simplifies to

$$y_0(t-n_y+1) = y_0(t-5) = v_2(t), \tag{C.10}$$

and from the use of (C.4) and (C.6)

$$\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{1,h} = -a_2\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{2,h-1} + b_2\left[\mathbf{R}_{\mathbf{u}_0}\right]_{5,h-1} \tag{C.11}$$

$$\begin{aligned}\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{2,h} =& -a_1\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{2,h-1} + b_1\left[\mathbf{R}_{\mathbf{u}_0}\right]_{5,h-1} \\ &+\left[\mathbf{R}_{\nu\mathbf{u}_0}\right]_{1,h-1}.\end{aligned} \tag{C.12}$$

The desired covariance elements for $h = 2,\dots,5$ are now found by applying recursive substitution to (C.11) and (C.12).

References

Bartholomew, David J., Knott, Martin, & Moustaki, Irini (2011). *Latent variable models and factor analysis: a unified approach* (3rd ed.). London: John Wiley & Sons, Ltd..

Beghelli, Sergio, Castaldi, Paolo, Guidorzi, Roberto P., & Soverini, Umberto (1993). A robust criterion for model selection in identification from noisy data. In *Proc. 9th international conference on systems engineering* (pp. 480–484).

Beghelli, Sergio, Guidorzi, Roberto P., & Soverini, Umberto (1990). The frisch scheme in dynamic system identification. *Automatica*, *26*(1), 171–176.
Bernstein, Dennis S. (2009). *Matrix mathematics* (2nd ed.). Princeton, NJ: Princeton University Press.
Bollen, Kenneth A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons, Inc.
Browne, Michael W. (1974). Generalized least-square estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1–24.
Browne, Michael W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.
Browne, Michael W., & Nesselroade, John R. (2005). Representing psychological processes with dynamic factor models: some promising uses and extensions of autogressive moving average time series models. In Albert Maydeu-Olivares, & John J McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. Mcdonald* (pp. 415–452). Mahwah, NJ: Lawrence Erlbaum Associates.
Browne, Michael W., & Zhang, Guangjian (2007). Developments in the factor analysis of individual time series. In Robert Cudeck, & Robert C MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 265–291). Mahwah, NJ: Lawrence Erlbaum Associates.
Cheng, Chi-Lun, & Van Ness, John W. (1999). *Statistical regression with measurement error, Vol. 6*. London, New York: Arnold, Oxford University Press.
Diversi, Roberto P., & Guidorzi, Roberto (2012). A covariance-matching criterion in the frisch scheme identification of mimo eiv models. In *Proc. 16th IFAC symposium on system identification, 2012* (pp. 1647–1652).
Diversi, Roberto, Guidorzi, Roberto P., & Soverini, Umberto (2003). A new criterion in eiv identification and filtering applications. In *Proc. 13th IFAC symposium on system identification* (pp. 1993–1998).
Diversi, Roberto, Guidorzi, Roberto P., & Soverini, Umberto (2004). Frisch scheme-based algorithms for eiv identification. In *Proc. 12th IEEE mediterranean conference on control and automation*.
Diversi, Roberto, Guidorzi, Roberto P., & Soverini, Umberto (2006). Yule-walker equations in the frisch scheme solution of errors-in-variables identification problems. In *Proc. 17th international symposium on mathematical theory of networks and systems, MTNS* (pp. 391–395).
du Toit, Stephen H. C., & Browne, Michael W. (2007). Structural equation modeling of multivariate time series. *Multivariate Behavioral Research*, *42*(1), 67–101.
Ekman, Mats (2005). Identification of linear system with errors in variables using separable nonlinear least-squares. In *Proc. 16th IFAC world congress* (pp. 136–136).
Ekman, Mats, Hong, Mei, & Söderström, Torsten (2006). Identification of linear system with errors in variables using separable nonlinear least-squares. In *Proc. 14th IFAC symposium on system identification* (pp. 178–183.
Fuller, Wayne A. (2006). *Measurement error models*. Hoboken, NJ: John Wiley & Sons, Inc.
Geweke, John F., & Singleton, Kenneth J. (1981). Maximum likelihood confirmatory factor analysis of economic time series. *International Economic Review*, *22*(1), 37–54.
Golub, Gene H., & Pereyra, Victor (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, *10*(2), 413–432.
Golub, Gene, & Pereyra, Victor (2003). Separable nonlinear least squares: The variable projection method and its applications. *Inverse Probl.*, *19*(2), R1–R26.
Guidorzi, Roberto, Diversi, Roberto P., & Soverini, Umberto (2008). The frisch scheme in algebraic and dynamic identification problems. *Kybernetika*, *44*(5), 585–616.
Hamaker, Ellen L., Dolan, Conor V., & Molenaar, Peter C. M. (2002). On the nature of sem estimates of arma parameters. *Structural Equation Modeling*, *9*(3), 347–368.
Jöreskog, Karl G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482.
Jöreskog, Karl G. (1970). A general method for analysis of covariance structure. *Biometrika*, *57*(2), 239–251.
Jöreskog, Karl G., & Goldberger, Arthur S. (1972). Factor analysis by generalized least squares. *Psychometrika*, *37*(3), 243–260.
Jöreskog, Karl G., & Sörbom, Dag (1996). *LISREL 8: User's reference guide*. Chicago, Il: Scientific Software International.
Kreiberg, David, Söderström, Torsten, & Yang-Wallentin, Fan (2013). Errors-in-variables identification using covariance matching and structural equation modeling. In *Proc. decision and control, CDC, 2013 IEEE 52nd annual conference on* (pp. 5852–5857).
Lee, Sik-Yum (2007). *Structural equation modeling: A Bayesian approach*. London: John Wiley & Sons, Ltd..
Ljung, Lennart (1999). *System identification—theory for the user* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR, Inc.
Mahata, Kaushik (2007). An improved bias-compensation approach for errors-in-variables model identification. *Automatica*, *43*(8), 1339–1354.
Mathworks. Matlab Optimization Toolbox User's Guide R2014a.
Molenaar, Peter C. M. (1985). A dynamic factor analysis model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202.
Molenaar, Peter C. M., & Nesselroade, John R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-Toeplitz matrices representing single subject multivariate time series. *Multivariate Behavioral Research*, *33*(3), 313–342.
Mossberg, Magnus, & Söderström, Torsten (2011). Covariance matching for continuous-time errors-in-variables problems. *IEEE Transactions on Automatic Control*, *56*(6), 1478–1483.
Mossberg, Magnus, & Söderström, Torsten (2012). On covariance matching for multiple input multiple output errors-in-variables systems. In *Proc. 16th IFAC symposium on system identification, 2012* (pp. 1371–1376).
Söderström, Torsten (1981). Identification of stochastic linear systems in presence of input noise. *Automatica*, *17*(5), 713–725.
Söderström, Torsten (2007). Error-in-variables methods in system identification. *Automatica*, *43*(6), 939–958.
Söderström, Torsten (2008). Extending the frisch scheme for errors-in-variables identification to correlated output noise. *International Journal of Adaptive Control and Signal Processing*, *22*(1), 55–73.
Söderström, Torsten (2012). System identification for the errors-in-variables problem. *Trans. Inst. Meas. Control*, *34*(7), 780–792.
Söderström, Torsten, Kreiberg, David, & Mossberg, Magnus (2014). Extended accuracy analysis of a covariance matching approach for identifying errors-in-variables systems. *Automatica*, *50*(10), 2597–2605.
Söderström, Torsten, & Mossberg, Magnus (2011). Accuracy analysis of a covariance matching approach for identifying errors-in-variables systems. *Automatica*, *47*(2), 272–282.
Söderström, Torsten, Mossberg, Magnus, & Hong, Mei (2009). A covariance matching approach for identifying errors-in-variables systems. *Automatica*, *45*(9), 2018–2031.
Söderström, Torsten, & Stoica, Petre (1989). *System identification* (2nd ed.). Hemel Hempstead, UK: Prentice-Hall, Inc.
van Buuren, Stef (1997). Fitting arma time series by structural equation models. *Psychometrika*, *62*(2), 215–236.
Zheng, Wei Xing (1998). Transfer function estimation from noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, *12*(4), 365–380.
Zheng, Wei Xing (2002). A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, *47*(7), 1142–1147.

David Kreiberg received his M.Sc. degree in financial economics from the Norwegian Business School, Oslo Norway. Currently, David is holding a lectureship in statistics at the Norwegian Business School, while being a Ph.D. candidate in statistics at Uppsala University, Sweden. His main research interests include multivariate statistical analysis and system identification.

Torsten Söderström received the M.Sc. degree (“civilingenjör”) in engineering physics in 1969 and the Ph.D. degree in automatic control in 1973, both from Lund Institute of Technology, Lund, Sweden. He is a Life Fellow of IEEE, and an IFAC Fellow.

During 1967–1974 he held various teaching positions at the Lund Institute of Technology. Since 1974, he has been with the Department of Systems and Control, Uppsala University, Uppsala, Sweden. He was a professor of automatic control 1982–2011, and is now a professor emeritus.

Dr. Söderström is the author or coauthor of many technical papers. His main research interests are in the fields of system identification, signal processing, and control. He is the (co)author of four books: “Theory and Practice of Recursive Identification”, MIT Press, 1983 (with L. Ljung), “The Instrumental Variable Methods for System Identification”, Springer-Verlag, 1983 (with P. Stoica), “System Identification”, Prentice–Hall, 1989 (with P. Stoica) and “Discrete-Time Stochastic Systems”, Prentice–Hall, 1994; second edition, Springer-Verlag, 2002. In 1981 he was, with coauthors, given an Automatica Paper Prize Award.

Within IFAC he has served in several capacities including vice-chairman of the TC on Modeling, Identification and Signal Processing, (1993–1999), IPC chairman of the IFAC SYSID'94 Symposium, Council member (1996–2002), Executive Board member (1999–2002) and Awards Committee Chair (1999–2002). He was an associate editor (1984–1991), guest associate editor and editor for four special issues with *Automatica* and has been the editor for the area of System Parameter Estimation since 1992.

Fan Yang-Wallentin is a professor in statistics at the Department of Statistics, Uppsala University. She earned her Ph.D. in statistics at Uppsala University, Sweden, 1997. She is a recipient of Arnberg Prize from the Swedish Royal Academy of Sciences. Dr. Wallentin's program of research focuses on the theory and applications of Latent Variable Modeling and other types of Multivariate Statistical Analysis, particularly their applications in the social and behavioral sciences. She published research articles in several leading statistics and psychometrics journals on these and related topics. She has also been teaching courses in Structural Equation models in Sweden, Europe, USA and China. She has broad experiences in statistical consultation for researchers in social and behavioral sciences.