UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 200

A covariance structure analysis approach to the errors-in-variables estimation problem

DAVID KREIBERG

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1652-9030
ISBN 978-91-513-1583-6
URN urn:nbn:se:uu:diva-482081

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Thursday, 13 October 2022 at 08:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Håkan Hjalmarsson (Department of Intelligent Systems, Kungliga Tekniska Högskolan Stockholm).

Abstract
Kreiberg, D. 2022. A covariance structure analysis approach to the errors-in-variables estimation problem. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 200. 37 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1583-6.

It is a well-known fact that standard regression techniques, when applied to errors-in-variables (EIV) models, lead to biased and inconsistent parameter estimation. The work presented in this thesis address the EIV estimation problem using covariance structure analysis (CSA). When performing CSA, the standard implementation of the minimum distance (MD) estimator is to apply computationally demanding nonlinear least squares (NLLS). This thesis provides a solution to this problem by proposing a computationally less demanding separable nonlinear least squares (SNLLS) implementation of the estimator.

The thesis consists of four papers. The first paper presents a covariance matching (CM) approach for identifying the single-input single-output (SISO) EIV model. The outlined approach extends previous known results by deriving an asymptotic covariance matrix of the jointly estimated system parameters, noise variances and auxiliary parameters. The second paper introduces two formulations of the SISO EIV model using structural equation modeling (SEM). The two formulations allow for quick implementation using standard SEM-based software. The third paper propose a numerically more efficient implementation of the MD estimator for estimating confirmatory factor analysis (CFA) models. The implementation uses an SNLLS approach, which allows part of the parameter vector to be estimated using numerically efficient linear techniques. The fourth and final paper presents a CFA-EIV modeling approach that allows for colored output noise. The presentation extends previous work by including a detailed treatment of the theoretical aspects of the MD estimator. All four papers use simulation examples to illustrate the outlined procedures.

Keywords: System identification, errors-in-variables models, structural equation modeling, confirmatory factor analysis, minimum distance estimator, separable nonlinear least squares

David Kreiberg, Department of Statistics, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-513-1583-6
URN urn:nbn:se:uu:diva-482081 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-482081)

Dedication

I would like to dedicate the work presented in this PhD thesis to my mother, Tove Arnkilde, who sadly passed away before the submission of the thesis. My mother was a woman who believed that education and the pursuit of knowledge was a lifelong endeavor. Although lack of maturity prevented me from embracing these beliefs at a young age, her ideas stuck with me and have since inspired me to pursue several academic degrees, including this PhD.

David Kreiberg

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals (in chronological order):

I. Söderström, Torsten, David Kreiberg, and Magnus Mossberg. "Extended accuracy analysis of a covariance matching approach for identifying errors-in-variables systems." *Automatica* 50, no. 10 (2014): 2597–2605.

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

III. Kreiberg, David, Katerina Marcoulides, and Ulf Henning Olsson. "A faster procedure for estimating CFA models applying minimum distance estimators with a fixed weight matrix." *Structural Equation Modeling: A Multidisciplinary Journal* 28, no. 5 (2021): 725–739.

IV. Kreiberg, David. "A confirmatory factor analysis approach for addressing the errors-in-variables problem with colored output noise." Unpublished manuscript (2022).

Reprints were made with permission from the respective publishers.

Contents

1 Introduction and research objectives

System identification is an empirical approach in which mathematical models of dynamic systems are fitted to measured input and output data. The system identification toolkit is a set of modeling procedures developed for estimation problems involving time-dependent data. This thesis focuses on parametric linear time series models for which the input and output processes are corrupted by additive measurement noise. Such models are referred to as errors-in-variables (EIV) models in the system identification literature. It is a well-known fact that standard regression techniques, when applied to EIV models, lead to biased and inconsistent parameter estimation. Over the years, extensive research has been devoted to the problem of consistently estimating EIV models (typically referred to as the "EIV problem").

The primary research objective in this thesis is to propose methods for addressing the EIV problem. Combining the insights of various strands of literature, the purpose is to outline frameworks that allow EIV models to be consistently estimated by applying covariance structure analysis (CSA). Broadly, CSA refers to a number of statistical procedures that analyze structures or patterns in a covariance matrix. Such procedures have proven particularly useful in the social and behavioral sciences, in which problems typically involve latent (i.e., unobserved) processes and measurement errors. The nature of the EIV problem, in which the noise-free input and output processes are latent, makes CSA an interesting proposition for addressing this problem. However, applying CSA to time-dependent data is not without complications. A well-known problem, addressed later in the thesis, is how to account for the dynamics before the time of the first observation. Solutions to this problem involve introducing latent auxiliary processes into the specification.

CSA-based estimation works by choosing the parameters that minimize the distance between the covariance structure observed in the data and the covariance structure implied by the model. Minimum distance (MD) estimators result from expressing the distance as a squared difference. Applying the MD criteria in the context of CSA can be viewed as a generalized method of moments (GMM) estimation problem. Over the years, GMM has become increasingly popular in econometric analysis. Extensive research on GMM has led to many general results regarding how to address estimation problems involving time-dependent data. The work presented here seeks to use insights from the econometrics literature. Doing so allows for a more in-depth treatment of the theoretical aspects of CSA-based estimation applying the MD criteria for addressing the EIV problem.

When performing CSA, the standard approach to model estimation is to apply computationally demanding nonlinear optimization techniques. In the context of CSA, only limited attention in the literature has been paid to improving the numerical implementation of estimators. A secondary research objective is to propose a numerically more efficient implementation of the MD estimator. Insights from the engineering science literature suggest that for a certain type of nonlinear estimation problems, reformulating the objective function allows part of the parameter vector to be estimated using computationally less demanding linear techniques.

The following section provides a more detailed presentation of the background to the objectives outlined above, including further clues about the content of the constituent studies of this thesis.

2 The EIV problem

2.1 Model formulation

The single-input single-output (SISO) EIV model consists of the following three equations:

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

$$u(t) = u_0(t) + \tilde{u}(t), \tag{2}$$

$$y(t) = y_0(t) + \tilde{y}(t). \tag{3}$$

In these equations, t represents the time unit and q^{-1} denotes the backshift operator, which performs the operation $q^{-m}x(t) = x(t-m)$ for some integer number m. Moreover, $u_0(t)$ and $y_0(t)$ are the unobserved noise-free input and output processes, respectively. Correspondently, $\tilde{u}(t)$ and $\tilde{y}(t)$ are the input and output measurement noise processes, respectively, and $u(t)$ and $y(t)$ are the observed noise-corrupted input and output processes, respectively. The system polynomials take the form

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}, \tag{4}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}. \tag{5}$$

The parameter vector containing the polynomial coefficients in (4) and (5) is

$$\boldsymbol{\vartheta} = \begin{pmatrix} a_1 & \dots & a_{n_a} & b_1 & \dots & b_{n_b} \end{pmatrix}. \tag{6}$$

The EIV model is further characterized by the covariance structure associated with the noise processes $\tilde{u}(t)$ and $\tilde{y}(t)$.

The following assumption outlines the characteristics of the EIV system and its components:

Assumption 1 (processes and model characteristics)

i. All processes are zero-mean ergodic and stationary.

ii. The polynomials $A(q^{-1})$ and $B(q^{-1})$ are coprime (i.e., the polynomials have no common roots) and their respective degrees, as given by n_a and n_b, are known.

iii. The noise-free input process $u_0(t)$ is unknown as are its second-order properties (i.e., $\phi_{u_0}(\omega)$ is unobserved).

iv. The noise processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are both mutually uncorrelated with $u_0(t)$, and mutually uncorrelated with each other.

v. Finally, the noise process $\tilde{u}(t)$ is white, whereas the noise process $\tilde{y}(t)$ may or may not be colored.

2.2 The problem of consistently estimating the EIV model

Based on the available processes $u(t)$ and $y(t)$, the problem is to consistently estimate the true parameter vector $\boldsymbol{\vartheta}_0$. To illustrate the EIV problem further, it will be useful to introduce the following notation:

$$\boldsymbol{v}(t) = \left(-y(t-1) \ \dots \ -y(t-n_a) \ u(t-1) \ \dots \ u(t-n_b)\right)^T, \tag{7}$$

$$\boldsymbol{v}_0(t) = \left(-y_0(t-1) \ \dots \ -y_0(t-n_a) \ u_0(t-1) \ \dots \ u_0(t-n_b)\right)^T, \tag{8}$$

$$\tilde{\boldsymbol{v}}(t) = \left(-\tilde{y}(t-1) \ \dots \ -\tilde{y}(t-n_a) \ \tilde{u}(t-1) \ \dots \ \tilde{u}(t-n_b)\right)^T. \tag{9}$$

The vectors in (7)–(9) all have $p_v = n_a + n_b$ entries. The associated $p_v \times p_v$ covariance matrices are

$$\begin{aligned} &\boldsymbol{R}_v = E[\boldsymbol{v}(t)\boldsymbol{v}^T(t)], \qquad \boldsymbol{R}_{v_0} = E[\boldsymbol{v}_0(t)\boldsymbol{v}_0^T(t)], \\ &\boldsymbol{R}_{\tilde{v}} = E[\tilde{\boldsymbol{v}}(t)\tilde{\boldsymbol{v}}^T(t)], \end{aligned} \tag{10}$$

where E is the expectation operator and the superscript T is the transpose of a vector or a matrix. From Assumption 1(iv), we have

$$\boldsymbol{R}_v = \boldsymbol{R}_{v_0} + \boldsymbol{R}_{\tilde{v}}. \tag{11}$$

The covariance functions associated with the noise processes are written as

$$r_{\tilde{u}}(\tau) = E[\tilde{u}(t+\tau)\tilde{u}(t)], \qquad r_{\tilde{y}}(\tau) = E[\tilde{y}(t+\tau)\tilde{y}(t)],$$
$$\text{for} = 0, \pm 1, \pm 2, \dots . \tag{12}$$

Due to Assumption 1(v), $r_{\tilde{u}}(\tau) = 0$ for all $\tau \neq 0$.

For the sake of simplicity, let the noise processes be white. It then follows that $\boldsymbol{R}_{\tilde{v}}$ is a diagonal matrix given by

$$\boldsymbol{R}_{\tilde{v}} = \begin{pmatrix} r_{\tilde{y}}(0)\boldsymbol{I}_{n_a} & \boldsymbol{0} \\ \boldsymbol{0} & r_{\tilde{u}}(0)\boldsymbol{I}_{n_b} \end{pmatrix}, \tag{13}$$

where $\boldsymbol{I}_{n_a}$ and $\boldsymbol{I}_{n_b}$ are identity matrices with dimensions n_a and n_b, respectively. Now, suppose that we attempt to estimate the EIV model applying least squares (LS) to the noise-corrupted input and output processes. Consider the following linear regression model derived from the EIV equations in (1)–(3):

$$y(t) = \boldsymbol{v}^T(t)\boldsymbol{\vartheta} + e(t), \tag{14}$$

where the noise process $e(t)$ takes the form

$$e(t) = A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \tag{15}$$

In violation of the standard LS assumptions, it is clear from (14) and (15) that $e(t)$ is correlated with the processes of $\boldsymbol{v}(t)$. Consequently, estimates obtained from applying LS to the noise-corrupted processes are biased and inconsistent. The following computation shows the inconsistency of LS:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_{LS} \xrightarrow{p} \boldsymbol{R}_v^{-1}\boldsymbol{r}_{vy} &= \boldsymbol{R}_v^{-1}\boldsymbol{R}_{v_0}\boldsymbol{\vartheta}_0 \\ &= \left(\boldsymbol{R}_{v_0} + \boldsymbol{R}_{\tilde{v}}\right)^{-1}\boldsymbol{R}_{v_0}\boldsymbol{\vartheta}_0 \\ &\neq \boldsymbol{\vartheta}_0, \end{aligned} \tag{16}$$

where "$\xrightarrow{p}$" denotes convergence in probability and $\boldsymbol{r}_{vy} = E[\boldsymbol{v}(t)y(t)]$. On the right-hand side of (16), we have used $\boldsymbol{r}_{vy} = \boldsymbol{R}_{v_0}\boldsymbol{\vartheta}_0$, which follows from the fact that $\tilde{u}(t)$ and $\tilde{y}(t)$ are white and mutually uncorrelated. This clearly shows that consistent estimation of $\boldsymbol{\vartheta}_0$ is a problem of somewhat greater difficulty.

If an estimate $\hat{\boldsymbol{R}}_{\tilde{v}}$ can be obtained, an unbiased and consistent estimate of $\boldsymbol{\vartheta}_0$ is computed by compensating for the bias arising from applying LS. The resulting estimator is known as bias compensating least squares (BCLS), and is given by

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}_{BCLS} &= \left(\hat{\boldsymbol{R}}_v - \hat{\boldsymbol{R}}_{\tilde{v}}\right)^{-1}\hat{\boldsymbol{r}}_{vy} \\ &= \hat{\boldsymbol{R}}_{v_0}^{-1}\hat{\boldsymbol{r}}_{vy}.\end{aligned} \tag{17}$$

Over the years, numerous estimation procedures for addressing the EIV problem have been presented in the literature. Söderström (2007, 2012) provides an overview of different procedures. Note that due to Assumption 1(iii), maximum likelihood (ML) and prediction error methods (PEMs) are not applicable. Recent developments have focused on applying CSA to the EIV problem. Consistent estimation of EIV models using CSA is the main theme of this thesis.

3 Applying CSA to address the EIV problem

Below, we introduce two related CSA approaches for addressing the EIV problem. The two approaches are known as covariance matching (CM) and structural equation modeling (SEM). The difference between the two is that the CM approach is specific to the EIV problem, whereas SEM is an all-purpose modeling approach that can be formulated to accommodate the EIV problem. Both approaches works by matching the covariance structure observed in the data to the covariance structure implied by the EIV model.

3.1 CM

Except for some minor notational differences, the presentation below follows that of Söderström et al. (2009, 2011).

Start by introducing the auxiliary process

$$z_0(t) = \frac{1}{A(q^{-1})} u_0(t). \tag{18}$$

From (18), it is possible to express the observed processes using

$$u(t) = A(q^{-1}) z_0(t) + \tilde{u}(t), \tag{19}$$

$$y(t) = B(q^{-1}) z_0(t) + \tilde{y}(t). \tag{20}$$

Beyond what is stated in Assumption 1, make the assumption that both $\tilde{u}(t)$ and $\tilde{y}(t)$ are white. The covariance and cross-covariance functions of $u(t)$ and $y(t)$ are now written in terms of

$$r_{z_0}(\tau) = E[z_0(t+\tau) z_0(t)], \tag{21}$$

and the model parameters using

$$
\begin{aligned}
r_u(\tau) &= E[u(t+\tau)u(t)] \\
&= \sum_{i=0}^{n_a}\sum_{j=0}^{n_a} a_i a_j r_{z_0}(\tau - i + j)\,, \quad \text{for } \tau > 0 \text{ and } a_0 = 1,
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
r_y(\tau) &= E[y(t+\tau)y(t)] \\
&= \sum_{i=1}^{n_b}\sum_{j=1}^{n_b} b_i b_j r_{z_0}(\tau - i + j)\,, \quad \text{for } \tau > 0,
\end{aligned} \tag{23}
$$

$$
\begin{aligned}
r_{yu}(\tau) &= E[y(t+\tau)u(t)] \\
&= \sum_{i=1}^{n_b}\sum_{j=0}^{n_a} b_i a_j r_{z_0}(\tau - i + j).
\end{aligned} \tag{24}
$$

Define the vectors

$$
\boldsymbol{r}_y \triangleq \begin{pmatrix} r_y(1) \\ \vdots \\ r_y(p_y) \end{pmatrix}, \qquad \boldsymbol{r}_u \triangleq \begin{pmatrix} r_u(1) \\ \vdots \\ r_u(p_u) \end{pmatrix}, \qquad \boldsymbol{r}_{yu} \triangleq \begin{pmatrix} r_{yu}(p_1) \\ \vdots \\ r_{yu}(p_2) \end{pmatrix}, \tag{25}
$$

and

$$
\boldsymbol{r}_{z_0} \triangleq \begin{pmatrix} r_{z_0}(0) \\ \vdots \\ r_{z_0}(k) \end{pmatrix}. \tag{26}
$$

The integers $p_y, p_u \geq 1$ and $p_1 \leq 0 \leq p_2$ are user-chosen quantities that determine the scope of the implementation. The integer k is determined by

$$
k = \max\big(p_y + n_b - 1,\ p_u + n_a,\ \max(-p_1 + n_b,\ p_2 + n_a - 1)\big). \tag{27}
$$

Stacking the vectors in (25) gives

$$
\boldsymbol{r} \triangleq \begin{pmatrix} \boldsymbol{r}_y \\ \boldsymbol{r}_u \\ \boldsymbol{r}_{yu} \end{pmatrix}. \tag{28}
$$

Based on the right-hand side of (22)–(24), it is possible to express $\boldsymbol{r}$ in terms of the covariance vector $\boldsymbol{r}_{z_0}$ and the model parameters using

$$\boldsymbol{r}(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0}) = \boldsymbol{F}(\boldsymbol{\vartheta})\boldsymbol{r}_{z_0}, \tag{29}$$

where $\boldsymbol{F}(\boldsymbol{\vartheta})$ is a tall matrix of full column rank[1]; see Söderström et al. (2009) for the full derivation of $\boldsymbol{F}(\boldsymbol{\vartheta})$.

3.1.1 Estimation

Suppose that a sample of N data points on $u(t)$ and $y(t)$ is available. Based on the data, an estimate, denoted $\hat{\boldsymbol{r}}$, of the true covariance vector can be obtained. CM-based estimation works by choosing the set of parameters that minimizes the distance between $\hat{\boldsymbol{r}}$ and $\boldsymbol{r}(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0})$. Formally, the minimization problem is stated as

$$\{\hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{r}}_{z_0}\} = \underset{\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0}}{\arg\min}\; V(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0}). \tag{30}$$

In this expression, $V(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0})$ is a scalar function expressing the distance between $\hat{\boldsymbol{r}}$ and $\boldsymbol{r}(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0})$. Based on Golub and Pereyra (1973, 2003), the implementation of the estimator uses a separable nonlinear least squares (SNLLS) approach. Later sections provide a more detailed treatment of the work of Golub and Pereyra (1973). The key to applying SNLLS is the separation of parameters in (29), in which $\boldsymbol{r}(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0})$ is written as a product of $\boldsymbol{F}(\boldsymbol{\vartheta})$ and $\boldsymbol{r}_{z_0}$. Let the objective function be given by

$$V(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0}) = \left\|\hat{\boldsymbol{r}} - \boldsymbol{F}(\boldsymbol{\vartheta})\boldsymbol{r}_{z_0}\right\|_{\boldsymbol{V}}^2, \tag{31}$$

where $\|.\|$ is the Euclidean norm and $\boldsymbol{V}$ is a symmetric positive definite weighting matrix. For a given $\boldsymbol{\vartheta}$, minimizing $V(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0})$ w.r.t. $\boldsymbol{r}_{z_0}$ is an LS problem with the straightforward solution

$$\hat{\boldsymbol{r}}_{z_0} = \left(\boldsymbol{F}^T(\boldsymbol{\vartheta})\boldsymbol{V}\boldsymbol{F}(\boldsymbol{\vartheta})\right)^{-1}\boldsymbol{F}^T(\boldsymbol{\vartheta})\boldsymbol{V}\hat{\boldsymbol{r}}. \tag{32}$$

Inserting the right-hand side of (32) into (31), it can be shown that

[1] Recall that a tall matrix is a matrix with more rows than columns.

$$V(\boldsymbol{\vartheta}) = \hat{\boldsymbol{r}}^T \boldsymbol{V} \hat{\boldsymbol{r}} - \hat{\boldsymbol{r}}^T \boldsymbol{V} \boldsymbol{F}(\boldsymbol{\vartheta}) \left(\boldsymbol{F}^T(\boldsymbol{\vartheta}) \boldsymbol{V} \boldsymbol{F}(\boldsymbol{\vartheta}) \right)^{-1} \boldsymbol{F}^T(\boldsymbol{\vartheta}) \boldsymbol{V} \hat{\boldsymbol{r}}. \tag{33}$$

It is clear from this expression that the objective function is now entirely written as a function of $\boldsymbol{\vartheta}$, and it follows that the estimation problem simplifies to

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

Remark 1

- For the estimation problem to be feasible, it is necessary that p_y, p_u, p_1, and p_2 be chosen so that the number of equations represented by (29) is at least as large as the number of parameters to be estimated. This condition is referred to as the order condition in the econometrics literature. Specifically, applying the CM approach to the EIV problem, the order condition is that

$$p_y + p_u + (-p_1 + p_2 + 1) \geq n_a + n_b + k + 1. \tag{35}$$

It is of interest to study the precision of the estimated parameters. Söderström and Mossberg (2011) derived the following expression for the asymptotic covariance matrix:

$$\begin{aligned} \boldsymbol{C}_{CM} &= \lim_{N \to \infty} N Cov(\hat{\boldsymbol{\vartheta}}) \\ &= (\boldsymbol{G}^T \boldsymbol{P} \boldsymbol{G})^{-1} \boldsymbol{G}^T \boldsymbol{P} \boldsymbol{\Omega} \boldsymbol{P} \boldsymbol{G} (\boldsymbol{G}^T \boldsymbol{P} \boldsymbol{G})^{-1}, \end{aligned} \tag{36}$$

where

$$\boldsymbol{G} = \frac{\partial \boldsymbol{r}(\boldsymbol{\vartheta}, \boldsymbol{r}_{z_0,0})}{\partial \boldsymbol{\vartheta}^T}, \tag{37}$$

$$\boldsymbol{P} = \boldsymbol{V} - \boldsymbol{V} \boldsymbol{F}(\boldsymbol{\vartheta}) \left(\boldsymbol{F}^T(\boldsymbol{\vartheta}) \boldsymbol{V} \boldsymbol{F}(\boldsymbol{\vartheta}) \right)^{-1} \boldsymbol{F}^T(\boldsymbol{\vartheta}) \boldsymbol{V}, \tag{38}$$

$$\boldsymbol{\Omega} = \lim_{N \to \infty} N\, E[(\hat{\boldsymbol{r}} - \boldsymbol{r})(\hat{\boldsymbol{r}} - \boldsymbol{r})^T]. \tag{39}$$

In (37), $\boldsymbol{r}_{z_0,0}$ denotes the true value of $\boldsymbol{r}_{z_0}$. As shown by Söderström and Mossberg (2011), $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$ is the optimal weighting matrix, in the sense that

$$\boldsymbol{C}_{CM}|_{\boldsymbol{V}} \geq \boldsymbol{C}_{CM}|_{\boldsymbol{V} = \boldsymbol{\Omega}^{-1}}. \tag{40}$$

Although CM has been applied to a number of problems in the system identification literature, it is still a novel approach to addressing the EIV problem. The first study in this thesis extends the CM approach by allowing $p_y, p_u \geq 0$. The study additionally investigates the precision of the estimated parameters for different $\boldsymbol{V}$.

3.2 SEM

SEM is a multivariate modeling approach that combines factor analysis and multiple regression analysis to analyze structural relationships between measured and latent processes. Over the years, SEM has become increasingly popular in the social and behavioral sciences. The major reason for this popularity stems from the versatility of the SEM framework.

3.2.1 Notation

Before presenting the SEM equations, it will be useful to introduce the following (somewhat customized) notation. Let $\boldsymbol{x}(t)$ be a $p_x \times 1$ zero-mean vector process, and let $\boldsymbol{R}_x$ be the associated $p_x \times p_x$ covariance matrix given by

$$\boldsymbol{R}_x = E[\boldsymbol{x}(t)\boldsymbol{x}^T(t)]. \tag{41}$$

The number of non-redundant elements of $\boldsymbol{R}_x$ is $q = 2^{-1}p_x(p_x + 1)$, given that no restrictions other than symmetry are placed on the elements of $\boldsymbol{R}_x$.[2] The $q \times 1$ covariance vector containing these elements is

$$\boldsymbol{r}_x = vech(\boldsymbol{R}_x), \tag{42}$$

where $vech$ is the operation of vectorising the non-redundant elements of $\boldsymbol{R}_x$. The vectorization in (42) can be accomplished by the following computation:

$$\boldsymbol{r}_x = \boldsymbol{K}_x^T vec(\boldsymbol{R}_x). \tag{43}$$

In this expression, vec is the operation of vectorising the elements of a matrix by stacking its columns, and $\boldsymbol{K}_x$ is a $p_x^2 \times q$ matrix obtained from

$$\boldsymbol{K}_x = \boldsymbol{L}_x(\boldsymbol{L}_x^T\boldsymbol{L}_x)^{-1}, \tag{44}$$

[2] The non-redundant elements of a symmetric matrix are the elements on and below the diagonal of the matrix.

where $\boldsymbol{L}_x$ is a $p_x^2 \times q$ selection matrix.[3] In the SEM literature, $\boldsymbol{L}_x$ is typically referred to as the duplication matrix.

Expanding on the previous notation, let $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ be $p_{x_1} \times 1$ and $p_{x_2} \times 1$ zero-mean vector processes, respectively. A $p_x = p_{x_1} + p_{x_2}$ vector is obtained by

$$\boldsymbol{x}(t) = \begin{pmatrix} \boldsymbol{x}_1(t) \\ \boldsymbol{x}_2(t) \end{pmatrix}. \tag{45}$$

The associated block covariance matrix is

$$\boldsymbol{R}_x = \begin{pmatrix} \boldsymbol{R}_{x_1} & \boldsymbol{R}_{x_2,x_1}^T \\ \boldsymbol{R}_{x_2,x_1} & \boldsymbol{R}_{x_2} \end{pmatrix}, \tag{46}$$

where the individual blocks are given by

$$\begin{aligned} &\boldsymbol{R}_{x_1} = E[\boldsymbol{x}_1(t)\boldsymbol{x}_1^T(t)], \qquad \boldsymbol{R}_{x_2} = E[\boldsymbol{x}_2(t)\boldsymbol{x}_2^T(t)], \\ &\boldsymbol{R}_{x_2,x_1} = E[\boldsymbol{x}_2(t)\boldsymbol{x}_1^T(t)]. \end{aligned} \tag{47}$$

The dimensions of $\boldsymbol{R}_{x_1}$, $\boldsymbol{R}_{x_2}$, and $\boldsymbol{R}_{x_2,x_1}$ are $p_{x_1} \times p_{x_1}$, $p_{x_2} \times p_{x_2}$, and $p_{x_2} \times p_{x_1}$, respectively.

3.2.2 Model formulation

The SEM framework consists of the following three model equations (excluding constant terms):

$$\boldsymbol{\eta}(t) = \boldsymbol{B}\boldsymbol{\eta}(t) + \boldsymbol{\Gamma}\boldsymbol{\xi}(t) + \boldsymbol{\delta}(t), \tag{48}$$

$$\boldsymbol{x}_1(t) = \boldsymbol{\Lambda}_1\boldsymbol{\eta}(t) + \boldsymbol{\epsilon}_1(t), \tag{49}$$

$$\boldsymbol{x}_2(t) = \boldsymbol{\Lambda}_2\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}_2(t). \tag{50}$$

The first equation is the structural equation, which specifies the causal relationships among the latent processes. In this equation, $\boldsymbol{\eta}(t)$ and $\boldsymbol{\xi}(t)$ are respectively $p_\eta \times 1$ and $p_\xi \times 1$ vectors of latent processes, $\boldsymbol{\delta}(t)$ is a $p_\eta \times 1$ vector of noise processes, and $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are respectively $p_\eta \times p_\eta$ and $p_\eta \times p_\xi$ parameter matrices relating the latent processes. The last two equations are measurement equations specifying the relationships between the latent and observed processes. In these

[3] This is a matrix containing only ones and zeros.

equations, $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ are respectively $p_{x_1} \times 1$ and $p_{x_2} \times 1$ vectors of observed processes, $\boldsymbol{\epsilon}_1(t)$ and $\boldsymbol{\epsilon}_2(t)$ are vectors of measurement noise processes of similar dimensions, and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are respectively $p_{x_1} \times p_\eta$ and $p_{x_2} \times p_\xi$ parameter matrices relating the latent and observed processes. All processes are zero-mean.

The model characteristics are:

- The matrix $\boldsymbol{H} = \boldsymbol{I}_{p_\eta} - \boldsymbol{B}$ is non-singular, such that $\boldsymbol{\eta}(t)$ is uniquely determined by $\boldsymbol{\xi}(t)$ and $\boldsymbol{\delta}(t)$.
- The processes of $\boldsymbol{\delta}(t)$ are mutually uncorrelated with the processes of $\boldsymbol{\xi}(t)$, and the processes of $\boldsymbol{\epsilon}_1(t)$ and $\boldsymbol{\epsilon}_2(t)$ are mutually uncorrelated with the processes of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\xi}(t)$, respectively.
- The processes of $\boldsymbol{\epsilon}_1(t)$ and $\boldsymbol{\epsilon}_2(t)$ may or may not correlate.[4]

The specification additionally includes the following covariance matrices:

$$
\begin{aligned}
&\boldsymbol{R}_\xi = E[\boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t)], \qquad \boldsymbol{R}_\delta = E[\boldsymbol{\delta}(t)\boldsymbol{\delta}^T(t)], \\
&\boldsymbol{R}_{\epsilon_1} = E[\boldsymbol{\epsilon}_1(t)\boldsymbol{\epsilon}_1^T(t)], \quad \boldsymbol{R}_{\epsilon_2} = E[\boldsymbol{\epsilon}_2(t)\boldsymbol{\epsilon}_2^T(t)], \\
&\boldsymbol{R}_{\epsilon_2,\epsilon_1} = E[\boldsymbol{\epsilon}_2(t)\boldsymbol{\epsilon}_1^T(t)].
\end{aligned}
\tag{51}
$$

The dimensions of $\boldsymbol{R}_\xi$, $\boldsymbol{R}_\delta$, $\boldsymbol{R}_{\epsilon_1}$, $\boldsymbol{R}_{\epsilon_2}$, and $\boldsymbol{R}_{\epsilon_2,\epsilon_1}$ are $p_\xi \times p_\xi$, $p_\eta \times p_\eta$, $p_{x_1} \times p_{x_1}$, $p_{x_2} \times p_{x_2}$, and $p_{x_2} \times p_{x_1}$, respectively.

The parameters contained in $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{R}_\xi$, $\boldsymbol{R}_\delta$, $\boldsymbol{R}_{\epsilon_1}$, $\boldsymbol{R}_{\epsilon_2}$, and $\boldsymbol{R}_{\epsilon_2,\epsilon_1}$ are either free or constrained. A parameter is said to be free if it is unknown and has to be estimated. A parameter is said to be constrained if it is assigned a specific value or if it is a function (linear or nonlinear) of other parameters. Models are formed by appropriately constraining the parameters. An important application of SEM is to evaluate the appropriateness of the model. This is done by jointly testing the validity of the imposed constraints that form the model.

Let $\boldsymbol{\vartheta}$ be a vector containing the free parameters. The covariance matrix implied by (48)–(50) is

[4] The covariance matrix $\boldsymbol{R}_{\epsilon_2,\epsilon_1}$ is typically omitted from the specification.

$$\boldsymbol{R}_x(\boldsymbol{\vartheta}) = \begin{pmatrix} \boldsymbol{R}_{x_1}(\boldsymbol{\vartheta}) & \boldsymbol{R}^T_{x_2,x_1}(\boldsymbol{\vartheta}) \\ \boldsymbol{R}_{x_2,x_1}(\boldsymbol{\vartheta}) & \boldsymbol{R}_{x_2}(\boldsymbol{\vartheta}) \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{\Lambda}_1\boldsymbol{H}^{-1}\left(\boldsymbol{\Gamma}\boldsymbol{R}_\xi\boldsymbol{\Gamma}^T + \boldsymbol{R}_\delta\right)\boldsymbol{H}^{-T}\boldsymbol{\Lambda}_1^T + \boldsymbol{R}_{\epsilon_1} & \boldsymbol{\Lambda}_1\boldsymbol{H}^{-1}\boldsymbol{\Gamma}\boldsymbol{R}_\xi\boldsymbol{\Lambda}_2^T + \boldsymbol{R}^T_{\epsilon_2,\epsilon_1} \\ \boldsymbol{\Lambda}_2\boldsymbol{R}_\xi\boldsymbol{\Gamma}^T\boldsymbol{H}^{-T}\boldsymbol{\Lambda}_1^T + \boldsymbol{R}_{\epsilon_2,\epsilon_1} & \boldsymbol{\Lambda}_2\boldsymbol{R}_\xi\boldsymbol{\Lambda}_2^T + \boldsymbol{R}_{\epsilon_2} \end{pmatrix}. \tag{52}$$

3.2.3 Confirmatory factor analysis (CFA)

The measurement equations in (49) and (50) are modeling devices in their own right. A CFA model takes the form

$$\boldsymbol{x}(t) = \boldsymbol{\Lambda}\boldsymbol{\xi}(t) + \boldsymbol{\epsilon}(t). \tag{53}$$

In this model, $\boldsymbol{x}(t)$ is $p_x \times 1$ vector of observed processes, $\boldsymbol{\xi}(t)$ is a $p_\xi \times 1$ vector of latent processes and $\boldsymbol{\epsilon}(t)$ is a $p_x \times 1$ vector of measurement noise processes. The $p_x \times p_\xi$ matrix $\boldsymbol{\Lambda}$ contains the parameters that relate the latent and observed processes. As previously, all processes are zero-mean. The processes of $\boldsymbol{\epsilon}(t)$ are assumed to be mutually uncorrelated with the processes of $\boldsymbol{\xi}(t)$, but are allowed to correlate among themselves.

The specification additionally includes the following covariance matrices:

$$\boldsymbol{R}_\xi = E[\boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t)], \qquad \boldsymbol{R}_\epsilon = E[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}^T(t)], \tag{54}$$

where the dimensions of $\boldsymbol{R}_\xi$ and $\boldsymbol{R}_\epsilon$ are $p_\xi \times p_\xi$ and $p_x \times p_x$, respectively. The parameter vector $\boldsymbol{\vartheta}$ is composed of the free elements of $\boldsymbol{\Lambda}$, $\boldsymbol{R}_\xi$, and $\boldsymbol{R}_\epsilon$. The model-implied covariance matrix now takes the simpler form

$$\boldsymbol{R}_x(\boldsymbol{\vartheta}) = \boldsymbol{\Lambda}\boldsymbol{R}_\xi\boldsymbol{\Lambda}^T + \boldsymbol{R}_\epsilon. \tag{55}$$

3.2.4 Estimation

Suppose that a sample of N data points on $\boldsymbol{x}(t)$ is available. Based on the data, an estimate of $\boldsymbol{R}_x$ is obtained by

$$\widehat{\boldsymbol{R}}_x = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^T(t). \tag{56}$$

Given $\widehat{\boldsymbol{R}}_x$, the aim is to estimate the true parameter vector $\boldsymbol{\vartheta}_0$. This is accomplished by

$$\widehat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} V(\boldsymbol{\vartheta}), \tag{57}$$

where $V(\boldsymbol{\vartheta})$ is a scalar function expressing the distance between the observed covariance structure and the model-implied covariance structure. There are various formulations of $V(\boldsymbol{\vartheta})$ depending on the estimator. Below, we focus on the MD estimator with the objective function

$$V(\boldsymbol{\vartheta}) = \left(\hat{\boldsymbol{r}}_x - \boldsymbol{r}_x(\boldsymbol{\vartheta})\right)^T \widehat{\boldsymbol{V}}\left(\hat{\boldsymbol{r}}_x - \boldsymbol{r}_x(\boldsymbol{\vartheta})\right). \tag{58}$$

In this expression, $\widehat{\boldsymbol{V}}$ is a $q \times q$ positive definite weighting matrix and the vectors $\hat{\boldsymbol{r}}_x$ and $\boldsymbol{r}_x(\boldsymbol{\vartheta})$ are

$$\hat{\boldsymbol{r}}_x = vech\left(\widehat{\boldsymbol{R}}_x\right), \qquad \boldsymbol{r}_x(\boldsymbol{\vartheta}) = vech\left(\boldsymbol{R}_x(\boldsymbol{\vartheta})\right). \tag{59}$$

Remark 2

- For the estimation problem to be feasible, it is necessary that the number of matching elements between $\hat{\boldsymbol{r}}_x$ and $\boldsymbol{r}_x(\boldsymbol{\vartheta})$ be at least as large as the number of parameters to be estimated (i.e., the order condition must be satisfied).
- General conditions for consistency are that $V(\boldsymbol{\vartheta})$ converge uniformly in probability to a deterministic function that is uniquely minimized at $\boldsymbol{\vartheta}_0$. Given that the general conditions are satisfied, $\widehat{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}_0$.
- Different choices of $\widehat{\boldsymbol{V}}$ lead to different estimators. Consistency does not depend on $\widehat{\boldsymbol{V}}$ as long as $\widehat{\boldsymbol{V}}$ converges in probability to a symmetric positive definite matrix $\boldsymbol{V}$. However, as shown below, the asymptotic precision of $\widehat{\boldsymbol{\vartheta}}$ depends on the weighting matrix.

The asymptotic covariance matrix of $\widehat{\boldsymbol{\vartheta}}$ is computed using

$$\boldsymbol{C}_{SEM} = (\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{\Omega}\boldsymbol{V}\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{G})^{-1}, \tag{60}$$

where

$$\boldsymbol{G} = E\left[\frac{\partial \boldsymbol{r}_x(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}^T}\right], \tag{61}$$

$$\boldsymbol{\Omega} = E\left[\left(\hat{\boldsymbol{r}}_x - \boldsymbol{r}_x(\boldsymbol{\vartheta}_0)\right)\left(\hat{\boldsymbol{r}}_x - \boldsymbol{r}_x(\boldsymbol{\vartheta}_0)\right)^T\right]. \tag{62}$$

It can be shown that $\boldsymbol{C}_{SEM}$ reaches its lower bound when $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$, which is formally stated as

$$\boldsymbol{C}_{SEM}|_{\boldsymbol{V}} \geq \boldsymbol{C}_{SEM}|_{\boldsymbol{V}=\boldsymbol{\Omega}^{-1}}. \tag{63}$$

3.2.5 Time series analysis in SEM

Although SEM is typically applied to static problems, numerous studies have demonstrated how to apply SEM to dynamic problems. For instance, Buuren (1997) showed how to implement autoregressive moving-average (ARMA) processes in SEM. Lyhagen (2005) outlined a SEM-based framework for analyzing stationary multivariate time series. Browne and Nesselroade (2005) and Browne and Zhang (2007) introduced the dynamic factor analysis (DFA) model, while du Toit and Browne (2007) extended the ARMA implementation to the multivariate case (VARMA). Sy-Miin Chow et al. (2010) discussed similarities and differences between SEM and state-space (SS) models. More studies could have been added to the list.

An important obstacle to applying SEM to time-dependent data is how to account for the dynamics before the time of the first observation. To help understand this problem better, it is useful to consider an example.

Example 1

Let $x(t)$ be an ARMA(1,1) process expressed by

$$x(t) - ax(t-1) = e(t) + be(t-1). \tag{64}$$

Given stationarity, the model is expanded into the following system of three equations:

$$\begin{aligned} x(t) &= ax(t-1) + e(t) + be(t-1) \\ x(t-1) &= ax(t-2) + e(t-1) + be(t-2) \\ x(t-2) &= ax(t-3) + e(t-2) + be(t-3). \end{aligned} \tag{65}$$

In this system, the time of the first observation is $t-2$, since the third equation is the first equation lagged two periods. Straightforward manipulation allows us to write the system in matrix form

$$\begin{aligned}\begin{pmatrix} x(t) \\ x(t-1) \\ x(t-2) \end{pmatrix} &= \begin{pmatrix} 0 & a & 0 \\ 0 & 0 & a \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x(t) \\ x(t-1) \\ x(t-2) \end{pmatrix} \\ &+ \begin{pmatrix} 1 & b & 0 & 0 \\ 0 & 1 & b & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} e(t) \\ e(t-1) \\ e(t-2) \\ z(t) \end{pmatrix},\end{aligned} \tag{66}$$

where $z(t) = ax(t-3) + be(t-3)$ is a latent auxiliary process that captures the dynamics before time $t-2$. What is not immediately obvious is that the model in (66) can be fitted into SEM by appropriately constraining the elements of the parameter matrices.

The SEM implementation of the EIV model is subject to the same problem as illustrated in Example 1. That is, when specifying the model, one must account for any dynamics before the time of the first observation. The second study in this thesis contributes to the literature by outlining two different implementations of the EIV model when fitted into SEM. The two implementations are distinguished by how the latent auxiliary processes are formulated.

3.2.6 Asymptotic properties of the MD estimator

When studying the asymptotic properties of the MD estimator of SEMs, it is useful to regard SEM as a special case of GMM. The appeal of the GMM framework is that it can handle a wide range of estimation problems, including semiparametric problems (i.e., problems for which the complete shape of the data distribution is unknown). Hansen (1982) initially introduced GMM. Since then, a vast body of literature has accumulated on how to model time-dependent data using GMM. Most notably is the work of White and Domowitz (1984) and Newey and West (1987). Newey and McFadden (1994) provided a thorough introduction to GMM estimation, including a treatment of GMM asymptotics and hypothesis testing. The presentation below provides a brief overview of the GMM framework.

Let $\boldsymbol{x}(t)$ be a $p_x \times 1$ vector process, and let $\boldsymbol{\rho}(\boldsymbol{x}(t), \boldsymbol{\vartheta})$ be a $q \times 1$ vector-valued function that satisfies

$$E[\boldsymbol{\rho}(\boldsymbol{x}(t), \boldsymbol{\vartheta}_0)] = \mathbf{0}, \tag{67}$$

where, as previously, $\boldsymbol{\vartheta}_0$ is the true parameter vector. The set of equations represented by (67) is referred to as the population moment conditions. Suppose that a set of N data points on $\boldsymbol{x}(t)$ is available. Based on the data, the sample mean of $\boldsymbol{\rho}(\boldsymbol{x}(t), \boldsymbol{\vartheta})$ is computed using

$$\widehat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}) = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\rho}(\boldsymbol{x}(t), \boldsymbol{\vartheta}). \tag{68}$$

Any statistical problem that fits into (67) and (68) can be considered a GMM problem. Estimation is performed using

$$\widehat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \ V(\boldsymbol{\vartheta}), \tag{69}$$

where $V(\boldsymbol{\vartheta})$ is the objective function expressed by the quadratic form

$$V(\boldsymbol{\vartheta}) = \widehat{\boldsymbol{\rho}}^T(\boldsymbol{\vartheta})\widehat{\boldsymbol{V}}\widehat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}). \tag{70}$$

In this expression, $\widehat{\boldsymbol{V}}$ is a $q \times q$ positive definite weighting matrix.

Remark 3

- When analyzing time-dependent data, it is typically assumed that $\boldsymbol{x}(t)$ is stationary and ergodic.
- For the estimation problem to be feasible, it is necessary that the number of elements of $\boldsymbol{\rho}(\boldsymbol{x}(t), \boldsymbol{\vartheta})$ be at least as large as the number of parameters to be estimated (i.e., the order condition must be satisfied).
- General conditions for consistency are that $V(\boldsymbol{\vartheta})$ converge uniformly in probability to a deterministic function that is uniquely minimized at $\boldsymbol{\vartheta}_0$. Given that the general conditions are satisfied, $\widehat{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}_0$.
- Different choices of $\widehat{\boldsymbol{V}}$ lead to different estimators. Consistency does not depend on $\widehat{\boldsymbol{V}}$ as long as $\widehat{\boldsymbol{V}}$ converges in probability to a symmetric positive definite matrix $\boldsymbol{V}$. However, as shown below, the asymptotic precision of $\widehat{\boldsymbol{\vartheta}}$ depends on the weighting matrix.

The asymptotic covariance matrix of $\hat{\boldsymbol{\vartheta}}$ is computed using

$$\boldsymbol{C}_{GMM} = (\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{\Omega}\boldsymbol{V}\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{V}\boldsymbol{G})^{-1}, \tag{71}$$

where

$$\boldsymbol{G} = E\left[\frac{\partial\boldsymbol{\rho}(\boldsymbol{x}(t),\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}^T}\right], \tag{72}$$

$$\boldsymbol{\Omega} = E[\boldsymbol{\rho}(\boldsymbol{x}(t),\boldsymbol{\vartheta}_0)\boldsymbol{\rho}^T(\boldsymbol{x}(t),\boldsymbol{\vartheta}_0)]. \tag{73}$$

It can be shown that $\boldsymbol{C}_{GMM}$ reaches its lower bound when $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$, which is formally stated as

$$\boldsymbol{C}_{GMM}|_{\boldsymbol{V}} \geq \boldsymbol{C}_{GMM}|_{\boldsymbol{V}=\boldsymbol{\Omega}^{-1}}. \tag{74}$$

The connection between SEM and GMM is seen from

$$\boldsymbol{\rho}(\boldsymbol{x}(t),\boldsymbol{\vartheta}) = \boldsymbol{K}_x^T vec\big(\boldsymbol{x}(t)\boldsymbol{x}^T(t) - \boldsymbol{R}_x(\boldsymbol{\vartheta})\big). \tag{75}$$

Inserting the right-hand side of (75) into (68) yields $\hat{\boldsymbol{\rho}}(\boldsymbol{\vartheta}) = \hat{\boldsymbol{r}}_x - \boldsymbol{r}_x(\boldsymbol{\vartheta})$. It follows that SEM-based estimation applying the MD criteria coincides with the GMM estimator.

Many results from the econometrics literature are directly applicable to SEM. The advantage of viewing SEM as a special case of GMM is that we can use the full GMM statistical machinery when studying the properties of the MD estimator. Following the presentation of Newey and McFadden (1994), the fourth and final study in the thesis outlines conditions for consistency and asymptotic normality of the estimator.

3.2.7 A numerically more efficient implementation of the estimator

SEM-based estimation works by searching the parameter space for the set of estimates that minimizes the objective function. The standard implementation of SEM estimators involves minimizing the objective function using nonlinear optimization techniques. Estimation applying the MD criteria is typically referred to as nonlinear least squares (NLLS).

When estimating SEMs, there may be situations in which NLLS is inconvenient due to the computational load, for instance, when the estimation problem involves some form of simulation or resampling scheme. It is therefore worthwhile investigating the possibility of applying SNLLS. The presentation below outlines the idea of SNLLS.

Given data (x_i, y_i), for $i = 1, \dots, N$, Golub and Pereyra (1973) consider nonlinear estimation problems for which the objective function is written as

$$\begin{aligned} V(\mathbf{a}, \boldsymbol{\alpha}) &= \sum_{i=1}^{N} \big(y_i - \varrho(\mathbf{a}, \boldsymbol{\alpha}, x_i)\big)^2 \\ &= \|\boldsymbol{y} - \boldsymbol{\varrho}(\mathbf{a}, \boldsymbol{\alpha})\|^2, \end{aligned} \tag{76}$$

where $\boldsymbol{\varrho}(\mathbf{a}, \boldsymbol{\alpha})$ is a nonlinear function of the disjoint parameter sets $\mathbf{a}$ and $\boldsymbol{\alpha}$. The estimation problem is

$$\{\hat{\mathbf{a}}, \widehat{\boldsymbol{\alpha}}\} = \arg\min_{\mathbf{a}, \boldsymbol{\alpha}} V(\mathbf{a}, \boldsymbol{\alpha}). \tag{77}$$

The idea of parameter separation is to express $\boldsymbol{\varrho}(\mathbf{a}, \boldsymbol{\alpha})$ in a way that allows $\mathbf{a}$ to enter the objective function in a linear fashion

$$V(\mathbf{a}, \boldsymbol{\alpha}) = \|\boldsymbol{y} - \boldsymbol{\Phi}(\boldsymbol{\alpha})\mathbf{a}\|^2. \tag{78}$$

In this expression, $\boldsymbol{\Phi}(\boldsymbol{\alpha})$ is a tall matrix-valued function of full column rank. Given $\boldsymbol{\alpha}$, an LS solution for minimizing (78) is given by

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) = \big(\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{\Phi}(\boldsymbol{\alpha})\big)^{-1}\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{y}. \tag{79}$$

Theorem 2.1 in Golub and Pereyra (1973) provides the justification for replacing $\mathbf{a}$ in (78) with the right-hand side of (79), for which it can be shown that

$$\begin{aligned} V(\boldsymbol{\alpha}) &= \left\|\boldsymbol{y} - \boldsymbol{\Phi}(\boldsymbol{\alpha})\big(\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{\Phi}(\boldsymbol{\alpha})\big)^{-1}\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{y}\right\|^2 \\ &= \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Phi}(\boldsymbol{\alpha})\big(\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{\Phi}(\boldsymbol{\alpha})\big)^{-1}\boldsymbol{\Phi}^T(\boldsymbol{\alpha})\boldsymbol{y}. \end{aligned} \tag{80}$$

It is clear from (80) that the objective function is purely a function of $\boldsymbol{\alpha}$. Thus, minimizing (80) represents a lower dimensional optimization problem that may considerably reduce the computational load. Applying SNLLS involves estimating the complete set of parameters in two steps. In the first step, $\widehat{\boldsymbol{\alpha}}$ is obtained by minimizing (80). In the second step, $\hat{\mathbf{a}}$ is computed by (79) using $\widehat{\boldsymbol{\alpha}}$ from the first step. As shown by Golub and Pereyra (1973), SNLLS yields the exact same estimation results as does NLLS. One major benefit of SNLLS is that fewer starting values are required. Moreover, studies have shown that SNLLS possesses a number of benefits, which include faster convergence and better performance when the estimation problem is ill-conditioned (i.e., problems in which the ratio between the largest and the smallest singular value of the covariance matrix is large); see for instance, Sjöberg and Viberg (1997) and Dattner et al. (2020).

The question is how to adapt SNLLS to accommodate SEM-based MD estimators. The third study in the thesis addresses part of this problem by proposing an SNLLS implementation for estimating CFA models. Although the results are applicable in an EIV context, the content of the third study does not explicitly involve EIV models.

4 Summary of research

4.1 Paper I

The first study presents a more general CM framework for identifying EIV models. The presentation extends previous results of Söderström et al. (2009) and Söderström and Mossberg (2011) by allowing $p_y, p_u \geq 0$ (rather than $p_y, p_u \geq 1$). Based on the extended framework, the study derives the asymptotic covariance matrix of the parameter estimates. Simulation examples are used to investigate how well the empirical variance agrees with the theoretical variance derived from asymptotic theory. Although it is difficult to generalize the results of simulation research, the most important conclusions from the examples are the following:

- In general, there is good agreement between the empirical variance and the theoretical variance. Agreement is slightly more regular when $\boldsymbol{V} = \boldsymbol{I}$ (equal weighting) than when $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$ (optimal weighting).
- Increasing the autocovariance lag length improves parameter precision up to a certain point. When $\boldsymbol{V} = \boldsymbol{I}$, adding lags beyond this point increases the variance of the parameter estimates. When $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$, adding more lags beyond this point marginally decreases the variance of the parameter estimates.
- Applying $\boldsymbol{V} = \boldsymbol{I}$, there exists a choice for the autocovariance lag length that leads to nearly the same parameter precision as does applying $\boldsymbol{V} = \boldsymbol{\Omega}^{-1}$.

4.2 Paper II

The second study seeks to address the EIV problem by applying SEM. The presentation first outlines the general SEM framework along with various procedures for estimating parameters. The presentation goes on to introduce two SEM-based formulations of the EIV model. The first formulation uses a CFA model in which the latent auxiliary process is similar to that used in the CM approach. The second formulation uses the complete SEM. In this formulation, the latent auxiliary processes follow the logic illustrated in Example 1. Specifically targeting the two SEM formulations of the EIV model, the study outlines how to estimate the parameters by applying an SNLLS implementation of the MD estimator.

Simulation examples demonstrate the implementation of the two SEM formulations. For comparison, the examples also include the CM approach. As expected, the results of the simulations show that SEM and CM-based estimation provide parameter estimates of similar quality.

4.3 Paper III

The third study presents an SNLLS implementation of SEM-based MD estimators for estimating CFA models. The study shows how to appropriately modify the objective function to accommodate SNLLS. The proposed implementation is a two-step procedure. In the first step, $\widehat{\boldsymbol{\vartheta}}_{\lambda}$ is obtained by means of nonlinear optimization, where $\widehat{\boldsymbol{\vartheta}}_{\lambda}$ contains the estimates of the true elements of $\boldsymbol{\Lambda}$. In a second step, $\widehat{\boldsymbol{\vartheta}}_{r_{\xi},r_{\epsilon}}$ is computed using numerically efficient LS, where $\widehat{\boldsymbol{\vartheta}}_{r_{\xi},r_{\epsilon}}$ contains the estimates of the true elements of $\boldsymbol{R}_{\xi}$ and $\boldsymbol{R}_{\epsilon}$. It is conjectured that the benefits of SNLLS in terms of faster estimation increase with the ratio $t_{\boldsymbol{\vartheta}_{r_{\xi},r_{\epsilon}}}/t_{\boldsymbol{\vartheta}_{\lambda}}$, where $t_{\boldsymbol{\vartheta}_{r_{\xi},r_{\epsilon}}}$ and $t_{\boldsymbol{\vartheta}_{\lambda}}$ are the numbers of elements of $\boldsymbol{\vartheta}_{r_{\xi},r_{\epsilon}}$ and $\boldsymbol{\vartheta}_{\lambda}$, respectively.

Four examples compare the numerical efficiency across the two implementations, SNLLS and NLLS. Numerical efficiency is here measured as the average (or median) estimation time over a large number of runs. Three of the four examples use well-known models from the literature. The fourth example is a simulation experiment in which the estimation time is evaluated for a range of models that systematically vary in size and in the ratio $t_{\boldsymbol{\vartheta}_{r_{\xi},r_{\epsilon}}}/t_{\boldsymbol{\vartheta}_{\lambda}}$. The main conclusions from the examples are the following:

- Estimation using SNLLS is faster than estimation using NLLS, and in some cases by a considerable margin.
- As conjectured, the benefit of applying SNLLS in terms of reduced estimation time increases with the ratio $t_{\boldsymbol{\vartheta}_{r_{\xi},r_{\epsilon}}}/t_{\boldsymbol{\vartheta}_{\lambda}}$

4.4 Paper IV

The fourth study aims at further refining the CFA approach for addressing the EIV problem. The study contributes to the literature by providing a more comprehensive description that seeks to deal with the theoretical aspects of SEM-based estimation applying the MD criteria. Specifically, the presentation extends the CFA formulation by allowing the output noise process to be colored. Based on GMM theory, the presentation outlines conditions for consistency and asymptotic normality. Finally, the SNLLS procedure is adapted to accommodate colored output noise.

A simulation example demonstrates the implementation of the outlined framework. To simplify matters, estimation is performed applying $\boldsymbol{V} = \boldsymbol{I}$. The example includes investigating the effect of the autocovariance lag length. The following items summarize the results of the simulation example:

- The empirical mean of the parameter estimates closely approximate the true parameter values.
- Generally, there is close agreement between the empirical standard errors of the parameter estimates and the theoretical standard errors derived from asymptotic theory.

Final comments offer users some practical guidance on how to prevent numerical problems when implementing the estimator.

5 Conclusion and future research

The main research objective in this thesis has been to outline procedures for addressing the EIV problem by applying CSA. The overall conclusion from the work is that CSA is a viable option for addressing this problem. Although the results so far represent important steps, more work is necessary to make CSA fully workable. A secondary research objective has been to suggest a numerically more efficient estimation of SEMs. This objective was accomplished in part by proposing an SNLLS implementation of the MD estimator of CFA models. The results of this work suggest that a considerable reduction in estimation time can be realized by applying SNLLS. One major limitation is that SNLLS is not yet applicable to the complete SEM.

The presentation below proposes several research problems related to the work in this thesis.

5.1 Obtaining an empirical optimal weight matrix

When applying CSA to the EIV problem, one important question is how to obtain an empirical optimal weight matrix. Mossberg and Söderström (2011) suggested an iterative procedure for estimating the weight matrix based on the model parameters. The procedure works by updating the weight matrix at every iteration using the estimated parameters from the previous iteration. Although appealing, an important drawback is the lack of stability. There is no mechanism to insure that the resulting weight matrix is positive definite. Another drawback is the lack of generality. The procedure is only applicable in the case of white output noise.

Insight from the econometrics literature may help address these issues. For instance, Newey and West (1987) suggested a heteroscedasticity and autocorrelation consistent (HAC) estimator of the weight matrix based on the Bartlett kernel. Another example is Den Haan and Levin (1998) who introduced the vector autoregressive HAC (VARHAC) estimator. This estimator involves obtaining the weight matrix from estimating a finite-order vector autoregression (VAR). Both these procedures provide a weight matrix that satisfies the definiteness condition and are applicable for a wide range of model formulations.

5.2 Generalizing the SNLLS implementation

A natural next step is to extend the SNLLS implementation to allow estimation of the complete SEM. For such an extension, parameter separation is more involved due to the block form of the model-implied covariance matrix. This work is well on its way, and a study presenting the SNLLS implementation of the MD estimator of SEMs is about to be published.

The SNLLS implementation of the MD estimator applies a fixed weighting matrix (i.e., the weight matrix is not a function of the free parameters), which initially appears to rule out ML estimation. The question is whether it is possible to rework SNLLS to facilitate ML estimation of SEMs, while retaining its benefits.

Acknowledgements

The path to finishing this PhD turned out to be somewhat of a “rocky road”. A major challenge in preparing this thesis has been working across different disciplines, such as informatics, econometrics, and psychometrics. To what extent I have succeeded in meeting this challenge is for others to decide. Fortunately, I had good help from three tremendously talented academics at Uppsala University. I wish to express my sincerest gratitude to Professor Torsten Söderström at the Department of Information Technology, and to my supervisors Professor Fan Yang Wallentin and Professor Johan Lyhagen at the Department of Statistics. To have had the opportunity to work with someone like Torsten Söderström has been a privilege and an honor. I thank Fan Yang Wallentin and Johan Lyhagen for accepting me as a PhD student and for their involvement in my work. I am deeply grateful to all three for their help and support over the years.

I would also like to express my appreciation to my employer, the Norwegian Business School (BI), for supporting me financially, and to Professor Ulf Henning Olson at the Economics Department at BI for his support over the years. Finally, I would like to thank my fellow PhD students and other colleagues at the Department of Statistics at Uppsala University for a pleasant stay at the Department.

References

Browne, Michael W., and Guangjian Zhang. "Developments in the factor analysis of individual time series." In *Factor Analysis at 100*, pp. 279–306. Routledge, 2007.

Browne, Michael W., and John R. Nesselroade. "Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models." In *A. Maydeu-Olivares, & J. J. McArdle (Eds.), Contemporary psychometrics: A festschrift for Roderick P. McDonald*, pp. 415–452. Erlbaum, 2005.

Chow, Sy-Miin, Moon-ho R. Ho, Ellen L. Hamaker, and Conor V. Dolan. "Equivalence and differences between structural equation modeling and state-space modeling techniques." *Structural Equation Modeling* 17, no. 2 (2010): 303–332.

Dattner, Itai, Harold Ship, and Eberhard O. Voit. "Separable nonlinear least-squares parameter estimation for complex dynamic systems." *Complexity* 2020 (2020).

Den Haan, Wouter J., and A. Levin. "Vector autoregressive covariance matrix estimation." *University of California, San Diego, manuscript* (1998).

du Toit, Stephen HC, and Michael W. Browne. "Structural equation modeling of multivariate time series." *Multivariate Behavioral Research* 42, no. 1 (2007): 67–101.

Golub, Gene H., and Victor Pereyra. "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate." *SIAM Journal on numerical analysis* 10, no. 2 (1973): 413–432.

Golub, Gene, and Victor Pereyra. "Separable nonlinear least squares: the variable projection method and its applications." *Inverse problems* 19, no. 2 (2003): R1.

Hansen, Lars Peter. "Large sample properties of generalized method of moments estimators." *Econometrica: Journal of the Econometric Society* (1982): 1029–1054.

Lyhagen, Johan. "The exact covariance matrix of dynamic models with latent variables." *Statistics & probability letters* 75, no. 2 (2005): 133–139.

Mossberg, Magnus, and Torsten Söderström. "Approximative weighting for a covariance- matching approach for identifying errors- in- variables systems." *International Journal of Adaptive Control and Signal Processing* 25, no. 6 (2011): 535–543.

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

Newey, Whitney K., and Kenneth D. West. "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica: Journal of the Econometric Society* 55, no. 3 (1987): 703–708

Sjoberg, Jonas, and Mats Viberg. "Separable non-linear least-squares minimization-possible improvements for neural net fitting." In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pp. 345–354. IEEE, 1997.

Söderström, Torsten, and Magnus Mossberg. "Accuracy analysis of a covariance matching approach for identifying errors-in-variables systems." *Automatica* 47, no. 2 (2011): 272–282.

Söderström, Torsten, Magnus Mossberg, and Mei Hong. "A covariance matching approach for identifying errors-in-variables systems." *Automatica* 45, no. 9 (2009): 2018–2031.

Söderström, Torsten. "Errors-in-variables methods in system identification." *Automatica* 43, no. 6 (2007): 939–958.

Söderström, Torsten. "System identification for the errors-in-variables problem." *Transactions of the Institute of Measurement and Control* 34, no. 7 (2012): 780–792.

Van Buuren, Stef. "Fitting ARMA time series by structural equation models." *Psychometrika* 62, no. 2 (1997): 215–236.

White, Halbert, and Ian Domowitz. "Nonlinear regression with dependent observations." *Econometrica: Journal of the Econometric Society* (1984): 143–161.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 200

Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-482081

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022