Automatica 156 (2023) 111187

Contents lists available at ScienceDirect

Automatica

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

Brief paper

A confirmatory factor analysis approach for addressing the errors-in-variables problem with colored output noise☆

David Kreiberg

Department of Economics, BI Norwegian Business School, Oslo, Norway

ARTICLE INFO

Article history:
Received 25 April 2022
Received in revised form 15 March 2023
Accepted 8 June 2023
Available online 3 August 2023

Keywords:
System identification
Errors-in-variables
Confirmatory factor analysis
Colored output noise
Minimum distance estimation

ABSTRACT

Over the years, errors-in-variables (EIV) system identification has attracted considerable research interest. Among the many proposed approaches for identifying EIV models is confirmatory factor analysis (CFA), here referred to as EIV-CFA. This study extends previous research by presenting a EIV-CFA modeling framework that allows for colored output noise. Considerable attention is paid to the theoretical aspects of the minimum distance (MD) estimator. The finite sample performance of the MD estimator is briefly evaluated using simulation. The results suggest that model parameters are well estimated.

1. Introduction

The system identification toolkit consists of a wide range of parametric procedures specifically intended for fitting models to time-dependent input and output data. This study deals with identifying errors-in-variables (EIV) models. What distinguishes EIV models from other types of linear parametric models is that the input process is corrupted by additive measurement noise. The identification of such models has been subject to extensive research, and numerous estimation procedures have been suggested over the years. Among the procedures are bias eliminating least squares (BELS) (Zheng, 1998, 2002), the Frisch estimator (Beghelli, Guidorzi, & Soverini, 1990; Diversi, Guidorzi, & Soverini, 2004, 2006), bias compensation (Ekman, 2005), and instrumental variables (IV) methods (Söderström, 2007, 2011). More recent developments have focused on techniques based on covariance structure analysis (CSA), namely, covariance matching (CM) (Mossberg & Söderström, 2012; Söderström, Kreiberg, & Mossberg, 2014; Söderström & Mossberg, 2011; Söderström, Mossberg, & Hong, 2009), and structural equation modeling (SEM) (Kreiberg, Söderström, & Yang-Wallentin, 2016; Söderström, 2018), with the latter including confirmatory factor analysis (CFA) as a special case.

1.1. Contribution

The motivation for this study is to further extend the EIV-CFA modeling approach, and to revise previous work by increasing the level of detail. The contributions to the existing literature are threefold. First, we introduce a framework that allows for colored output noise. Second, we provide a more thorough theoretical treatment of the minimum distance (MD) estimator. Specifically, using established theory from the econometrics literature, we outline conditions for consistency and asymptotic normality. Third, we present an expression for the asymptotic covariance matrix of the estimated parameters.

1.2. Organization

The study is organized as follows. Section 2 describes the EIV model with colored output noise. Section 3 outlines how to fit the EIV model into the CFA modeling framework. Section 4 introduces the estimator, its asymptotic properties and how to implement it using a separable nonlinear least squares (SNLLS) approach. Section 5 demonstrates the finite sample properties of the presented EIV-CFA framework. Finally, in Section 6, concluding remarks are given.

2. The EIV model with colored output noise

The considered model is a linear single-input single-output (SISO) system of the form

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

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Adrian Wills under the direction of Editor Alessandro Chiuso.

E-mail address: david.kreiberg@bi.no.

https://doi.org/10.1016/j.automatica.2023.111187

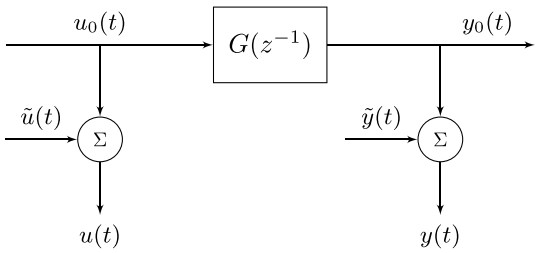

Fig. 1. The EIV model with $\tilde{y}(t)$ being an arbitrary process.

where $y_0(t)$ and $u_0(t)$ denote the unobserved noise-free input and output processes, respectively. Moreover, $A\left(z^{-1}\right)$ and $B\left(z^{-1}\right)$ are polynomials in the backward shift operator z^{-1}, which are explicitly written as

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

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

The parameter vector containing the polynomial coefficients is given by

$$\boldsymbol{\vartheta}_G = \begin{pmatrix} a_1 & \dots & a_{n_a} & b_1 & \dots & b_{n_b} \end{pmatrix}^T. \tag{4}$$

The noise-free processes are corrupted by additive measurement noises denoted $\tilde{y}(t)$ and $\tilde{u}(t)$. It follows that the available processes are given by

$$y(t) = y_0(t) + \tilde{y}(t), \quad u(t) = u_0(t) + \tilde{u}(t). \tag{5}$$

The presentation below considers the more general case of colored output noise with white noise as a special case. Fig. 1 summarizes the problem.

The following assumptions outline the characteristics of the EIV model and its components:

- **A1**. All processes are zero-mean ergodic and stationary.
- **A2**. The polynomials $A\left(z^{-1}\right)$ and $B\left(z^{-1}\right)$ are coprime and their respective degrees, given by n_a and n_b, are known.
- **A3**. The noise-free input process $u_0(t)$ is unknown as are its second-order properties.
- **A4**. The noise processes $\tilde{y}(t)$ and $\tilde{u}(t)$ are both mutually uncorrelated with the noise-free input process $u_0(t)$, and mutually uncorrelated with each other.
- **A5**. The output noise process $\tilde{y}(t)$ may or may not be colored, whereas the input noise process $\tilde{u}(t)$ is white.

Given the stated assumptions, the purpose is to outline a framework for estimating the parameter vector in (4) along with other parameters characterizing the EIV model.

3. Implementing the EIV-CFA modeling framework

3.1. Notation

First, we briefly introduce some general notation used throughout the presentation. Let $v(t)$ be a zero-mean random process with covariance function

$$r_v(\tau) = \mathbb{E}\left[v(t)v(t-\tau)\right], \quad \tau = 0, \pm 1, \pm 2, \dots, \tag{6}$$

where $\mathbb{E}$ is the expectation operator. Moreover, let $\mathbf{v}(t)$ be a zero-mean $p_v \times 1$ random vector. The associated $p_v \times p_v$ covariance matrix is then given by

$$\mathbf{R}_v = \mathbb{E}\left[\mathbf{v}(t)\mathbf{v}^T(t)\right]. \tag{7}$$

In this expression, T is the transpose of a vector or a matrix. For later, it will be useful to eliminate the redundancy that arise from the structure characterizing $\mathbf{R}_v$. For this purpose, let $\mathbf{K}_v$ be a matrix such that

$$\mathbf{r}_v = \mathbf{K}_v^T \operatorname{vec}\left(\mathbf{R}_v\right). \tag{8}$$

Here, $\mathbf{r}_v$ is a $q_v \times 1$ covariance vector containing the non-redundant elements of $\mathbf{R}_v$, and vec is the operation of vectorizing the elements of a matrix. The matrix $\mathbf{K}_v$ is obtained by

$$\mathbf{K}_v = \mathbf{L}_v\left(\mathbf{L}_v^T\mathbf{L}_v\right)^{-1}, \tag{9}$$

where $\mathbf{L}_v$ is a selection matrix of full column rank. Another useful expression involving $\mathbf{L}_v$ is

$$\operatorname{vec}\left(\mathbf{R}_v\right) = \mathbf{L}_v\mathbf{r}_v. \tag{10}$$

Appendix A outlines how to obtain $\mathbf{L}_v$ for different covariance structures relevant to this study.

3.2. The covariance structure characterizing the EIV model

Define the random vectors

$$\mathbf{y}(t) = \begin{pmatrix} y(t) & \dots & y(t-p_y+1) \end{pmatrix}^T, \tag{11}$$

$$\mathbf{u}(t) = \begin{pmatrix} u(t-1) & \dots & u(t-p_u) \end{pmatrix}^T. \tag{12}$$

In (11), p_y is an integer value determining the number of entries in $\mathbf{y}(t)$ starting at lag 0. Correspondingly, in (12), p_u is an integer value determining the number of entries in $\mathbf{u}(t)$ starting at lag 1. Vectors of similar structure are formed for the noise-free processes, denoted $\mathbf{y}_0(t)$ and $\mathbf{u}_0(t)$, and the noise processes, denoted $\tilde{\mathbf{y}}(t)$ and $\tilde{\mathbf{u}}(t)$. Next, introduce

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{y}(t) \\ \mathbf{u}(t) \end{pmatrix}, \quad \mathbf{x}_0(t) = \begin{pmatrix} \mathbf{y}_0(t) \\ \mathbf{u}_0(t) \end{pmatrix}, \quad \tilde{\mathbf{x}}(t) = \begin{pmatrix} \tilde{\mathbf{y}}(t) \\ \tilde{\mathbf{u}}(t) \end{pmatrix}, \tag{13}$$

where all vectors have $p_x = p_y + p_u$ entries. The covariance matrices associated with $\mathbf{x}(t)$, $\mathbf{x}_0(t)$ and $\tilde{\mathbf{x}}(t)$ all have a block Toeplitz structure. By assumption, the relation between the covariance matrices is

$$\mathbf{R}_x = \mathbf{R}_{x_0} + \mathbf{R}_{\tilde{x}}. \tag{14}$$

Here, $\mathbf{R}_{\tilde{x}}$ takes the form

$$\mathbf{R}_{\tilde{x}} = \begin{pmatrix} \mathbf{R}_{\tilde{y}} & \mathbf{0} \\ \mathbf{0} & r_{\tilde{u}}(0)\mathbf{I}_{p_u} \end{pmatrix}, \tag{15}$$

where $\mathbf{I}_{p_u}$ is a p_u-dimensional identity matrix.

3.3. Fitting the EIV model into the CFA modeling framework

The CFA model is given by the equation

$$\mathbf{x}(t) = \boldsymbol{\Lambda}\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}(t). \tag{16}$$

In the model, $\mathbf{x}(t)$ is a $p_x \times 1$ observed random vector, whereas $\boldsymbol{\xi}(t)$ and $\boldsymbol{\epsilon}(t)$ are $p_\xi \times 1$ and $p_x \times 1$ unobserved random vectors, respectively. The $p_x \times p_\xi$ matrix $\boldsymbol{\Lambda}$ contains the parameters relating the processes of $\boldsymbol{\xi}(t)$ to the processes of $\mathbf{x}(t)$. The processes of $\boldsymbol{\epsilon}(t)$ are measurement errors that are assumed to be mutually uncorrelated with the processes of $\boldsymbol{\xi}(t)$, but can be correlated among themselves. Given the model and the assumptions, the covariance structure becomes

$$\mathbf{R}_x(\boldsymbol{\vartheta}) = \boldsymbol{\Lambda}\mathbf{R}_\xi\boldsymbol{\Lambda}^T + \mathbf{R}_\epsilon, \tag{17}$$

where $\boldsymbol{\vartheta}$ is a parameter vector composed of the free elements of $\boldsymbol{\Lambda}$, $\mathbf{R}_\xi$ and $\mathbf{R}_\epsilon$.

When fitting the EIV model into (16), it will be convenient to introduce the following auxiliary process (see Söderström et al., 2009):

$$\xi(t) = \frac{1}{A\left(z^{-1}\right)}u_0(t). \tag{18}$$

The use of (18) allows us to express $y(t)$ and $u(t)$ using

$$y(t) = B\big(z^{-1}\big)\xi(t) + \tilde{y}(t), \tag{19}$$

$$u(t) = A\big(z^{-1}\big)\xi(t) + \tilde{u}(t). \tag{20}$$

Expanding (19) and (20), based on the choice of p_y and p_u, gives the equation system

$$\begin{aligned} y(t) &= B\big(z^{-1}\big)\xi(t) + \tilde{y}(t) \\ &\vdots \\ y(t-p_y+1) &= B\big(z^{-1}\big)\xi(t-p_y+1) + \tilde{y}(t-p_y+1) \\ u(t-1) &= A\big(z^{-1}\big)\xi(t-1) + \tilde{u}(t-1) \\ &\vdots \\ u(t-p_u) &= A\big(z^{-1}\big)\xi(t-p_u) + \tilde{u}(t-p_u). \end{aligned} \tag{21}$$

In (21), the collection of left-hand side terms forms the vector $\mathbf{x}(t)$ and the collection of right-hand side noise terms forms the vector $\boldsymbol{\epsilon}(t) = \tilde{\mathbf{x}}(t)$. The matrix $\boldsymbol{\Lambda}$ depends on the difference between $p_y + n_b - 1$ and $p_u + n_a$. In the case when $p_y + n_b - 1 \geq p_u + n_a$, $\boldsymbol{\Lambda}$ takes the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} b_1 & \cdots & \cdots & b_{n_b} & 0 & \cdots & 0 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & b_1 & \cdots & \cdots & b_{n_b} \\ 1 & \cdots & a_{n_a} & 0 & \cdots & 0 & \mathbf{0} \\ 0 & \ddots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & a_{n_a} & \mathbf{0} \end{pmatrix}, \tag{22}$$

where the $\mathbf{0}$'s are row vectors of zeros with number of elements equal to $p_y+n_b-1-(p_u+n_a)$. In the case when $p_u+n_a \geq p_y+n_b-1$, $\boldsymbol{\Lambda}$ becomes

$$\boldsymbol{\Lambda} = \begin{pmatrix} b_1 & \cdots & b_{n_b} & 0 & \cdots & 0 & \mathbf{0} \\ 0 & \ddots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & b_1 & \cdots & b_{n_b} & \mathbf{0} \\ 1 & \cdots & \cdots & a_{n_a} & 0 & \cdots & 0 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & \cdots & a_{n_a} \end{pmatrix}, \tag{23}$$

where the $\mathbf{0}$'s are row vectors of zeros with number of elements equal to $p_u + n_a - (p_y + n_b - 1)$.

The auxiliary vector is specified by

$$\boldsymbol{\xi}(t) = \begin{pmatrix} \xi(t-1) & \ldots & \xi(t-p_\xi) \end{pmatrix}^T. \tag{24}$$

The number of entries in $\boldsymbol{\xi}(t)$ is determined by

$$p_\xi = \max\big(p_y + n_b - 1, p_u + n_a\big). \tag{25}$$

Using (8) and (9), the covariance vectors containing the non-redundant elements of $\mathbf{R}_\xi$ and $\mathbf{R}_\epsilon = \mathbf{R}_{\tilde{\mathbf{x}}}$ are respectively

$$\mathbf{r}_\xi = \mathbf{K}_\xi^T \operatorname{vec}(\mathbf{R}_\xi), \qquad \mathbf{r}_\epsilon = \mathbf{K}_\epsilon^T \operatorname{vec}(\mathbf{R}_\epsilon). \tag{26}$$

From the preceding expressions, the parameter vector becomes

$$\boldsymbol{\vartheta} = \begin{pmatrix} \boldsymbol{\vartheta}_G^T & \boldsymbol{\vartheta}_{r_\xi, r_\epsilon}^T \end{pmatrix}^T, \tag{27}$$

where $\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}$ is a vector formed by stacking $\mathbf{r}_\xi$ and $\mathbf{r}_\epsilon$. The number of parameters in $\boldsymbol{\vartheta}$ is

$$n_\vartheta = n_a + n_b + q_\xi + q_\epsilon. \tag{28}$$

For the purpose of identification, p_y and p_u must be chosen so that the number of non-redundant elements of $\mathbf{R}_x$ is at least the size of n_ϑ (i.e., $q_x \geq n_\vartheta$).

4. The estimator

4.1. Minimum distance estimation

The MD estimator applied in this study belongs to a broader class known as extremum estimators. Over the years, this class of estimators has been thoroughly studied, and many general results have been presented in the literature. An overview was presented by Amemiya (1985).

Suppose that a set of data points of sample length N is available. Based on the data, an estimate of $\mathbf{R}_x$ is obtained by

$$\hat{\mathbf{R}}_x = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^T(t). \tag{29}$$

Applying the MD estimator involves searching the parameter space for the vector that minimizes the sum of the squared distances between corresponding elements of $\hat{\mathbf{R}}_x$ and $\mathbf{R}_x(\boldsymbol{\vartheta})$. Before stating the objective function, it is necessary to introduce the following $q_x \times 1$ vector-valued function

$$\boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big) = \mathbf{K}_x^T \operatorname{vec}\big(\mathbf{x}(t)\mathbf{x}^T(t) - \mathbf{R}_x(\boldsymbol{\vartheta})\big). \tag{30}$$

Taking the expectation of (30) leads to

$$\begin{aligned} \boldsymbol{\rho}_E(\boldsymbol{\vartheta}) &= \mathbb{E}\left[\boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)\right] \\ &= \mathbf{K}_x^T \operatorname{vec}\big(\mathbf{R}_x - \mathbf{R}_x(\boldsymbol{\vartheta})\big) \\ &= \mathbf{r}_x - \mathbf{r}_x(\boldsymbol{\vartheta}), \end{aligned} \tag{31}$$

which is estimated by the sample mean

$$\begin{aligned} \hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}) &= \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big) \\ &= \mathbf{K}_x^T \operatorname{vec}\big(\hat{\mathbf{R}}_x - \mathbf{R}_x(\boldsymbol{\vartheta})\big) \\ &= \hat{\mathbf{r}}_x - \mathbf{r}_x(\boldsymbol{\vartheta}). \end{aligned} \tag{32}$$

The objective function for estimating the true parameter vector $\boldsymbol{\vartheta}_0$ is now expressed as

$$\hat{V}(\boldsymbol{\vartheta}) = \hat{\boldsymbol{\rho}}^T(\boldsymbol{\vartheta})\widehat{\mathbf{W}}\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}), \tag{33}$$

where $\widehat{\mathbf{W}}$ is a $q_x \times q_x$ symmetric positive definite weighting matrix. This matrix is typically, but not necessarily, obtained from the data. The estimation problem becomes

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}\in\Theta}\big(\hat{V}(\boldsymbol{\vartheta})\big). \tag{34}$$

In this expression, Θ denotes the parameter space, which is a subset of the Euclidean space $\mathbb{R}^{n_\vartheta}$.

4.2. Asymptotic properties

Observe that the objective function in (33) is identical to the one used when applying generalized method of moments (GMM) (see Hansen, 1982). Thus, when investigating the asymptotic properties of the estimator, we can conveniently apply existing results from the econometrics literature. Texts of particular interest are those of Amemiya (1985), Hayashi (2011), Newey and McFadden (1994).

4.2.1. Consistency

Consider the deterministic function

$$V_E(\boldsymbol{\vartheta}) = \boldsymbol{\rho}_E^T(\boldsymbol{\vartheta})\mathbf{W}\boldsymbol{\rho}_E(\boldsymbol{\vartheta}), \tag{35}$$

where $\mathbf{W}$ is a symmetric positive definite matrix. Given that the EIV model is specified correctly (i.e., $\boldsymbol{\rho}_E(\boldsymbol{\vartheta}_0) = \mathbf{0}$), $V_E(\boldsymbol{\vartheta})$ is a non-negative function that attains its minimum value of 0 at $\boldsymbol{\vartheta}_0$. Moreover, if $\boldsymbol{\rho}_E(\boldsymbol{\vartheta})$ is a continuous function, then $V_E(\boldsymbol{\vartheta})$ is also continuous.

The estimator is consistent provided that $\hat{V}(\boldsymbol{\vartheta})$ converges in probability to $V_E(\boldsymbol{\vartheta})$ for all $\boldsymbol{\vartheta} \in \Theta$, and that $V_E(\boldsymbol{\vartheta})$ is uniquely minimized at $\boldsymbol{\vartheta}_0$. For this to happen, some technical requirements regarding the parameter space Θ, and the components making up $V_E(\boldsymbol{\vartheta})$, must be satisfied. The theorem below outlines the necessary conditions for $\hat{\boldsymbol{\vartheta}}$ to be a consistent estimator of $\boldsymbol{\vartheta}_0$.

Theorem 1 (*Consistency*)**.** *Suppose that; (i) The parameter space Θ is compact (i.e., closed and bounded); (ii) $\widehat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$, where $\xrightarrow{p}$ denotes convergence in probability; (iii) The function $\boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)$ is continuous for all $\boldsymbol{\vartheta} \in \Theta$; (iv) $\boldsymbol{\rho}_E(\boldsymbol{\vartheta}) \neq \mathbf{0}$ for all $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$; (v) $\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}) - \boldsymbol{\rho}_E(\boldsymbol{\vartheta})\right\| \xrightarrow{p} 0$, where $\|.\|$ is the Euclidean norm. Then*

$$\hat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_0. \tag{36}$$

Proof. See Amemiya (1985) and Newey and McFadden (1994) □

Remark 1.

- Conditions (i) and (iii) are technical requirements that play an important role in the proof of the theorem, and in the verification of other conditions.
- Condition (ii) ensures that $\widehat{\mathbf{W}}$ converges in probability to a symmetric positive definite matrix. Several valid choices for the weighting matrix satisfy this condition. For instance, a simple and easy-to-implement choice is the identity matrix.
- Condition (iv) guarantees that $\boldsymbol{\vartheta}_0$ is uniquely identified; that is, $\boldsymbol{\vartheta}_0$ is the sole value that satisfies the equation $\boldsymbol{\rho}_E(\boldsymbol{\vartheta}) = \mathbf{0}$. This condition can be equivalently formulated as $\mathbf{r}_x \neq \mathbf{r}_x(\boldsymbol{\vartheta})$ for all $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$. Verifying global identification is typically difficult. Local identification is verified given some mild technical requirements and that the Jacobian matrix of $\boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)$ has full column rank.
- Condition (v) is a statement that ensures that $\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta})$ converges uniformly in probability to $\boldsymbol{\rho}_E(\boldsymbol{\vartheta})$. This condition is verified given (i), (iii), and the existence of certain fourth-order moments.

In what follows, Conditions (i) and (iv) in Theorem 1 are assumed to hold. Condition (ii) holds by choosing a suitable weighting matrix and Condition (iii) is verified by inspection. In Appendix B, it is demonstrated that Condition (v) holds under some mild moment conditions.

4.2.2. Asymptotic normality

The principal tool for deriving asymptotic normality is the mean value theorem (MVT). Applying the MVT involves the following $q_x \times n_\vartheta$ Jacobian matrix

$$\mathbf{J}(\boldsymbol{\vartheta}) = \frac{\partial \boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)}{\partial \boldsymbol{\vartheta}^T}. \tag{37}$$

The expectation of (37) is given by

$$\mathbf{J}_E(\boldsymbol{\vartheta}) = \mathbb{E}\left[\frac{\partial \boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)}{\partial \boldsymbol{\vartheta}^T}\right], \tag{38}$$

and the sample mean is

$$\begin{aligned}\hat{\mathbf{J}}(\boldsymbol{\vartheta}) &= \frac{1}{N}\sum_{t=1}^{N} \frac{\partial \boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)}{\partial \boldsymbol{\vartheta}^T} \\ &= \frac{\partial \hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^T}.\end{aligned} \tag{39}$$

The theorem below summarizes the necessary conditions for asymptotic normality.

Theorem 2 (*Asymptotic Normality*)**.** *Suppose that; (i) $\hat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_0$; (ii) $\boldsymbol{\vartheta}_0$ is an interior point in Θ; (iii) $\widehat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$; (iv) $\boldsymbol{\rho}\big(\mathbf{x}(t), \boldsymbol{\vartheta}\big)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $\boldsymbol{\vartheta}_0$; (v) $\mathbf{J}_E(\boldsymbol{\vartheta}_0)$ has full column rank; (vi) $\sqrt{N}\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \boldsymbol{\Gamma}]$, where $\xrightarrow{d}$ denotes convergence in distribution and $\boldsymbol{\Gamma}$ is a $q_x \times q_x$ covariance matrix; (vii) $\sup_{\boldsymbol{\vartheta}\in\mathcal{N}} \left\|\hat{\mathbf{J}}(\boldsymbol{\vartheta}) - \mathbf{J}_E(\boldsymbol{\vartheta})\right\| \xrightarrow{p} 0$. Then*

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \boldsymbol{\Omega}], \tag{40}$$

where $\boldsymbol{\Omega}$ is the $n_\vartheta \times n_\vartheta$ asymptotic covariance matrix given by the expression

$$\boldsymbol{\Omega} = \big(\mathbf{J}_E^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}_E(\boldsymbol{\vartheta}_0)\big)^{-1}\mathbf{J}_E^T(\boldsymbol{\vartheta}_0)\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\mathbf{J}_E(\boldsymbol{\vartheta}_0)\big(\mathbf{J}_E^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}_E(\boldsymbol{\vartheta}_0)\big)^{-1}. \tag{41}$$

Proof. See Amemiya (1985) and Newey and McFadden (1994) □

Remark 2.

- Conditions (i), (ii), and (iv) are technical requirements necessary for applying the MVT.
- Condition (v) ensures that $\mathbf{J}_E^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}_E(\boldsymbol{\vartheta}_0)$ is non-singular.
- Condition (vi) is essential when determining the mean and variance of the asymptotic distribution. Given certain requirements, this condition holds when the observed processes $y(t)$ and $u(t)$ are linear with independent and identically distributed (*i.i.d.*) innovations.
- Condition (vii) is a statement that ensures that $\hat{\mathbf{J}}(\boldsymbol{\vartheta})$ converges uniformly in probability to $\mathbf{J}_E(\boldsymbol{\vartheta})$ in a neighborhood of $\boldsymbol{\vartheta}_0$.

In the following presentation, Conditions (i)–(vi) in Theorem 2 are assumed to hold. As demonstrated in Appendix B, Condition (vii) holds without further requirements on the data.

From (30), it is immediately clear that $\mathbf{J}(\boldsymbol{\vartheta})$ does not involve $\mathbf{x}(t)$. Thus, for the case at hand, the Jacobian matrix simplifies to

$$\mathbf{J}(\boldsymbol{\vartheta}) = -\frac{\partial \mathbf{r}_x(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^T}. \tag{42}$$

The asymptotic covariance matrix in (41) is then written using the somewhat simpler expression

$$\boldsymbol{\Omega} = \big(\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big)^{-1}\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big(\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big)^{-1}, \tag{43}$$

where $\mathbf{J}(\boldsymbol{\vartheta}_0)$ is (42) evaluated at $\boldsymbol{\vartheta}_0$. On a final note, it can be shown that $\boldsymbol{\Omega}$ has a lower bound. Specifically, $\boldsymbol{\Omega}$ satisfies the inequality

$$\begin{aligned}&\big(\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big)^{-1}\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big(\mathbf{J}^T(\boldsymbol{\vartheta}_0)\mathbf{W}\mathbf{J}(\boldsymbol{\vartheta}_0)\big)^{-1} \\ &\geq \big(\mathbf{J}^T(\boldsymbol{\vartheta}_0)\boldsymbol{\Gamma}^{-1}\mathbf{J}(\boldsymbol{\vartheta}_0)\big)^{-1}.\end{aligned} \tag{44}$$

Here, the right-hand side is obtained by setting $\mathbf{W}$ equal to $\boldsymbol{\Gamma}^{-1}$. Given the specification, it trivially follows that $\boldsymbol{\Gamma}^{-1}$ is an optimal weighting matrix. Obviously, in a practical setting, this matrix has to be estimated. In this study, we do not attempt to suggest how to estimate $\boldsymbol{\Gamma}^{-1}$.

4.3. A numerically more efficient implementation of the MD estimator

The standard implementation of the MD estimator is a one-step approach, where $\hat{V}(\boldsymbol{\vartheta})$ is minimized using nonlinear optimization techniques. This approach is typically referred to as nonlinear least squares (NLLS). In contrast, as demonstrated below, the SNLLS implementation is a two-step approach. Examples of how to apply SNLLS for estimating CFA models are given by Kreiberg, Marcoulides, and Olsson (2021).

Suppose that the vector-valued function $\mathbf{r}_x(\boldsymbol{\vartheta})$ can be written as

$$\mathbf{r}_x(\boldsymbol{\vartheta}) = \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}, \tag{45}$$

where $\mathbf{F}(\boldsymbol{\vartheta}_G)$ is a tall matrix-valued function assumed to have full column rank. Appendix C shows that $\mathbf{F}(\boldsymbol{\vartheta}_G)$ takes the form

$$\mathbf{F}(\boldsymbol{\vartheta}_G) = \begin{pmatrix} \mathbf{K}_x^T(\boldsymbol{\Lambda} \otimes \boldsymbol{\Lambda})\mathbf{L}_\xi & \mathbf{K}_x^T\mathbf{L}_\epsilon \end{pmatrix}. \tag{46}$$

Applying (45) allows us to express the objective function in (33) using

$$\hat{V}(\boldsymbol{\vartheta}) = \left(\hat{\mathbf{r}}_x - \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}\right)^T \widehat{\mathbf{W}} \left(\hat{\mathbf{r}}_x - \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}\right). \tag{47}$$

For some value of $\boldsymbol{\vartheta}_G$, the solution to the problem of minimizing $\hat{V}(\boldsymbol{\vartheta})$ w.r.t. to $\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}$ is a straightforward application of least squares

$$\hat{\boldsymbol{\vartheta}}_{r_\xi, r_\epsilon}(\boldsymbol{\vartheta}_G) = \left(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\right)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x. \tag{48}$$

As shown in Appendix C, the SNLLS objective function for estimating the parameters in $G(z^{-1})$ is derived from replacing $\boldsymbol{\vartheta}_{r_\xi, r_\epsilon}$ in (47) with the right-hand side of (48), which leads to

$$\tilde{V}(\boldsymbol{\vartheta}_G) = \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\hat{\mathbf{r}}_x - \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\left(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\right)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x. \tag{49}$$

Estimation is performed in two steps. First, $\hat{\boldsymbol{\vartheta}}_G$ is obtained by minimizing (49) using nonlinear optimization techniques. Second, $\hat{\boldsymbol{\vartheta}}_{r_\xi, r_\epsilon}$ is computed by (48) using $\hat{\boldsymbol{\vartheta}}_G$ from the first step. The theoretical justification underlying the two-step approach is attributable to Golub and Pereyra (1973), who showed that SNLLS provides the exact same minimum function value and parameter estimates as does NLLS. It obviously follows that the asymptotic properties of the estimator are maintained.

5. Simulation example

The purpose of this section is to provide a brief simulation example demonstrating the CFA implementation of the EIV model.

5.1. Model and setup

The considered model is a second-order system given by

$$y_0(t) = \frac{1.0z^{-1} - 0.7z^{-2}}{1.0 - 1.2z^{-1} + 0.5z^{-2}}u_0(t), \tag{50}$$

where $u_0(t)$ is an autoregressive moving average (ARMA) process of the form

$$u_0(t) = \frac{1.0 + 0.7z^{-1}}{1.0 - 0.5z^{-1}}\eta(t). \tag{51}$$

Colored output noise is generated by the following moving average (MA) process:

$$\tilde{y}(t) = \left(1.0 + 0.8z^{-1} + 0.5z^{-2}\right)\varepsilon(t). \tag{52}$$

The innovation processes $\eta(t)$ and $\varepsilon(t)$, and the input noise process $\tilde{u}(t)$, are Gaussian zero-mean white noise with variances

$$\mathbb{E}\left[\eta^2(t)\right] = 0.6, \quad \mathbb{E}\left[\varepsilon^2(t)\right] = 1.0, \quad \mathbb{E}\left[\tilde{u}^2(t)\right] = 0.8. \tag{53}$$

Empirical means and standard errors (std. errors) of the estimated parameters are obtained under the following conditions:

- The model is simulated 2000 times using a sample length of $N = 2000$. When simulating time-dependent data, a good practice is to eliminate initial data points up to the point at which the process has stabilized. For this example, the final half of 4000 simulated data points makes up the final sample.
- The user choices are

$$\{p_y, p_u\} = \{12 + j, 11 + j\} \quad \text{for } j = 0, \ldots, 4. \tag{54}$$

- Initial values for the minimization of (49) are obtained from the IV estimator using delayed input as instruments. The number of instruments is set to 100.
- The output noise is assumed to be an arbitrary linear process with an unknown covariance structure.
- For simplicity, estimation is performed by applying equal weighting; that is, $\mathbf{W} = \mathbf{I}$.
- Theoretical std. errors are obtained using (43). Söderström et al. (2014) provide expressions for how to compute the elements of $\boldsymbol{\Gamma}$ in the case of white output noise; only minor modifications to these expressions are necessary to accommodate colored output noise.
- All estimation is performed in Matlab using standard optimization techniques. Implementation is based on the SNLLS approach as previously outlined.

5.2. Results

For simplicity, the presentation below is limited to reporting results only for the parameters characterizing the transfer function $G(z^{-1})$ and the non-zero covariance parameters of $\mathbf{R}_\epsilon$.

Results 1. *The empirical means of the simulated parameters are summarized in Table 1. For the parameters in $G(z^{-1})$, it is seen from Table 1(a) that the mean estimates are close to the true values, indicating low estimation bias. For the covariance parameters characterizing the output noise process, Table 1(b) shows that the mean estimates are somewhat smaller than the true values, indicating some downward bias. There is an ever-so-slight deterioration for larger values of p_y and p_u. For the covariance parameter characterizing the input noise process, deviations from the true value are small.*

Results 2. *Figs. 2 and 3 summarize the empirical and theoretical std. errors. For the parameters in $G(z^{-1})$, it is seen from the diagrams in Fig. 2 that the discrepancies between the empirical and theoretical values are small. Note the increase in the std. errors as p_y and p_u get larger. For the covariance parameters characterizing the output noise process, the diagrams in Fig. 3 show that the empirical values tend to be smaller than the theoretical values. This is especially the case for $r_{\tilde{y}}(0)$ and $r_{\tilde{y}}(1)$ when p_y and p_u are small. For the covariance parameter characterizing the input noise process, there is close agreement between the empirical and theoretical values.*

5.3. Practical considerations

It is well known that nonlinear optimization is subject to numerical instability. To prevent optimizer failure, it is worthwhile to consider the following: (1) avoid choosing a too small number of instruments when computing the starting values, and (2) avoid choosing p_y and p_u too small. The term "too small" is somewhat ambiguous. The previous points are simply to highlight the fact that the number of instruments and the values of p_y and p_u have an impact on optimizer performance.

Table 1
Simulated means and standard errors.

(a) Parameters characterizing the transfer function $G(z^{-1})$

$\{p_y, p_u\}$	Parameter values			
	a_1	a_2	b_1	b_2
	−1.20	0.50	1.00	−0.70
{12, 11}	−1.1982 (±0.0397)	0.5011 (±0.0309)	1.0032 (±0.0688)	−0.6978 (±0.0804)
{13, 12}	−1.1977 (±0.0407)	0.5017 (±0.0308)	1.0032 (±0.0692)	−0.6963 (±0.0828)
{14, 13}	−1.1978 (±0.0410)	0.5022 (±0.0311)	1.0035 (±0.0698)	−0.6959 (±0.0853)
{15, 14}	−1.1984 (±0.0410)	0.5023 (±0.0315)	1.0042 (±0.0704)	−0.6971 (±0.0874)
{16, 15}	−1.1989 (±0.0409)	0.5021 (±0.0322)	1.0051 (±0.0711)	−0.6989 (±0.0894)

(b) Parameters characterizing the covariance vector $\mathbf{R}_\epsilon$

$\{p_y, p_u\}$	Parameter values			
	$r_{\tilde{y}}(0)$	$r_{\tilde{y}}(1)$	$r_{\tilde{y}}(2)$	$r_{\tilde{u}}(0)$
	1.89	1.20	0.50	0.80
{12, 11}	1.8599 (±0.1976)	1.1734 (±0.1541)	0.4845 (±0.0816)	0.8004 (±0.0810)
{13, 12}	1.8568 (±0.1996)	1.1706 (±0.1562)	0.4835 (±0.0828)	0.8017 (±0.0816)
{14, 13}	1.8531 (±0.2007)	1.1672 (±0.1572)	0.4820 (±0.0836)	0.8033 (±0.0821)
{15, 14}	1.8499 (±0.2013)	1.1643 (±0.1578)	0.4804 (±0.0840)	0.8076 (±0.0824)
{16, 15}	1.8484 (±0.2015)	1.1629 (±0.1581)	0.4794 (±0.0842)	0.8048 (±0.0825)

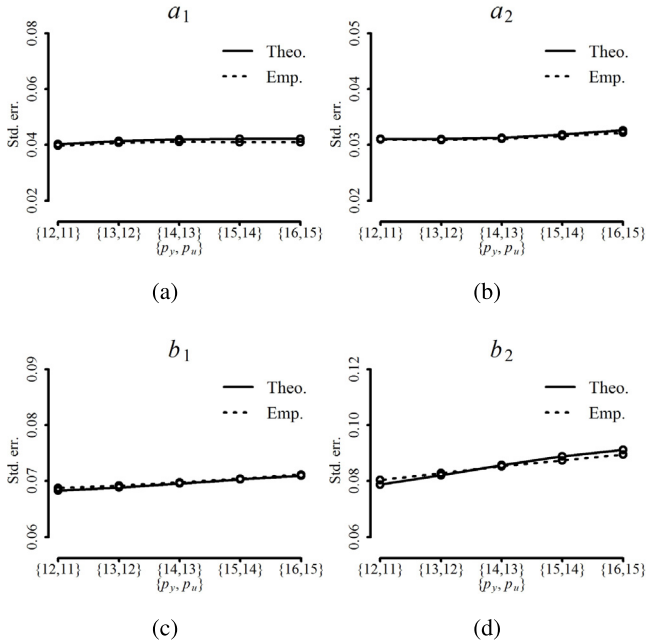

Fig. 2. Parameters characterizing the transfer function $G(z^{-1})$.

6. Concluding remarks

In this study, we have presented an EIV-CFA modeling framework that allows for colored output noise. The presentation included an in-depth treatment of the theoretical aspects of applying the MD estimator. Using established theory from the econometrics literature, we outlined conditions for consistency and asymptotic normality. The finite-sample performance of the estimator was briefly evaluated using simulated data. The simulation results suggest that the MD estimator performs well in reproducing the true model parameters.

We did not address the issue of how to compute an empirical weight matrix when applying the estimator. The effect of weighting on the precision of the estimated parameters is an important topic that deserves a more thorough treatment than what was possible here.

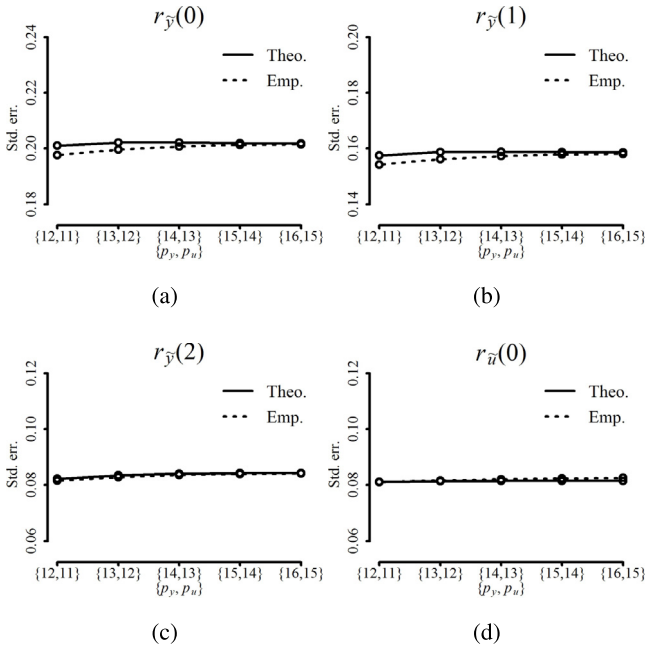

Fig. 3. Parameters characterizing the covariance matrix $\mathbf{R}_\epsilon$.

Acknowledgments

The author would like to thank Professor Torsten Söderström at Uppsala University, the Editor and the anonymous referees for their valuable comments and suggestions.

Appendix A. Deriving the matrix K

Let $\mathbf{R} = \{r_{i,j}\}$ be a $p \times p$ covariance matrix, where $r_{i,j}$ is the element at the ith row and the jth column. Moreover, let $\mathbf{K}$ be a $p^2 \times q$ matrix such that

$$\mathbf{r} = \mathbf{K}^T \operatorname{vec}(\mathbf{R}). \tag{55}$$

In this expression, $\mathbf{r}$ is a $q \times 1$ covariance vector containing the non-redundant elements of $\mathbf{R}$, and $\mathbf{K}$ is obtained from

$$\mathbf{K} = \mathbf{L}\left(\mathbf{L}^T\mathbf{L}\right)^{-1}, \tag{56}$$

where $\mathbf{L}$ is a $p^2 \times q$ selection matrix. It is now shown how to derive $\mathbf{L}$ for various covariance structures relevant to this study. We will start by introducing some necessary notation.

Let $\mathbf{E}(u, v) = \{e_{i,j}(u, v)\}$ denote a $p \times p$ matrix for $i = 1, \ldots, p$, $j = 1, \ldots, p$, and with elements

$$e_{i,j}(u, v) = \begin{cases} 1 & \text{if } r_{i,j} = r_{u,v} \\ 0 & \text{otherwise.} \end{cases} \tag{57}$$

Case 1: As a start, consider the case when $\mathbf{R}$ has a Toeplitz structure with no additional restrictions, other than symmetry,

placed on its elements. In this simple case, the first column of $\mathbf{R}$ makes up the $q = p$ non-redundant elements

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & \dots & r_{p,1} \end{pmatrix}^T. \tag{58}$$

Applying (57), the matrix $\mathbf{L}$ is formed by horizontally concatenating q column vectors in the following way

$$\mathbf{L} = \begin{pmatrix} \operatorname{vec}(\mathbf{E}(1,1)) & \dots & \operatorname{vec}(\mathbf{E}(p,1)) \end{pmatrix}. \tag{59}$$

Case 2: Next, consider the somewhat more involved case when $\mathbf{R}$ has a block Toeplitz structure with no additional restrictions, other than symmetry, placed on its elements. Let $\mathbf{R}$ be given by the expression

$$\mathbf{R} = \begin{pmatrix} \underset{(p_1\times p_1)}{\mathbf{R}_{1,1}} & \underset{(p_1\times p_2)}{\mathbf{R}_{1,2}} \\ \underset{(p_2\times p_1)}{\mathbf{R}_{2,1}} & \underset{(p_2\times p_2)}{\mathbf{R}_{2,2}} \end{pmatrix}, \tag{60}$$

where $p = p_1 + p_2$. The covariance vector containing the $q = 2p - 1$ non-redundant elements of $\mathbf{R}$ now becomes

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & \dots & r_{p,1} & r_{2,p_1+1} & \dots & r_{p,p_1+1} \end{pmatrix}^T. \tag{61}$$

As in the previous case, $\mathbf{L}$ is constructed by concatenating q column vectors

$$\begin{aligned} \mathbf{L} = \begin{pmatrix} \operatorname{vec}(\mathbf{E}(1,1)) & \dots & \operatorname{vec}(\mathbf{E}(p,1)) & \operatorname{vec}(\mathbf{E}(2,p_1+1)) \\ \dots & \operatorname{vec}(\mathbf{E}(p,p_1+1)) \end{pmatrix}. \end{aligned} \tag{62}$$

Case 3: Finally, consider the case when $\mathbf{R}$ has the structure

$$\mathbf{R} = \begin{pmatrix} \underset{(p_1\times p_1)}{\mathbf{R}_{1,1}} & \underset{(p_1\times p_2)}{\mathbf{0}} \\ \underset{(p_2\times p_1)}{\mathbf{0}} & r_{p_1+1}\mathbf{I}_{p_2} \end{pmatrix}. \tag{63}$$

First, suppose that $\mathbf{R}_{1,1}$ is a Toeplitz matrix that takes the form

$$\mathbf{R}_{1,1} = \begin{pmatrix} r_1 & \cdots & r_{p_1} \\ \vdots & \ddots & \vdots \\ r_{p_1} & \cdots & r_1 \end{pmatrix}. \tag{64}$$

The covariance vector consisting of the $q = p_1 + 1$ non-redundant elements of $\mathbf{R}$ becomes

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & \dots & r_{p_1,1} & r_{p_1+1,p_1+1} \end{pmatrix}^T. \tag{65}$$

It follows that

$$\mathbf{L} = \begin{pmatrix} \operatorname{vec}(\mathbf{E}(1,1)) & \dots & \operatorname{vec}(\mathbf{E}(p_1,1)) & \operatorname{vec}(\mathbf{E}(p_1+1,p_1+1)) \end{pmatrix}. \tag{66}$$

Second, suppose that $\mathbf{R}_{1,1}$ is a Toeplitz matrix with the additional restrictions that some elements are zero. Specifically, let $\mathbf{R}_{1,1}$ be a matrix formed by

$$\mathbf{R}_{1,1} = \begin{pmatrix} r_1 & \cdots & r_l & 0 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ r_l & & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & & r_l \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & r_l & \cdots & r_1 \end{pmatrix}, \tag{67}$$

where $l < p_1$. In this case, the covariance vector consisting of the $q = l + 1$ non-redundant elements of $\mathbf{R}$ is

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & \dots & r_{l,1} & r_{p_1+1,p_1+1} \end{pmatrix}^T. \tag{68}$$

The matrix $\mathbf{L}$ is now given by

$$\mathbf{L} = \begin{pmatrix} \operatorname{vec}(\mathbf{E}(1,1)) & \dots & \operatorname{vec}(\mathbf{E}(l,1)) & \operatorname{vec}(\mathbf{E}(p_1+1,p_1+1)) \end{pmatrix}. \tag{69}$$

Appendix B. Mathematical details regarding the asymptotic results

B.1. Theorem 1

Let $\boldsymbol{\rho}_E(\boldsymbol{\vartheta})$, $\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta})$, $\hat{V}(\boldsymbol{\vartheta})$ and $V_E(\boldsymbol{\vartheta})$ be functions as given in (31), (32), (33) and (35), respectively. Throughout the presentation, it is understood that $\hat{\boldsymbol{\vartheta}}$, $\hat{V}(\boldsymbol{\vartheta})$ and $V_E(\boldsymbol{\vartheta})$ are measurable functions (i.e., well-defined functions).

Consistent estimation requires that $\hat{V}(\boldsymbol{\vartheta})$ converge uniformly in probability to $V_E(\boldsymbol{\vartheta})$. As shown by Newey and McFadden (1994), this requirement is satisfied when

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\| \hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}) - \boldsymbol{\rho}_E(\boldsymbol{\vartheta}) \right\| \xrightarrow{p} 0. \tag{70}$$

In order to establish (70), it is useful to introduce the following lemma:

Lemma 1 (*Uniform Law of Large Numbers*). *Suppose that; (i) The parameter space Θ is compact; (ii) The process $\mathbf{x}(t)$ is ergodic and stationary; (iii) $\mathbf{g}(\mathbf{x}(t), \boldsymbol{\vartheta})$ is a continuous vector-valued function of $\boldsymbol{\vartheta}$ for all $\boldsymbol{\vartheta} \in \Theta$; (iv) There exists a dominance function $d(\mathbf{x}(t))$ such that $\left\|\mathbf{g}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\| \leq d(\mathbf{x}(t))$ for all $\boldsymbol{\vartheta} \in \Theta$ and $\mathbb{E}\left[d(\mathbf{x}(t))\right] < \infty$. Then, $\mathbb{E}\left[\mathbf{g}(\mathbf{x}(t), \boldsymbol{\vartheta})\right]$ is a continuous function of $\boldsymbol{\vartheta}$, and*

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left\| \frac{1}{N}\sum_{t=1}^{N} \mathbf{g}(\mathbf{x}(t), \boldsymbol{\vartheta}) - \mathbb{E}\left[\mathbf{g}(\mathbf{x}(t), \boldsymbol{\vartheta})\right] \right\| \xrightarrow{p} 0. \tag{71}$$

Applying Lemma 1, a function $d(\mathbf{x}(t))$, for which it can be established that $\mathbb{E}\left[d(\mathbf{x}(t))\right]$ is finite, is sought. By definition

$$\left\|\boldsymbol{\rho}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\| \leq \sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\boldsymbol{\rho}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\|. \tag{72}$$

Thus, an obvious choice for the dominance function is the right-hand side of (72), from which it follows that $\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta})$ converges uniformly in probability to $\boldsymbol{\rho}_E(\boldsymbol{\vartheta})$ provided that

$$\mathbb{E}\left[\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\boldsymbol{\rho}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\|\right] < \infty. \tag{73}$$

Since Θ is compact, the elements of $\mathbf{r}_x(\boldsymbol{\vartheta})$ are bounded. Then, by continuity, there exists a positive constant c such that

$$\begin{aligned} \sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\boldsymbol{\rho}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\| &= \sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\mathbf{K}_x^T \operatorname{vec}(\mathbf{x}(t)\mathbf{x}^T(t)) - \mathbf{r}_x(\boldsymbol{\vartheta})\right\| \\ &\leq \left\|\mathbf{K}_x^T \operatorname{vec}(\mathbf{x}(t)\mathbf{x}^T(t))\right\| + c. \end{aligned} \tag{74}$$

Showing that the expectation of (74) is finite, it needs to be established that the expectation of the first term on the right-hand side is finite. By Jensen's inequality, the expectation can be written as

$$\begin{aligned} \mathbb{E}\left[\sup_{\boldsymbol{\vartheta}\in\Theta} \left\|\boldsymbol{\rho}(\mathbf{x}(t), \boldsymbol{\vartheta})\right\|\right] &\leq \mathbb{E}\left[\left\|\mathbf{K}_x^T \operatorname{vec}(\mathbf{x}(t)\mathbf{x}^T(t))\right\|\right] + c \\ &\leq \sqrt{\mathbb{E}\left[\left\|\mathbf{K}_x^T \operatorname{vec}(\mathbf{x}(t)\mathbf{x}^T(t))\right\|^2\right]} + c, \end{aligned} \tag{75}$$

which is finite given that

$$\begin{aligned} &\mathbb{E}\left[y^2(t)y^2(t-\tau)\right] < \infty, \quad \tau = 0, \dots, p_y - 1, \\ &\mathbb{E}\left[y^2(t)u^2(t-\tau)\right] < \infty, \quad \tau = 1, \dots, p_u, \\ &\mathbb{E}\left[u^2(t-1)y^2(t-\tau)\right] < \infty, \quad \tau = 1, \dots, p_y - 1, \\ &\mathbb{E}\left[u^2(t-1)u^2(t-\tau)\right] < \infty, \quad \tau = 1, \dots, p_u. \end{aligned} \tag{76}$$

B.2. Theorem 2

Let $\mathbf{J}(\boldsymbol{\vartheta})$ be the Jacobian matrix as given in (37). Condition (vii) in Theorem 2 is easily verified by applying Lemma 1. By the same logic as in (72) and (73), it is sufficient to establish that

$$\mathbb{E}\left[\sup_{\boldsymbol{\vartheta}\in\mathcal{N}}\|\mathbf{J}(\boldsymbol{\vartheta})\|\right] < \infty. \tag{77}$$

Since $\mathbf{J}(\boldsymbol{\vartheta})$ does not involve $\mathbf{x}(t)$, it immediately follows that (77) is satisfied given that $\mathbf{J}(\boldsymbol{\vartheta})$ is continuous in a neighborhood of $\boldsymbol{\vartheta}_0$.

Appendix C. Mathematical details regarding the SNLLS implementation of the estimator

Start by showing that the vector-valued function $\mathbf{r}_x(\boldsymbol{\vartheta})$ can be written in the form

$$\mathbf{r}_x(\boldsymbol{\vartheta}) = \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}, \tag{78}$$

where $\mathbf{F}(\boldsymbol{\vartheta}_G)$ is given by (46). Applying (8) to the expression in (17), and by standard matrix operations, we have

$$\begin{aligned}\mathbf{r}_x(\boldsymbol{\vartheta}) &= \mathbf{K}_x^T \operatorname{vec}\big(\mathbf{R}_x(\boldsymbol{\vartheta})\big)\\ &= \mathbf{K}_x^T \operatorname{vec}\big(\boldsymbol{\Lambda}\mathbf{R}_\xi\boldsymbol{\Lambda}^T + \mathbf{R}_\epsilon\big)\\ &= \mathbf{K}_x^T(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda})\operatorname{vec}(\mathbf{R}_\xi) + \mathbf{K}_x^T\operatorname{vec}(\mathbf{R}_\epsilon).\end{aligned} \tag{79}$$

Based on (10), use that

$$\operatorname{vec}(\mathbf{R}_\xi) = \mathbf{L}_\xi\mathbf{r}_\xi, \quad \operatorname{vec}(\mathbf{R}_\epsilon) = \mathbf{L}_\epsilon\mathbf{r}_\epsilon, \tag{80}$$

from which it follows that

$$\begin{aligned}\mathbf{r}_x(\boldsymbol{\vartheta}) &= \mathbf{K}_x^T(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda})\mathbf{L}_\xi\mathbf{r}_\xi + \mathbf{K}_x^T\mathbf{L}_\epsilon\mathbf{r}_\epsilon\\ &= \begin{pmatrix}\mathbf{K}_x^T(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda})\mathbf{L}_\xi & \mathbf{K}_x^T\mathbf{L}_\epsilon\end{pmatrix}\begin{pmatrix}\mathbf{r}_\xi\\ \mathbf{r}_\epsilon\end{pmatrix}\\ &:= \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}.\end{aligned} \tag{81}$$

It is now shown how the MD objective function can be expressed entirely as a function of $\boldsymbol{\vartheta}_G$. Expand (33) to get

$$\begin{aligned}\hat{V}(\boldsymbol{\vartheta}) &= \hat{\boldsymbol{\rho}}^T(\boldsymbol{\vartheta})\widehat{\mathbf{W}}\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta})\\ &= \big(\hat{\mathbf{r}}_x - \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}\big)^T\widehat{\mathbf{W}}\big(\hat{\mathbf{r}}_x - \mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}\big)\\ &= \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\hat{\mathbf{r}}_x - 2\hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon} + \boldsymbol{\vartheta}_{r_\xi,r_\epsilon}^T\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}.\end{aligned} \tag{82}$$

Replacing $\boldsymbol{\vartheta}_{r_\xi,r_\epsilon}$ in (82) with the right-hand side of (48) yields

$$\begin{aligned}\tilde{V}(\boldsymbol{\vartheta}_G) &= \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\hat{\mathbf{r}}_x - 2\hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x\\ &\quad + \Big(\big(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x\Big)^T\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\\ &\quad \times \big(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x\\ &= \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\hat{\mathbf{r}}_x - \hat{\mathbf{r}}_x^T\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big(\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\mathbf{F}(\boldsymbol{\vartheta}_G)\big)^{-1}\mathbf{F}^T(\boldsymbol{\vartheta}_G)\widehat{\mathbf{W}}\hat{\mathbf{r}}_x,\end{aligned} \tag{83}$$

which confirms (49).

References

Amemiya, Takeshi (1985). *Advanced econometrics*. Harvard University Press.

Beghelli, Sergio, Guidorzi, Roberto P., & Soverini, Umberto (1990). The Frisch scheme in dynamic system identification. *Automatica*, *26*(1), 171–176.

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).

Ekman, Mats (2005). Identification of linear system with errors in variables using separable nonlinear least-squares. *IFAC Proceedings Volumes*, *38*(1), 815–820.

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.

Hansen, Lars Peter (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Hayashi, Fumio (2011). *Econometrics*. Princeton University Press.

Kreiberg, David, Marcoulides, Katerina, & Olsson, Ulf Henning (2021). A faster procedure for estimating CFA models applying minimum distance estimators with a fixed weight matrix. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(5), 725–739.

Kreiberg, David, Söderström, Torsten, & Yang-Wallentin, Fan (2016). Errors-in-variables system identification using structural equation modeling. *Automatica*, *66*, 218–230.

Mossberg, Magnus, & Söderström, Torsten (2012). On covariance matching for multiple input multiple output errors-in-variables systems. *IFAC Proceedings Volumes*, *45*(16), 1371–1376.

Newey, Whitney K., & McFadden, Daniel (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. 4* (pp. 2111–2245).

Söderström, Torsten (2007). Error-in-variables methods in system identification. *Automatica*, *43*(6), 939–958.

Söderström, Torsten (2011). A generalized instrumental variable method for errors-in-variables. *Automatica*, *47*(8), 1656–1666.

Söderström, Torsten (2018). *Errors-in-variables methods in system identification*. Springer International Publishing.

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.

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 BI Norwegian Business School, Norway in 2010 and the Ph.D. degree in statistics from Uppsala University, Sweden in 2022. David is currently holding a lectureship in statistics at BI Norwegian Business School. His main research interests include multivariate statistical analysis, latent variable modeling and system identification.