Behaviormetrika manuscript No.
(will be inserted by the editor)

Comparisons among several consistent estimators of structural equation models

Yoshio Takane · Heungsun Hwang

Received: date / Accepted: date

Abstract With the advent of consistent partial least squares (PLSc), an interest has surged in comparing the quality of various estimation methods in structural equation models. Of particular interest are, beside PLSc, Bentler's non-iterative confirmatory factor analysis, Hägglund's instrumental variable (IV) estimation method, and Ihara-Kano's non-iterative uniqueness estimation method. All of these methods yield consistent estimates of parameters in measurement models (factor loadings and unique variances), but require additional steps to estimate parameters in structural models (covariances among latent variables (LVs) and path coefficients). These additional steps typically involve calculating LV scores, either correlating them or applying regression analysis, and correcting possible "biases" incurred by the use of LV scores as proxies of true LVs. In this paper, we conduct a Monte Carlo study to evaluate parameter recovery capabilities of the above LV extraction methods in conjunction with subsequent LV score construction and bias correction methods. We also compare these methods against more conventional estimation methods, such as the full least squares and maximum likelihood methods, that estimate parameters in both measurement and structural models simultaneously. In addition, we examine three methods of estimating standard errors (SEs) of estimated parameters from a single data set, the bootstrap method,

Y. Takane
Department of Psychology, University of Victoria
5173 Del Monte Avenue, Victoria, BC V8Y 1X3 Canada
Tel.: +250-744-0076
Fax: +250-744-0076
E-mail: yoshio.takane@mcgill.ca

H. Hwang
Department of Psychology, McGill University
2001 McGill College, Room 710, Montreal, QC H3A 1G1, Canada
Tel.: +514-398-8021
Fax: +514-398-4896
E-mail: heungsun.hwang@mcgill.ca

ordinary least squares (OLS) regression, and the inverse Hessian method. The SEs are important in assessing the reliability of parameter estimates, and in testing their significance. It was found that Hägglund's method used to extract one LV at a time from each block of observed variables, combined with Croon's bias correction method, worked best in both parameter recovery and resistance to improper solutions, and that the bootstrap method provided the most accurate estimates of SEs.

Keywords Bentler's non-iterative confirmatory factor analysis · Hägglund's instrumental variable (IV) estimation method · Ihara-Kano's non-iterative uniqueness estimation method · Latent variable (LV) scores · Croon's bias correction method · Skrondal-Laake's method · Consistent partial least squares (PLSc) · Blockwise and full least squares and maximum likelihood methods · Bollen's two-stage least squares (2SLS) method · The bootstrap method · Ordinary least squares (OLS) · The inverse Hessian method

Acknowledgements The work reported in this paper has been supported by research grants from the Natural Sciences and Engineering Research Council of Canada to both authors.

1 Introduction

Structural equation models (SEMs) provide useful analytic tools for data arising from research in many scientific disciplines by allowing statistical evaluations of *a priori* hypothesized relationships among observed variables. They are currently very popular in psychology, education, medicine, etc., and many other natural and social sciences. A wide variety of methods have been developed for estimating parameters in structural equation models (SEMs). This paper addresses the question of how the different methods of estimation compare with each other in their ability to recover parameters in SEMs.

Our initial interest in this problem has been inspired by the recent development of consistent partial least squares (PLSc; Dijkstra and Shermelleh-Engel, 2014). Specifically, we were interested in finding out how well this newly proposed method performed relative to more conventional estimation methods that are based on covariance structure analysis. These methods include Bentler's (1982) non-iterative confirmatory factor analysis, Hägglund's (1982) instrumental variable (IV) estimation method, Ihara-Kano's (1986) non-iterative uniqueness estimation method, and blockwise unweighted least squares (ULS) and maximum likelihood (ML) methods. Like PLSc, all of these methods require additional steps. After factor loadings and unique variances (parameters in measurement models) are estimated by these methods, latent variable (LV) scores are calculated, which are then used to estimate covariances (or correlations) and/or path coefficients among LVs (parameters in structural models). The estimates of structural parameters thus obtained are, however, often "biased," where the word "bias" means the difference between estimates of structural parameters calculated from LV scores and those for true LVs. Specifically, let t_i $(i = 1, \cdots, m)$ indicate the random variable representing the LV score for the ith LV denoted by θ_i. A bias in covariance between LVs arises when the covariance between t_i and t_j is not equal to the covariance between θ_i and θ_j. A bias in path coefficients arises when regression coefficients calculated from t_i's are not equal to the path coefficients between θ_i's. (Note that this use of "bias" is different from its standard usage in statistics, where the bias usually means the difference between the expected value of an estimator and the corresponding population parameter. We will encounter an instance of the word bias used in the traditional sense in Section 7.3.)

When the bias occurs, it must be corrected. Several methods for correcting bias have been proposed, including Croon's method (Croon, 2002), Skrondal-Laake's (SL) method (Skrondal and Laake, 2001), and Dijkstra's PLSc method (Dijkstra and Schermelleh-Engel, 2014). These methods have been proven useful in obtaining better (less biased) estimates of structural parameters (e.g., Lu, Kwan, Thomas, and Cedzynski, 2014). It is interesting to compare the initial LV extraction methods along with these bias correction methods, and see which combination of the methods works best in their parameter recovery. In this paper, we use a Monte Carlo study to achieve this goal.

All the methods mentioned so far are multi-step procedures. These methods are mostly non-iterative, and computationally simple and quick. There are,

however, one-step estimation procedures as well, such as full-ULS, full-GLS (generalized least squares), and full-ML methods, that estimate all parameters in SEMs simultaneously based on single optimization criteria. These methods are mostly iterative, involving much heavier computation. However, they are generally regarded as the state-of-the-art estimation methods in SEMs. Part of the reason for their popularity is due to the fact that some of these methods (full-GLS and full-ML) are known to provide estimates of parameters with better statistical properties. That is, they provide not only consistent but also asymptotically efficient estimates of parameters under some conditions (e.g., correctness of fitted models, a large sample). Apart from difficulties of meeting these conditions, however, one may wonder to what extent this theoretical expectation holds in finite samples. This question naturally leads to our second interest in this paper, that of comparing the multi-step methods against the one-step methods. Bollen's (1996) two-stage least squares (2SLS) method is added to the list of one-step procedures because this method directly obtains estimates of path coefficients without estimating any other parameters. It is of interest to see how this method compares against other methods in its ability to recover parameters.

Standard errors (SEs) are important quantities in assessing the reliability of estimated parameters, and in testing their significance (Devlieger, Mayer, and Rosseel, 2015). In our study, we obtain benchmark estimates of SEs from replicated samples of data generated from a population model. This is not feasible, however, in practical data analytic situations. We examine three methods of estimating SEs from a single data set, the bootstrap method, ordinary least squares (OLS) regression, and the inverse Hessian method, to see which method provides the most accurate estimates of SEs.

This paper is organized as follows. In the next section (Section 2), we lay out the basic SEMs we assume throughout this paper. We then (Section 3) discuss methods for initial LV extractions, starting from Bentler's non-iterative confirmatory factor analysis (Section 3.1), followed by Hägglund's instrumental variable (IV) estimation method (Section 3.2), Ihara-Kano's non-iterative uniqueness estimation method (Section 3.3), and other (blockwise) methods (Section 3.4). As has been noted above, these methods require additional steps to estimate parameters in structural models, namely calculations of LV scores, and of covariances and/or regression coefficients among them, and possible bias corrections. We discuss methods of constructing LV scores in Section 4, and methods of bias corrections in Section 5. In the latter, we discuss Croon's method (Section 5.1), Skrondal-Laake's method (Section 5.2), and PLSc (Section 5.3). In Section 6, we briefly discuss one-step methods, such as the full-ULS, full-GLS, and full-ML methods (Section 6.1), and Bollen's 2SLS method (Section 6.2). These one-step methods serve as benchmark methods against which the performance of the multi-step methods are assessed. In Section 7, we present our empirical study. We first describe assumed population parameters (Section 7.1), then exact implementations of the estimation methods to be compared (Section 7.2), performance and other measures of interest used in

the study (Section 7.3), and report the main results (Section 7.4). We conclude with a summary of the results (Section 8).

2 The Basic Structural Equation Model

This short section describes the basic SEM we assume throughout this paper. Let there be m sets of observed variables (indicators), each assumed to be generated by a single common LV (factor). Let $\mathbf{y}_i$ $(i = 1, \cdots, m)$ denote the random vector of p_i indicators in the ith set, and let θ_i denote the random variable for the ith LV. We may write

$$\mathbf{y}_i = \mathbf{a}_i\theta_i + \mathbf{e}_i, \tag{1}$$

where $\mathbf{a}_i$ is the p_i-component loading vector, and $\mathbf{e}_i$ is the random vector of unique variations. We assume that θ_i has 0 mean and unit variance, that $\mathbf{e}_i$ also has 0 means, that the elements of $\mathbf{e}_i$ are mutually independent, and that they are independent from θ_i. Notice that at this point no other distributional assumptions (such as normality) are made on either θ_i or on $\mathbf{e}_i$. It follows that

$$\boldsymbol{\Sigma}_{ii} = \mathrm{E}[\mathbf{y}_i\mathbf{y}_i'] = \mathrm{E}[\theta_i^2]\mathbf{a}_i\mathbf{a}_i' + \boldsymbol{\Delta}_i = \mathbf{a}_i\mathbf{a}_i' + \boldsymbol{\Delta}_i, \tag{2}$$

and

$$\boldsymbol{\Sigma}_{ij} = \mathrm{E}[\mathbf{y}_i\mathbf{y}_j'] = \mathrm{E}[\theta_i\theta_j]\mathbf{a}_i\mathbf{a}_j' = \rho_{ij}\mathbf{a}_i\mathbf{a}_j', \tag{3}$$

where the subscript j $(j \neq i)$ refers to another set of observed variables and the associated LV, E stands for expectation, $\mathrm{E}(\theta_i^2) = \rho_{ii}$ is the variance of θ_i (assumed unity), $\mathrm{E}[\mathbf{e}_i\mathbf{e}_i'] = \mathrm{V}[\mathbf{e}_i] = \boldsymbol{\Delta}_i$ (assumed diagonal), and $\mathrm{E}[\theta_i\theta_j] = \rho_{ij}$ is the correlation between θ_i and θ_j.

Let $\mathbf{Y}_i$ $(i = 1, \cdots, m)$ represent an $n-$cases by p_i-variables matrix of mean centered data (realized values of $\mathbf{y}_i$). Then, sample analogues of $\boldsymbol{\Sigma}_{ii}$ and $\boldsymbol{\Sigma}_{ij}$, denoted by $\mathbf{S}_{ii}$ and $\mathbf{S}_{ij}$, are calculated by $\mathbf{S}_{ii} = \mathbf{Y}_i'\mathbf{Y}_i/(n-1)$ and $\mathbf{S}_{ij} = \mathbf{Y}_i'\mathbf{Y}_j/(n-1)$. Estimates of $\mathbf{a}_i$ and $\boldsymbol{\Delta}_i$, denoted by placing a hat on them, are obtained by applying factor analysis to $\mathbf{S}_{ii}$.

It is sometimes useful to represent (2) and (3) simultaneously. Let

$$\boldsymbol{\Sigma}_{(ij)} = \begin{bmatrix} \boldsymbol{\Sigma}_{ii} & \boldsymbol{\Sigma}_{ij} \\ \boldsymbol{\Sigma}_{ij}' & \boldsymbol{\Sigma}_{jj} \end{bmatrix}, \tag{4}$$

$$\mathbf{A}_{ij} = \begin{bmatrix} \mathbf{a}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_j \end{bmatrix}, \tag{5}$$

$$\boldsymbol{\Phi}_{ij} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}, \tag{6}$$

and

$$\boldsymbol{\Delta}_{ij} = \begin{bmatrix} \boldsymbol{\Delta}_i & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Delta}_j \end{bmatrix}. \tag{7}$$

Then,

$$\boldsymbol{\Sigma}_{(ij)} = \mathbf{A}_{ij}\boldsymbol{\Phi}_{ij}\mathbf{A}_{ij}' + \boldsymbol{\Delta}_{ij}. \tag{8}$$

The above expressions for two blocks of variables can easily be extended to multiple blocks of variables.

3 Methods for Initial Latent Variable Extraction

In this section, we discuss methods for initial LV extraction in the group of multi-step procedures. In these methods, parameters in measurement models ($\mathbf{a}_i$ and $\boldsymbol{\Delta}_i$, and in some cases, a subset of $\boldsymbol{\Phi}_{ij}$) are estimated prior to the estimation of structural parameters (the other subset of $\boldsymbol{\Phi}$ and path coefficients). As noted above, there are a number of methods in this class. We discuss some of them here, namely Bentler's confirmatory factor analysis, Hägglund's IV estimation method, Ihara-Kano's non-iterative uniqueness estimation method, and blockwise ULS and ML methods. (PLSc will be separately treated in Section 5.3.)

3.1 Bentler's non-iterative confirmatory factor analysis

We first discuss Bentler's (1982) non-iterative confirmatory factor analysis method. This method is easy to program, quick to compute, and yet flexible enough to accommodate various situational demands required in our study. In particular, this method can be used for extracting one LV at a time from each block of indicators as well as for extracting multiple LVs simultaneously from multiple blocks of indicators. The former is required in Croon's bias correction method, while the latter in Skrondal-Laake's method.

We begin with the situation in which one LV is separately extracted from each block of indicators. We first rewrite (2) as

$$\boldsymbol{\Sigma}_{ii} = \mathbf{a}_i \phi_i \mathbf{a}_i' + \boldsymbol{\Delta}_i, \tag{9}$$

where $\phi_i = \mathrm{E}[\theta_i^2] = 1$. An observed variable in block i deemed most representative of the LV θ_i is called a pivotal variable. We assume that a pivotal variable has been identified (perhaps having the largest average correlation with other variables in the block), and that the pivotal variable is always placed in the leading position in the block by permuting the variables in the block. We then rescale $\mathbf{a}_i$ in such a way that the loading on the pivotal variable is unity. This rescaling can be done by dividing all elements of $\mathbf{a}_i$ by its first element a_{i1}, that is,

$$\tilde{\mathbf{a}}_i = \mathbf{a}_i a_{i1}^{-1} = \begin{pmatrix} a_{i1} \\ \mathbf{a}_{i2} \end{pmatrix} a_{i1}^{-1} = \begin{pmatrix} 1 \\ \mathbf{a}_{i2}^* \end{pmatrix}, \tag{10}$$

where $\mathbf{a}_{i2}^* = \mathbf{a}_{i2} a_{i1}^{-1}$. Then, we may rewrite (9) as

$$\boldsymbol{\Sigma}_{ii} = \begin{bmatrix} \sigma_{i11} & \boldsymbol{\sigma}_{i1}' \\ \boldsymbol{\sigma}_{i1} & \boldsymbol{\Sigma}_{ii}^* \end{bmatrix} = \begin{bmatrix} \phi_i^* + \delta_{i1} & \phi_i^* \mathbf{a}_{i2}^{*'} \\ \mathbf{a}_{i2}^* \phi_i^* & \mathbf{a}_{i2}^* \phi_i^* \mathbf{a}_{i2}^{*'} + \boldsymbol{\Delta}_i^* \end{bmatrix}, \tag{11}$$

where $\phi_i^* = a_{i1}^2$, and $\boldsymbol{\Delta}_i = \begin{bmatrix} \delta_{i1} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Delta}_i^* \end{bmatrix}$. Since

$$\mathbf{a}_{i2}^* = \boldsymbol{\sigma}_{i1} \phi_i^{*-1} \tag{12}$$

from the bottom left corner of $\boldsymbol{\Sigma}_{ii}$, we obtain

$$\boldsymbol{\Sigma}^*_{ii} = \boldsymbol{\sigma}_{i1}\kappa_i\boldsymbol{\sigma}'_{i1} + \boldsymbol{\Delta}^*_i, \tag{13}$$

where $\kappa_i = \phi_i^{*-1}$. The sample analogue of (11) is given by

$$\mathbf{S}_{ii} = \begin{bmatrix} s_{i11} & \mathbf{s}'_{i1} \\ \mathbf{s}_{i1} & \mathbf{S}^*_{ii} \end{bmatrix}. \tag{14}$$

We now assume that $\boldsymbol{\sigma}_{i1}$ is closely approximated by $\mathbf{s}_{i1}$, so that the former can be replaced by the latter in (13). We then have

$$\boldsymbol{\Sigma}^*_{ii} = \mathbf{s}_{i1}\kappa_i\mathbf{s}'_{i1} + \boldsymbol{\Delta}^*_i, \tag{15}$$

and

$$\mathbf{a}^*_{i2} = \mathbf{s}_{i1}\phi_i^{*-1}. \tag{16}$$

We estimate $\phi^*_i = \kappa_i^{-1}$ and $\boldsymbol{\Delta}^*_i$ in the above equation by minimizing the following criterion:

$$f = \mathrm{tr}[(\boldsymbol{\Sigma}^*_{ii} - \mathbf{S}^*_{ii})\mathbf{W}]^2, \tag{17}$$

where $\mathbf{W}$ is the weight matrix set equal to $\mathbf{I}_{p_i}$ for the unweighted least squares estimation (ULS), or to $\mathbf{S}^{*-1}_{ii}$ for the generalized least squares estimation (GLS). We allow both options. By differentiating f with respect to κ_i and setting the result equal to zero, we obtain

$$(\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1})\hat{\kappa}_i(\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1}) - \mathbf{s}'_{i1}\mathbf{W}(\mathbf{S}^*_{ii} - \boldsymbol{\Delta}^*_i)\mathbf{W}\mathbf{s}_{i1} = 0, \tag{18}$$

resulting in

$$\hat{\kappa}_i = (\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1})^{-1}\mathbf{s}'_{i1}\mathbf{W}(\mathbf{S}^*_{ii} - \boldsymbol{\Delta}^*_i)\mathbf{W}\mathbf{s}_{i1}(\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1})^{-1}, \tag{19}$$

or

$$\hat{\phi}^*_i = \hat{\kappa}^{-1} = \mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1}(\mathbf{s}'_{i1}\mathbf{W}(\mathbf{S}^*_{ii} - \boldsymbol{\Delta}^*_i)\mathbf{W}\mathbf{s}_{i1})^{-1}\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1}. \tag{20}$$

We now estimate $\boldsymbol{\Delta}^*_i$ in the above. Let $\boldsymbol{\delta}^*_i$ denote the column vector consisting of the diagonal elements of $\boldsymbol{\Delta}^*_i$. By differentiating f with respect to $\boldsymbol{\delta}^*_i$ and setting the result to a zero vector, we obtain

$$\frac{\partial \boldsymbol{\Delta}^*_i}{\partial \boldsymbol{\delta}^*_i}\mathrm{vec}[\mathbf{W}(\boldsymbol{\Sigma}^*_{ii} - \mathbf{S}^*_{ii})\mathbf{W}] = \mathbf{P}\mathrm{vec}[\mathbf{W}(\mathbf{s}_{i1}\kappa_i\mathbf{s}'_{i1} + \hat{\boldsymbol{\Delta}}^*_i - \mathbf{S}^*_{ii})\mathbf{W}] = \mathbf{0}, \tag{21}$$

where $\mathbf{P} = \frac{\partial \boldsymbol{\Delta}^*_i}{\partial \boldsymbol{\delta}^*_i}$ is a $p_i - 1$ by $(p_i - 1)^2$ matrix with a zero element everywhere except ones where row and column designate the same parameter, and where the vec operator strings out columns of a matrix into a tall vector. Note that $\mathbf{P}\mathrm{vec}(\boldsymbol{\Delta}^*_i) = \boldsymbol{\delta}^*_i$, or more generally,

$$\mathbf{P}\mathrm{vec}(\mathbf{B}) = \mathrm{diag}(\mathbf{B})\mathbf{1}, \tag{22}$$

where $\mathrm{diag}(\mathbf{B})$ indicates a diagonal matrix whose diagonal elements are equal to the diagonal elements of a square matrix $\mathbf{B}$, and $\mathbf{1}$ is a vector of ones. Note also that

$$\mathbf{P}'\boldsymbol{\delta}^*_i = \mathrm{vec}(\boldsymbol{\Delta}^*_i), \tag{23}$$

and $\mathbf{PP}' = \mathbf{I}_{p_i-1}$. By substituting the estimate of $\hat{\kappa}_i$ in (19) into (21), we obtain

$$\mathbf{P}\text{vec}[\mathbf{G}(\mathbf{S}^*_{ii} - \hat{\boldsymbol{\Delta}}^*_i)\mathbf{G} + \mathbf{W}\hat{\boldsymbol{\Delta}}^*_i\mathbf{W} - \mathbf{W}\mathbf{S}^*_{ii}\mathbf{W}] = \mathbf{0}, \tag{24}$$

where

$$\mathbf{G} = \mathbf{W}\mathbf{s}_{i1}(\mathbf{s}'_{i1}\mathbf{W}\mathbf{s}_{i1})^{-1}\mathbf{s}'_{i1}\mathbf{W}. \tag{25}$$

It follows that

$$\mathbf{P}\text{vec}(\mathbf{W}\hat{\boldsymbol{\Delta}}^*_i\mathbf{W} - \mathbf{G}\hat{\boldsymbol{\Delta}}^*_i\mathbf{G}) = \mathbf{P}\text{vec}(\mathbf{W}\mathbf{S}^*_{ii}\mathbf{W} - \mathbf{G}\mathbf{S}^*_{ii}\mathbf{G}), \tag{26}$$

and that

$$\mathbf{P}(\mathbf{W}\otimes\mathbf{W} - \mathbf{G}\otimes\mathbf{G})\text{vec}(\hat{\boldsymbol{\Delta}}^*_i) = \mathbf{P}\text{vec}(\mathbf{W}\mathbf{S}^*_{ii}\mathbf{W} - \mathbf{G}\mathbf{S}^*_{ii}\mathbf{G}), \tag{27}$$

where $\otimes$ indicates a Kronecker product (i.e., $\mathbf{A}\otimes\mathbf{B} = [a_{ij}\mathbf{B}]$). Note that we used a well known identity, $\text{vec}(\mathbf{ABC}) = (\mathbf{C}'\otimes\mathbf{A})\text{vec}(\mathbf{B})$, to obtain (27) from (26). We have

$$\hat{\boldsymbol{\delta}}^*_i = [\mathbf{P}(\mathbf{W}\otimes\mathbf{W} - \mathbf{G}\otimes\mathbf{G})\mathbf{P}']^{-1}\mathbf{P}\text{vec}(\mathbf{W}\mathbf{S}^*_{ii}\mathbf{W} - \mathbf{G}\mathbf{S}^*_{ii}\mathbf{G}). \tag{28}$$

Note that $\boldsymbol{\delta}^*_i = \mathbf{P}'\text{vec}(\boldsymbol{\Delta}^*_i)$ by (23). We then arrange the elements of $\hat{\boldsymbol{\delta}}^*_i$ into the diagonal matrix $\hat{\boldsymbol{\Delta}}^*_i$. Finally, the estimate of ϕ^*_i is obtained by putting $\hat{\boldsymbol{\Delta}}^*_i$ in (20).

It remains to obtain estimates of δ_{i1} and $\mathbf{a}_i$. We first obtain

$$\hat{\delta}_{i1} = s_{i11} - \hat{\phi}^*_i \tag{29}$$

and

$$\hat{\mathbf{a}}^*_{i2} = \mathbf{s}_{i1}\hat{\phi}^{*-1}_i = \mathbf{s}_{i1}\hat{\kappa}_i. \tag{30}$$

We then have

$$\hat{a}_{i1} = \hat{\phi}^{*1/2}_i, \tag{31}$$

and

$$\hat{\mathbf{a}}_{i2} = \hat{\mathbf{a}}^*_{i2}\hat{\phi}^{*1/2}_i = \mathbf{s}_{i1}\hat{\phi}^{*-1/2}_i. \tag{32}$$

This concludes Bentler's estimation method for the first situation.

We next discuss the situation in which multiple LVs are derived from multiple blocks of indicators. We still assume that each block of indicators are generated by a single LV (see Model (1)). For ease of exposition, we also assume that we derive two LVs from two blocks of indicators, so that the model can be stated as in (4) through (8). We further assume that pivotal variables in the two blocks are placed in the leading positions. This means that $\mathbf{A}_{ij}$ is of the form:

$$\mathbf{A}_{ij} = \begin{bmatrix} a_{i1} & 0 \\ 0 & a_{j1} \\ \mathbf{a}_{i2} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{j2} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_A \\ \mathbf{A}_{ij2} \end{bmatrix}, \tag{33}$$

where $\mathbf{D}_A$ is a diagonal matrix, and $\mathbf{A}_{ij2}$ is a block diagonal matrix. As before, we rescale the LVs in such a way that

$$\boldsymbol{\Sigma}_{(ij)} = (\mathbf{A}_{ij}\mathbf{D}_A^{-1})(\mathbf{D}_A\boldsymbol{\Phi}_{ij}\mathbf{D}_A)(\mathbf{D}_A^{-1}\mathbf{A}_{ij}') + \boldsymbol{\Delta}_{ij} = \mathbf{A}_{ij}^*\boldsymbol{\Phi}_{ij}^*\mathbf{A}_{ij}^{*'} + \boldsymbol{\Delta}_{ij}, \quad (34)$$

where

$$\mathbf{A}_{ij}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \mathbf{a}_{i2}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{j2}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{A}_{ij2}^* \end{bmatrix}, \text{ and } \mathbf{A}_{ij2}^* = \mathbf{A}_{ij2}\mathbf{D}_A^{-1}, \quad (35)$$

and $\boldsymbol{\Phi}_{ij}^* = \mathbf{D}_A\boldsymbol{\Phi}_{ij}\mathbf{D}_A$. We now partition $\boldsymbol{\Sigma}_{(ij)}$ and $\mathbf{S}_{(ij)}$ conformably with the partition of $\mathbf{A}_{ij}^*$. This is analogous to (11) and (14). The rest of the estimation procedure proceeds much the same way as in the previous situation with two notable exceptions. First of all, some scalar quantities (e.g., s_{i11}) and vectors (e.g., $\mathbf{s}_{i1}$) in the previous situation are now matrices, so that some inversion operations involve real matrix inversions (rather than mere reciprocals of scalars). Also, (29) must be replaced by

$$\hat{\boldsymbol{\Delta}}_{ij1} = \text{diag}(\mathbf{S}_{ij1} - \hat{\boldsymbol{\Phi}}_{ij}^*). \quad (36)$$

Secondly, in estimating $\mathbf{A}_{ij2}^*$, we must take into account its block diagonality. That is, analogously to $\mathbf{s}_{i1} = \mathbf{a}_i^*\phi^*$, we have

$$\mathbf{S}_{ij1} = \begin{bmatrix} \mathbf{S}_{i1} \\ \mathbf{S}_{j1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{i2}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{j2}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_i^{*'} \\ \boldsymbol{\phi}_j^{*'} \end{bmatrix} + \mathbf{E}_{ij}. \quad (37)$$

The least squares estimate of $\hat{\mathbf{a}}_{k2}^*$ $(k = i, j)$ is given by

$$\hat{\mathbf{a}}_{k2}^* = \mathbf{S}_{k1}\boldsymbol{\phi}_k^*(\boldsymbol{\phi}_k^{*'}\boldsymbol{\phi}_k^*)^{-1}. \quad (38)$$

Finally, $\hat{\mathbf{a}}_{k2}^*$ and $\hat{\boldsymbol{\Phi}}_{ij}^*$ are rescaled in a manner similar to (31) and (32), and in such a way that $\text{diag}(\hat{\boldsymbol{\Phi}}) = \mathbf{I}_2$.

3.2 Hägglund's instrumental variable (IV) estimation method

Hägglund's (1982) method is used only for extracting one LV at a time from each block of indicator variables, combined with Croon's bias correction method. As will be seen shortly, this method is similar to Bentler's method in many ways. Indeed, Bentler (1982) notes that his method was inspired by Hägglund's method. A major difference is that in Häggland's method, $\mathbf{a}_i^*$ is estimated first, based on which $\boldsymbol{\Delta}_i^*$ and ϕ_i^* are estimated. In Bentler's method, on the other hand, $\mathbf{s}_{i1}$ is used as a surrogate for $\mathbf{a}_{i2}^*$, based on which $\boldsymbol{\Delta}_i^*$ and ϕ_i^* are estimated first, followed by $\mathbf{a}_{i2}^*$. In what follows, we emphasize parallels between the two methods, wherever possible.

Let the observed variables in block i be partitioned into three subsets, one consisting of variable 1 (the pivotal variable), the second consisting of variable

j ($j = 2, \cdots, p_i$, i.e., any variable other than 1), and the set K consisting of all other variables in the block. We rescale $\mathbf{a}_i$ in the same way as in (10). Then, for $j = 2, \cdots, p_i$, we have

$$\begin{aligned}&\boldsymbol{\sigma}'_{Kj}\boldsymbol{\sigma}_{K1}(\boldsymbol{\sigma}'_{K1}\boldsymbol{\sigma}_{K1})^{-1} = \\ &\quad a^*_{ij}\phi^*_i\mathbf{a}^{*'}_{iK}\mathbf{a}^*_{iK}\phi^*_i a^*_{i1}(a^*_{i1}\phi^*_i\mathbf{a}^{*'}_{iK}\mathbf{a}^*_{iK}\phi^*_i a^*_{i1})^{-1} = a^*_{ij},\end{aligned} \tag{39}$$

where $\boldsymbol{\sigma}_{Kj}$ is the vector of covariances between variables in set K and variable j, and $a^*_{i1} = 1$. Similarly,

$$\boldsymbol{\sigma}'_{Kj}\boldsymbol{\Sigma}^{-1}_{KK}\boldsymbol{\sigma}_{K1}(\boldsymbol{\sigma}'_{K1}\boldsymbol{\Sigma}^{-1}_{KK}\boldsymbol{\sigma}_{K1})^{-1} = a^*_{ij}. \tag{40}$$

The above two identities show that sample analogues of the quantities on the left hand side of (39) and (40) may be used as estimates of a^*_{ij}. That is, for $j = 2, \cdots, p_i$,

$$\hat{a}^*_{ij} = \mathbf{s}'_{Kj}\mathbf{s}_{K1}(\mathbf{s}'_{K1}\mathbf{s}_{K1})^{-1}, \tag{41}$$

and

$$\hat{a}^*_{ij} = \mathbf{s}'_{Kj}\mathbf{S}^{-1}_{KK}\mathbf{s}_{K1}(\mathbf{s}'_{K1}\mathbf{S}^{-1}_{KK}\mathbf{s}_{K1})^{-1}, \tag{42}$$

where $\mathbf{s}$'s and $\mathbf{S}_{KK}$ are relevant portions of $\mathbf{S}_{ii}$. Hägglund (1982) calls the method that uses (41) FABIN2 (Factor Analysis By INstrumental variables), and the method that uses (42) FABIN3.

Once the estimate of $\mathbf{a}^*_i$ is obtained, the remaining parameters (ϕ^*_i and $\boldsymbol{\Delta}_i$) are estimated in a manner similar to Bentler's method. We fit $\hat{\mathbf{a}}^*_i\phi^*_i\hat{\mathbf{a}}^{*'}_i + \boldsymbol{\Delta}_i$ to $\mathbf{S}_{ii}$. Let

$$f^* = \mathrm{tr}(\mathbf{S}_{ii} - \hat{\mathbf{a}}^*_i\phi^*_i\hat{\mathbf{a}}^{*'}_i - \boldsymbol{\Delta}_i)^2. \tag{43}$$

(This is analogous to (17) with $\mathbf{W} = \mathbf{I}_{p_i}$ (ULS) in Bentler's method. The difference is that (17) is defined only for non-pivotal variables, while (43) is defined for all variables in the block.) By differentiating f^* with respect to ϕ^*_i and $\boldsymbol{\Delta}_i$, and setting the results equal to zero, we obtain

$$\hat{\mathbf{a}}^{*'}_i(\mathbf{S}_{ii} - \hat{\mathbf{a}}^*_i\hat{\phi}^*_i\hat{\mathbf{a}}^{*'}_i - \hat{\boldsymbol{\Delta}}_i)\hat{\mathbf{a}}^*_i = 0, \tag{44}$$

and

$$\mathrm{diag}(\mathbf{S}_{ii} - \hat{\mathbf{a}}^*_i\hat{\phi}^*_i\hat{\mathbf{a}}^{*'}_i - \hat{\boldsymbol{\Delta}}_i) = \mathbf{O}, \tag{45}$$

respectively. From (44), we obtain

$$\hat{\phi}^*_i = (\hat{\mathbf{a}}^{*'}_i\hat{\mathbf{a}}^*_i)^{-1}\hat{\mathbf{a}}^{*'}_i(\mathbf{S}_{ii} - \hat{\boldsymbol{\Delta}}_i)\hat{\mathbf{a}}^*_i(\hat{\mathbf{a}}^{*'}_i\hat{\mathbf{a}}^*_i)^{-1}. \tag{46}$$

This is analogous to (20) in Bentler's method. (See the relationship between $\mathbf{s}_{i1}$ and $\mathbf{a}^*_{i2}$ in (16).) By substituting (46) into (45), we obtain

$$\hat{\boldsymbol{\Delta}}_i = \mathrm{diag}[\mathbf{S}_{ii} - \mathbf{D}(\mathbf{S}_{ii} - \hat{\boldsymbol{\Delta}}_i)\mathbf{D}], \tag{47}$$

where $\mathbf{D} = \hat{\mathbf{a}}^*_i(\hat{\mathbf{a}}^{*'}_i\hat{\mathbf{a}}^*_i)^{-1}\hat{\mathbf{a}}^{*'}_i$. (This quantity is analogous to $\mathbf{G}$ in Bentler's method.) From (47), we further obtain

$$\hat{\boldsymbol{\Delta}}_i - \mathrm{diag}(\mathbf{D}\hat{\boldsymbol{\Delta}}_i\mathbf{D}) = \mathrm{diag}(\mathbf{S}_{ii} - \mathbf{D}\mathbf{S}_{ii}\mathbf{D}). \tag{48}$$

Let $\mathbf{E} = \mathbf{I} - \mathbf{D} * \mathbf{D}$, where "$*$" indicates an elementwise multiplication, and let $\hat{\boldsymbol{\delta}}_i$ and $\mathbf{g}$ denote column vectors of the diagonal elements of $\hat{\boldsymbol{\Delta}}_i$ and $\text{diag}(\mathbf{S}_{ii} - \mathbf{D}\mathbf{S}_{ii}\mathbf{D})$, respectively. Then (48) can be rewritten as

$$\mathbf{E}\hat{\boldsymbol{\delta}}_i = \mathbf{g}, \tag{49}$$

or

$$\hat{\boldsymbol{\delta}}_i = \mathbf{E}^{-1}\mathbf{g}. \tag{50}$$

The elements of $\hat{\boldsymbol{\delta}}_i$ are then put in the diagonal elements of $\hat{\boldsymbol{\Delta}}_i$, which is further substituted into (46) to obtain $\hat{\phi}_i^*$. We then rescale $\hat{\mathbf{a}}_i^*$ and $\hat{\phi}_i^*$ to satisfy $\hat{\phi}_i = 1$.

It may be pointed out in passing that (48) is analogous to (27). This may be seen by first noting that (48) can be rewritten as

$$\hat{\boldsymbol{\delta}}_i - \text{diag}(\mathbf{D}\hat{\boldsymbol{\Delta}}_i\mathbf{D})\mathbf{1} = \text{diag}(\mathbf{S}_{ii} - \mathbf{D}\mathbf{S}_{ii}\mathbf{D})\mathbf{1} = \mathbf{g}, \tag{51}$$

where $\mathbf{1}$ is a vector of ones of appropriate size. By noting (22) and (23), the second term on the left hand side of this equation can further be rewritten as

$$\begin{aligned}&\text{diag}(\mathbf{D}\hat{\boldsymbol{\Delta}}_i\mathbf{D})\mathbf{1} = \mathbf{P}\text{vec}(\mathbf{D}\hat{\boldsymbol{\Delta}}_i\mathbf{D}) = \\ &\mathbf{P}(\mathbf{D}\otimes\mathbf{D})\text{vec}(\hat{\boldsymbol{\delta}}_i) = \mathbf{P}(\mathbf{D}\otimes\mathbf{D})\mathbf{P}'\hat{\boldsymbol{\delta}}_i = (\mathbf{D}*\mathbf{D})\hat{\boldsymbol{\delta}}_i,\end{aligned} \tag{52}$$

and the right hand side as

$$\mathbf{g} = \mathbf{P}'\text{vec}(\mathbf{S}_{ii} - \mathbf{D}\mathbf{S}_{ii}\mathbf{D}). \tag{53}$$

The last equality in (52) follows from

$$\mathbf{P}(\mathbf{D}\otimes\mathbf{D})\mathbf{P}' = \mathbf{D}*\mathbf{D}. \tag{54}$$

3.3 Ihara-Kano's non-iterative uniqueness estimation method

Ihara-Kano's (1986) method is also applied to extract one LV from each block of indicators. In this method, unique variances ($\boldsymbol{\Delta}_i$) are estimated first, and then the loading vector ($\mathbf{a}_i$). This is similar to Bentler's method and opposite to Hägglund's method. Unlike both of these methods, Ihara-Kano's method does not rescale the loading vector for estimation.

For any variable index j ($j = 1, \cdots, p_i$), consider evaluating the quantity of the following form, $\sigma_{ju}\sigma_{vj}/\sigma_{uv}$, where σ's are appropriate elements in $\boldsymbol{\Sigma}_{ii}$, and u and $v > u$ are indices of any variables other than j in block i. Assuming the model in (1), we have

$$\sigma_{ju}\sigma_{vj}/\sigma_{uv} = a_{ij}a_{iu}a_{iv}a_{ij}/(a_{iu}a_{iv}) = a_{ij}^2. \tag{55}$$

This means that

$$\sigma_{jj} - \sigma_{ju}\sigma_{vj}/\sigma_{uv} = \delta_j, \tag{56}$$

where δ_j is the jth diagonal element of $\boldsymbol{\Delta}_i$, and that the sample analogue of the left hand side of this identity may be used to estimate the uniqueness of variable j. That is,

$$\hat{\delta}_j = s_{jj} - s_{ju}s_{vj}/s_{uv}, \tag{57}$$

for $j = 1, \cdots, p_i$. This estimate of δ_j, however, depends on the choice of u and v. Ihara and Kano (1986) recommend to choose $u = q$ and $v = r$, such that $|s_{qr}|$ is a maximum over all combinations of q and r such that $q \neq j$, $r \neq j$ and $q < r$.

Once $\boldsymbol{\Delta}_i$ is estimated, estimates of loadings are easily obtained by the eigen-decomposition of $\mathbf{S}_{ii} - \hat{\boldsymbol{\Delta}}_i$. Let d_1 represent the most dominant eigenvalue, and $\mathbf{f}_1$ the corresponding eigenvector of this matrix. Then

$$\hat{\mathbf{a}}_i = \mathbf{f}_1 d_1^{1/2}. \tag{58}$$

One thing we need to be careful in this process is that $-\mathbf{f}_1$ is as good as $\mathbf{f}_1$ as the eigenvector, and in general there is no way to identify which one we get. We thus must make sure the sign of $\hat{\mathbf{a}}_i$ is consistent with that of the population loading vector in evaluating goodness of parameter recovery.

3.4 Blockwise methods for initial latent variable extraction

Beside the non-iterative methods discussed above, there are several iterative estimation methods that can be used for initial LV extraction. We briefly discuss two of them here, ULS (unweighted least squares) and ML (maximum likelihood) methods, that are included in our comparison. We still assume that these methods are used to extract one LV at a time from each block of indicator variables. These methods are called blockwise-ULS and blockwise-ML in this paper.

All the methods we introduced so far estimate different groups of model parameters sequentially, so that parameters estimated at later stages are dependent on those estimated earlier, and although some optimization criteria are used to estimate some subsets of parameters, there were no single criteria optimized to estimate the entire set of parameters. In contrast, both blockwise-ULS and blockwise-ML estimate all parameters within blocks simultaneously. Specifically, blockwise-ULS minimizes

$$g_i = \mathrm{tr}(\mathbf{S}_{ii} - \boldsymbol{\Sigma}_{ii})^2, \tag{59}$$

and blockwise-ML minimizes

$$h_i = \mathrm{tr}(\mathbf{S}_{ii}\boldsymbol{\Sigma}_{ii}^{-1}) - \log|\mathbf{S}_{ii}\boldsymbol{\Sigma}_{ii}^{-1}| - p_i \tag{60}$$

with respect to $\mathbf{a}_i$ and $\boldsymbol{\Delta}_i$ $(i = 1, \cdots, m)$. Minimizations of these criteria require rather elaborate iterative algorithms (e.g., Jöreskog, 1977). We use the Gauss-Newton algorithm to minimize (59), and Fisher's scoring algorithm to minimize (60), developed earlier (Jung and Takane, 2008), although we give

no further details of these algorithms in this paper. Being iterative, they tend to require more computation time than the previous methods, although it is usually not too excessive because of the relatively small number of parameters involved in each block of observed variables.

4 The Methods for Constructing Latent Variable Scores

Once the estimates of loadings are obtained by one of the methods described above, LV scores are calculated. We first discuss methods in which scores are constructed for each LV separately, and then a method in which multiple LVs constructed simultaneously. The former are used in combination with Croon's bias correction method, while the latter with Skrondal-Laake's method.

There are four representative formulas for constructing LV scores (McDonald and Burr, 1967). They are all of the form $t_i = \mathbf{w}_i'\mathbf{y}_i$ $(i = 1, \cdots, m)$, where

$$\textbf{(i) Regression:} \quad \mathbf{w}_i = (\mathbf{a}_i\mathbf{a}_i' + \boldsymbol{\Delta}_i)^{-1}\mathbf{a}_i = \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i d_i, \tag{61}$$

with $d_i = (1 + \mathbf{a}_i'\boldsymbol{\Delta}_i^{-1}\mathbf{a}_i)^{-1}$,

$$\textbf{(ii-a) Least Squares (LS):} \quad \mathbf{w}_i = \mathbf{a}_i(\mathbf{a}_i'\mathbf{a}_i)^{-1}, \tag{62}$$

$$\textbf{(ii-b) Bartlett:} \quad \mathbf{w}_i = \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i(\mathbf{a}_i'\boldsymbol{\Delta}_i^{-1}\mathbf{a}_i)^{-1}, \tag{63}$$

$$\textbf{(iii-b) Anderson-Rubin (A-R):} \quad \mathbf{w}_i = \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i(\mathbf{a}_i'\boldsymbol{\Delta}_i^{-1}\boldsymbol{\Sigma}_{ii}\boldsymbol{\Delta}_i^{-1}\mathbf{a}_i)^{-1/2}. \tag{64}$$

Note that the so-called empirical Bayes method of estimating LV scores reduces to the Regression formula under the additional assumptions of normality on σ and $\mathbf{e}$ in (1) (Skrondal and Rabe-Hesketh, 2004, p. 228). Note also that the Bartlett formula reduces to the LS formula if $\boldsymbol{\Delta}_i$ is ignored (or set to an identity matrix). If the same thing is done on the A-R formula, a new formula results that might be called the Simplified A-R formula:

$$\textbf{(iii-a) Simplified A-R:} \quad \mathbf{w}_i = \mathbf{a}_i(\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}\mathbf{a}_i)^{-1/2}. \tag{65}$$

In both the original and simplified A-R formulas, the variance of t_i is constrained to be unity. It may also be pointed out that although $\boldsymbol{\Sigma}_{ii}$ in these formula are typically estimated by $\hat{\boldsymbol{\Sigma}}_{ii} = \hat{\mathbf{a}}_i\hat{\mathbf{a}}_i' + \hat{\boldsymbol{\Delta}}_i$, it may also be estimated by its sample analogue, i.e., $\mathbf{S}_{ii}$, as is done in PLSc. (See Section 5.3.) The second equality in (61) holds because $(\mathbf{a}_i\mathbf{a}_i' + \boldsymbol{\Delta}_i)^{-1} = \boldsymbol{\Delta}_i^{-1} - \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i d_i \mathbf{a}_i'\boldsymbol{\Delta}_i^{-1}$.

A couple of additional remarks are in order concerning the above formulas. First of all, $\mathbf{w}_i'\mathbf{a}_i = 1$ for the LS and Bartlett formulas. These formulas produce so-called "conditionally unbiased" estimates of LVs because it holds that $\mathrm{E}[t_i|\theta_i] = \theta_i$ (Croon, 2002, p. 215), where $\mathrm{E}[t_i|\theta_i]$ is the conditional expectation of t_i given θ_i. This is readily seen by noting that $\mathrm{E}[\mathbf{y}_i|\theta_i] = \mathbf{a}_i\theta_i$ from (1), so that

$$\mathrm{E}[t_i|\theta_i] = \mathrm{E}[\mathbf{w}_i'\mathbf{y}_i|\theta_i] = \mathbf{w}_i'\mathrm{E}[\mathbf{y}_i|\theta_i] = \mathbf{w}_i'\mathbf{a}_i\theta_i = \theta_i. \tag{66}$$

It also holds that $\mathrm{E}[t_i|\mathbf{x}] = \mathrm{E}[\theta_i|\mathbf{x}]$ for any external variables $\mathbf{x}$. This follows from

$$\begin{aligned}\mathrm{E}[t_i|\mathbf{x}] &= \int t_i f(t_i|\mathbf{x})dt_i = \int t_i[\int g(t_i|\theta_i)h(\theta_i|\mathbf{x})d\theta_i]dt_i \\ &= \int [\int t_i g(t_i|\theta_i)dt_i]h(\theta_i|\mathbf{x})d\theta_i = \int \mathrm{E}[t_i|\theta_i]h(\theta_i|\mathbf{x})d\theta_i \\ &= \int \theta_i h(\theta_i|\mathbf{x})d\theta_i = \mathrm{E}[\theta_i|\mathbf{x}],\end{aligned}$$

where f, g, and h are conditional densities (Croon, 2002, p. 204). Notice that in the above derivation, we used the fact that $f(t_i|\mathbf{x}) = \int g(t_i|\theta_i)h(\theta_i|\mathbf{x})d\theta_i$, and assumed that the order of integrations with respect to t_i and θ_i was exchangeable.

Secondly, the weight vectors in the LS and simplified A-R formula are proportional to the loading vector ($\mathbf{w}_i \propto \mathbf{a}_i$), while those in the other formulas are proportional to $\boldsymbol{\Delta}_i^{-1}\mathbf{a}_i$ ($\mathbf{w}_i \propto \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i$). We call the former the series (A) formulas, while the latter the series (B) formulas. Weight vectors within the same series are proportional to each other, and are expected to yield identical results after bias corrections. Thus, the distinction between the two series is more important than the distinction within the series. The series (B) formulas tend to be more often used with initial LV extraction methods that yield asymptotically efficient estimates in measurement models, such as the GLS and ML methods, while the opposite is true for the series (A) formulas. We later take this into account in choosing a LV score construction method in Croon's bias correction method. Note, however, that the use of the series (B) formulas used in conjunction with the efficient factor extraction methods usually does not provide efficient estimates of structural parameters even with proper bias correction methods. The following table classifies these formulas according to two criteria: The proportionality and the normalization conditions.

Table 1 Classification of the formulas for constructing LV scores.

Normalization Conditions	Proportionality (A) $\mathbf{w}_i \propto \mathbf{a}_i$	(B) $\mathbf{w}_i \propto \boldsymbol{\Delta}_i^{-1}\mathbf{a}_i$
I. Unit regression $(\mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i)^{-1}\mathbf{w}_i'\mathbf{a}_i = \phi_i^{-1} = 1$		Regression
II. Conditionally unbiased $\mathbf{w}_i'\mathbf{a}_i = 1$	Least Squares (LS)	Bartlett
III. Unit variance scores $\mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i = 1$	Simplified A-R	A-R

As alluded to earlier, more than one set of LV scores may have to be constructed simultaneously in Skrondal-Laake's method, when the path models include more than one predictor variable. Assume that there are two predictor

LVs, θ_i and θ_j. Let

$$\mathbf{t}_{ij} = \begin{pmatrix} t_i \\ t_j \end{pmatrix}, \text{ and } \mathbf{y}_{ij} = \begin{pmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{pmatrix},$$

and assume that we use the Regression formula. Then, $\mathbf{t}_{ij}$ is obtained by $\mathbf{t}_{ij} = \mathbf{W}'_{ij}\mathbf{y}_{ij}$, where

$$\mathbf{W}_{ij} = (\mathbf{A}_{ij}\boldsymbol{\Phi}_{ij}\mathbf{A}'_{ij} + \boldsymbol{\Delta}_{ij})^{-1}\mathbf{A}_{ij}\boldsymbol{\Phi}_{ij}, \tag{67}$$

where $\mathbf{A}_{ij}$, $\boldsymbol{\Phi}_{ij}$, and $\boldsymbol{\Delta}_{ij}$ are as defined in (5), (6), and (7), respectively.

5 Methods of Bias Correction

5.1 Croon's method

Any weight vectors obtained in the previous section may be used to construct LV scores, $t_i = \mathbf{w}'_i\mathbf{y}_i$ for $i = 1, \cdots, m$ in Croon's (2002) method. The covariance between t_i and t_j $(i \neq j)$ is related to the correlation between θ_i and θ_j by

$$\sigma_{t_it_j} = \mathbf{w}'_i\mathrm{E}[\mathbf{y}_i\mathbf{y}'_j]\mathbf{w}_j = \mathbf{w}'_i\boldsymbol{\Sigma}_{ij}\mathbf{w}_j = \mathbf{w}'_i\mathbf{a}_i\rho_{ij}\mathbf{a}'_j\mathbf{w}_j = Q_iQ_j\rho_{ij}, \tag{68}$$

or

$$\rho_{ij} = \mathbf{w}'_i\boldsymbol{\Sigma}_{ij}\mathbf{w}_j/Q_iQ_j = \sigma_{t_it_j}/Q_iQ_j, \tag{69}$$

where $Q_i = \mathbf{w}'_i\mathbf{a}_i$ and $Q_j = \mathbf{w}'_j\mathbf{a}_j$. As noted earlier, Q_i and Q_j are unity for conditionally unbiased formula for calculating LV scores (the LS and Bartlett formula), in which case, $\rho_{ij} = \sigma_{t_it_j}$. The variance of t_i, on the other hand, is given by

$$\sigma^2_{t_i} = \mathbf{w}'_i\mathrm{E}[\mathbf{y}_i\mathbf{y}'_i]\mathbf{w}_i = \mathbf{w}'_i\boldsymbol{\Sigma}_{ii}\mathbf{w}_i = (\mathbf{w}'_i\mathbf{a}_i)^2\sigma^2_i + \mathbf{w}'_i\boldsymbol{\Delta}_i\mathbf{w}_i, \tag{70}$$

or

$$\sigma^2_i = \frac{\mathbf{w}'_i\boldsymbol{\Sigma}_{ii}\mathbf{w}_i - \mathbf{w}'_i\boldsymbol{\Delta}_i\mathbf{w}_i}{Q^2_i} = \frac{\sigma^2_{t_i} - \mathbf{w}'_i\boldsymbol{\Delta}_i\mathbf{w}_i}{Q^2_i}. \tag{71}$$

The latter is equal to $\frac{Q^2_i}{Q^2_i} = 1$, as expected.

In practice, all relevant quantities have to be estimated from the data. As noted earlier, estimates of $\mathbf{a}_i$ and $\boldsymbol{\Delta}_i$ are obtained by applying initial LV extraction methods to $\mathbf{S}_{ii}$, from which estimates of the weight vectors $\hat{\mathbf{w}}_i$ $(i = 1, \cdots, m)$ are derived. Vectors of realizations of t_i are obtained by $\hat{\mathbf{t}}_i = \mathbf{Y}_i\hat{\mathbf{w}}_i$ $(i = 1, \cdots, m)$, from which covariances between $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{t}}_j$ $(i, j = 1, \cdots, m)$ are calculated. Estimates of covariances between true LVs, θ_i's, are then obtained using sample analogues of (69), namely

$$\hat{\sigma}_{ij} = \frac{\hat{\mathbf{w}}'_i\mathbf{S}_{ij}\hat{\mathbf{w}}_j}{\hat{Q}_i\hat{Q}_j}, \tag{72}$$

where $\hat{Q}_i = \hat{\mathbf{w}}_i'\hat{\mathbf{a}}_i$ and $\hat{Q}_j = \hat{\mathbf{w}}_j'\hat{\mathbf{a}}_j$. Note that here we have "covariances" because the sample analogue of the variances of the true LVs are usually not equal to unity, as indicated by

$$\hat{\sigma}_i^2 = \frac{\hat{\mathbf{w}}_i'\mathbf{S}_{ii}\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_i'\hat{\boldsymbol{\Delta}}_i\hat{\mathbf{w}}_i}{\hat{Q}_i^2}, \tag{73}$$

which is not equal to unity unless $\mathbf{S}_{ii} = \hat{\mathbf{a}}_i\hat{\mathbf{a}}_i' + \hat{\boldsymbol{\Delta}}_i$ holds exactly. Once the estimates of the variances and covariances are obtained, estimates of path coefficients (regression coefficients for predicting LVs from other sets of LVs) can be easily derived from them.

5.2 Skrondal and Laake's method

In the Skrondal and Laake's (2001) method, "unbiased" estimates of parameters in structural models are directly obtained by combining two particular formulas for constructing LV scores, the Regression formula for predictor LVs, and the Bartlett formula for criterion LVs. When there are more than a single structural model, each one is estimated separately. Each structural model usually has only one criterion variable. We first consider the case in which there is a single predictor variable, and then extend it to a multiple-predictor case. Suppose that LV i (θ_i) is the predictor variable, while θ_j is the criterion variable. That is, we estimate the regression coefficient $\beta_{j|i}$ for predicting θ_j from θ_i. We define t_i using the weight vector in the Regression formula, while t_j using the weight vector in the Bartlett formula. We then regress t_j onto t_i. Let $b_{t_j|t_i}$ denote the regression coefficient. Then,

$$b_{t_j|t_i} = (\mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i)^{-1}\mathbf{w}_i'\boldsymbol{\Sigma}_{ij}\mathbf{w}_j = (\mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i)^{-1}\mathbf{w}_i'\mathbf{a}_i\rho_{ij}\mathbf{a}_j'\mathbf{w}_j, \tag{74}$$

where ρ_{ij} is the correlation between θ_i and θ_j. Since $\mathbf{a}_j'\mathbf{w}_j = 1$ (the Bartlett estimate is conditionally unbiased), and

$$\begin{aligned}&(\mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i)^{-1}\mathbf{w}_i'\mathbf{a}_i \\ &\quad = (\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{ii}^{-1}\mathbf{a}_i)^{-1}\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}^{-1}\mathbf{a}_i = 1,\end{aligned} \tag{75}$$

where $\boldsymbol{\Sigma}_{ii} = \mathbf{a}_i\mathbf{a}_i' + \boldsymbol{\Delta}_i$, (74) reduces to

$$b_{t_j|t_i} = \rho_{ij} = \rho_{ij}/\sigma_i^2 = \beta_{j|i}, \tag{76}$$

where $\sigma_i^2 = \rho_{ii} = 1$, and $\beta_{j|i}$ is the regression coefficient for predicting θ_j from θ_i. The above relation indicates that $b_{t_j|t_i}$ is an "unbiased" estimate of $\beta_{j|i}$. Note also that any conditionally unbiased weight vector (e.g., the LS formula) may be used for $\mathbf{w}_j$ in (74) to obtain the relation in (76). Estimates of $\mathbf{a}_i$, $\boldsymbol{\Delta}_i$, and $\mathbf{w}_j$, and realized values of t_j are obtained similarly to the previous section.

When there are more than one predictor variable in a structural model, the above derivation has to be changed only slightly (Lu et al., 2011). Suppose,

for simplicity, there are two predictor variables, θ_i and θ_j, and one criterion variable, θ_k. The predictor LVs $\mathbf{t}_{ij}$ is calculated using the weight matrix defined in (67) and t_k using a weight vector in any conditionally unbiased formula. By regressing t_k on $\mathbf{t}_{ij}$, we obtain

$$\begin{aligned}\mathbf{b}_{t_k|t_it_j} &= (\mathbf{W}_{ij}'\boldsymbol{\Sigma}_{ij}\mathbf{W}_{ij})^{-1}\mathbf{W}_{ij}'\boldsymbol{\Sigma}_{ij,k}\mathbf{w}_k \\ &= (\mathbf{W}_{ij}'\boldsymbol{\Sigma}_{ij}\mathbf{W}_{ij})^{-1}\mathbf{W}_{ij}'\mathbf{A}_{ij}\boldsymbol{\Phi}_{ik,jk}\mathbf{a}_k'\mathbf{w}_k, \end{aligned} \tag{77}$$

where $\boldsymbol{\Sigma}_{ij} = \mathrm{E}[\mathbf{y}_{ij}\mathbf{y}_{ij}']$, $\boldsymbol{\Sigma}_{ij,k} = \mathrm{E}[\mathbf{y}_{ij}y_k]$, and $\boldsymbol{\Phi}_{ik,jk} = \begin{pmatrix}\rho_{ik}\\ \rho_{jk}\end{pmatrix}$. Since $\mathbf{a}_k'\mathbf{w}_k = 1$, and

$$\begin{aligned}&(\mathbf{W}_{ij}'\boldsymbol{\Sigma}_{ij}\mathbf{W}_{ij})^{-1}\mathbf{W}_{ij}'\mathbf{A}_{ij} \\ &= (\boldsymbol{\Phi}_{ij}\mathbf{A}_{ij}'\boldsymbol{\Sigma}_{(ij)}^{-1}\boldsymbol{\Sigma}_{(ij)}\boldsymbol{\Sigma}_{(ij)}^{-1}\mathbf{A}_{ij}\boldsymbol{\Phi}_{ij})^{-1}\boldsymbol{\Phi}_{ij}\mathbf{A}_{ij}'\boldsymbol{\Sigma}_{(ij)}^{-1}\mathbf{A}_{ij} \\ &= \boldsymbol{\Phi}_{ij}^{-1}, \end{aligned} \tag{78}$$

where $\boldsymbol{\Sigma}_{(ij)} = \mathbf{A}_{ij}\boldsymbol{\Phi}_{ij}\mathbf{A}_{ij}' + \boldsymbol{\Delta}_{ij}$, (77) reduces to

$$\mathbf{b}_{t_k|t_it_j} = \boldsymbol{\Phi}_{ij}^{-1}\boldsymbol{\Phi}_{ik,jk} = \boldsymbol{\beta}_{k|ij}. \tag{79}$$

An extension to more than two predictor variables is rather straightforward, following a similar line of derivation as above.

5.3 PLSc

The consistent Partial Least Squares (PLSc) method (Dijkstra and Schermelleh-Engel, 2014) is unique in that it obtains weight vectors first, and then loading vectors, while in all other methods discussed so far, loading vectors are obtained first, from which weight vectors are derived.

The weight vectors are derived as a convergence point of the following iterative procedure:

$$\hat{\mathbf{w}}_i \propto \sum_{j\in B_i} \mathrm{sign}_{ij}\mathbf{S}_{ij}\hat{\mathbf{w}}_j, \tag{80}$$

where $j \in B_i$ if the jth block of variables are connected to block i, and $\mathrm{sign}_{ij} = \pm 1$. (Note that following Dijkstra and Schermelleh-Engel (2014), we only consider the so-called Mode A algorithm in PLS, which these authors seem to favor.) Under (2) and (3), the probability limit of $\hat{\mathbf{w}}_i$ as the sample size n increases infinitely, denoted as $\mathbf{w}_i$, is proportional to $\mathbf{a}_i$, since

$$\mathbf{w}_i \propto \sum_{j\in B_i} \mathrm{sign}_{ij}\boldsymbol{\Sigma}_{ij}\mathbf{w}_j = (\sum_{j\in B_i} \mathrm{sign}_{ij}\rho_{ij}\mathbf{a}_j'\mathbf{w}_j)\mathbf{a}_i \propto \mathbf{a}_i. \tag{81}$$

(Note that the term enclosed in parentheses in the formula above is a scalar which is dependent only on i.) That is,

$$\mathbf{a}_i = c_i\mathbf{w}_i \tag{82}$$

for some c_i. We require, as a normalization convention, that

$$\sigma_{t_i}^2 = \mathrm{E}[(\mathbf{w}_i'\mathbf{y}_i)^2] = \mathbf{w}_i'\boldsymbol{\Sigma}_{ii}\mathbf{w}_i = 1, \tag{83}$$

from which it follows that

$$\mathbf{w}_i = (\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}\mathbf{a}_i)^{-1/2}\mathbf{a}_i \tag{84}$$

or

$$c_i = (\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}\mathbf{a}_i)^{1/2}. \tag{85}$$

The sample analogues of (82) through (85) are:

$$\hat{\mathbf{a}}_i = \hat{c}_i\hat{\mathbf{w}}_i, \tag{86}$$

$$\hat{\mathbf{w}}_i'\mathbf{S}_{ii}\hat{\mathbf{w}}_i = 1, \tag{87}$$

$$\hat{\mathbf{w}}_i = (\hat{\mathbf{a}}_i'\mathbf{S}_{ii}\hat{\mathbf{a}}_i)^{-1/2}\hat{\mathbf{a}}_i, \tag{88}$$

and

$$\hat{c}_i = (\hat{\mathbf{a}}_i'\mathbf{S}_{ii}\hat{\mathbf{a}}_i)^{1/2}. \tag{89}$$

We determine $\hat{c}_i$ in such way that its probability limit is equal to c_i. Dijkstra and Schermelleh-Engel (2014) suggest to use

$$\hat{c}_i = \left(\frac{\hat{\mathbf{w}}_i'(\mathbf{S}_{ii} - \mathrm{diag}(\mathbf{S}_{ii}))\hat{\mathbf{w}}_i}{\hat{\mathbf{w}}_i'(\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i' - \mathrm{diag}(\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'))\hat{\mathbf{w}}_i}\right)^{1/2}. \tag{90}$$

It may be seen that in the above formula, the matrix in the numerator, $\mathbf{S}_{ii} - \mathrm{diag}(\mathbf{S}_{ii})$, is $\mathbf{a}_i'\boldsymbol{\Sigma}_{ii}\mathbf{a}_i$ times as large as the matrix in the denominator, $\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i' - \mathrm{diag}(\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i')$, if $\mathbf{S}_{ii}$ and $\hat{\mathbf{w}}_i$ are replaced by their population counterparts, i.e., by $\boldsymbol{\Sigma}_{ii}$ and $\mathbf{w}_i$. Note that (84) and (88) indicate that the relationship between loadings and weights in PLSc is similar to the Simplified A-R formula for constructing LV scores. We say "similar", not "identical," because $\mathbf{S}_{ii}$ is used for $\boldsymbol{\Sigma}_{ii}$ (rather than $\hat{\boldsymbol{\Sigma}}_{ii} = \hat{\mathbf{a}}_i\hat{\mathbf{a}}_i' + \hat{\boldsymbol{\Delta}}_i$) in (88). See Dijkstra and Schermelleh-Engel (2014) for more detailed justifications of $\hat{c}_i$ above.

Let $\rho_{t_it_j}$ represent the correlation between $t_i = \mathbf{w}_i'\mathbf{y}_i$ and $t_j = \mathbf{w}_j'\mathbf{y}_j$. (We say "correlation" here, because the variance of LV scores is constrained to be unity in the A-R formula.) Then,

$$\rho_{t_it_j} = \mathbf{w}_i'\boldsymbol{\Sigma}_{ij}\mathbf{w}_j = Q_iQ_j\rho_{ij}, \tag{91}$$

or

$$\rho_{ij} = \mathbf{w}_i'\boldsymbol{\Sigma}_{ij}\mathbf{w}_j/Q_iQ_j = \rho_{t_it_j}/Q_iQ_j, \tag{92}$$

where $Q_i = \mathbf{w}_i'\mathbf{a}_i$ and $Q_j = \mathbf{w}_j'\mathbf{a}_j$ as defined previously. Note that (91) and (92) are identical in form to (68) and (69), respectively. The numerator of (92) can be consistently estimated by $\hat{\mathbf{w}}_i'\mathbf{S}_{ij}\hat{\mathbf{w}}_j$, and the denominator by $\hat{Q}_i\hat{Q}_j = \hat{\mathbf{w}}_i'\hat{\mathbf{a}}_i \cdot \hat{\mathbf{w}}_j'\hat{\mathbf{a}}_j$. The $\hat{\rho}_{ii}(= 1)$ and $\hat{\rho}_{ij}$ $(i, j = 1, \cdots, m)$ may further be used to estimate parameters in structural models that describe hypothesized relationships among θ_i's. Dijkstra and Schermelleh-Engel (2014) further

proposed to fit polynomial regression models among LVs using estimated ρ_{ij}'s and Q_i's.

Dijkstra (2013) proposed to apply the PLS iteration algorithm (80) only once to estimate $\mathbf{w}_i$ rather than iterating until convergence, and showed that the resultant estimators are still consistent. This procedure is non-iterative, although it is a two-step procedure and yields different estimates starting from different initial estimates of the weights. Since these estimates are intermediate results obtained during the fully iterated PLS procedure, it is easy to include them in our comparison. This method will be referred to as PLSc1 in the sequel. The initial weights used are equal weights across all variables as in the fully iterated PLSc.

As may be noticed earlier, the weight vector is adjusted by one-parameter to obtain the corresponding loading vector in both PLSc and PLSc1 (see (82) and (86)). This means that the choice of the value of this parameter is very important. There are a number of possible choices, and (90) is just one of them. Dijkstra (2016) proposed to choose $\hat{c}_i$ that minimizes

$$g_i = \sum_{a,b\neq a} \left[\frac{1}{2} \log \left(\frac{1+s_{ab}}{1-s_{ab}} \right) - \frac{1}{2} \log \left(\frac{1+c_i^2 \hat{w}_a \hat{w}_b}{1-c_i^2 \hat{w}_a \hat{w}_b} \right) \right]^2, \tag{93}$$

where a and b index variables in block i, and s_{ab} is an element in $\mathbf{S}_{ii}$. A minimization of the above criterion requires an iterative procedure. The consistent PLS method that uses the above adjustment will be denoted as PLSc*.

6 One-Step Estimation Methods to Be Compared

In this section, we discuss one-step estimation methods to be compared. These methods include full-ULS, full-GLS, and full-ML methods, and Bollen's two-stage least squares (2SLS) method. All of these methods directly estimate $\boldsymbol{\Phi}$ and/or path coefficients without calculating LV scores, and consequently no bias corrections are required. These methods are included in our study because it is of interest to find out how well those methods that require bias corrections (the multi-step procedures) work relative to those requiring no corrections.

6.1 Full-ULS, full-GLS, and full-ML methods

The full-ULS, full-GLS, and full-ML methods minimize

$$g = \mathrm{tr}(\mathbf{S} - \boldsymbol{\Sigma})^2, \tag{94}$$

$$r = \mathrm{tr}[(\mathbf{S} - \boldsymbol{\Sigma})\mathbf{S}^{-1}]^2, \tag{95}$$

and

$$h = \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - p, \tag{96}$$

respectively, where p is the total number of observed variables (i.e., $p = \sum_{i=1}^{m} p_i$). The minimizations of these criteria typically require elaborate iterative algorithms. For convenience, we use a MATLAB version of Cudeck, Klebe, and Henly's (1993) Gauss-Newton algorithm with numerical approximations to the Hessian matrix. Unfortunately, a MATLAB program implementing the same with an analytic Hessian matrix is unavailable. The full-GLS and full-ML methods are known to provide not only consistent but also asymptotically efficient estimates of parameters if the fitted model is correct. They also provide estimates of standard errors (SEs) as the square root of the diagonal elements of the inverse of the Hessian matrix. It is interesting to compare the estimates of SEs thus obtained against other methods of estimating SEs. It should be noted that none of the three methods discussed in this section are capable of *a priori* equating the variance of the criterion LVs, which renders equality constraints on path coefficients meaningless.

6.2 Bollen's 2SLS method

Bollen's (1993) 2SLS method directly estimates pass coefficients without estimating loadings, unique variances, or correlations among LVs. In this method, one (called the pivotal variable) of the observed variables in each block of variables is chosen as a proxy of the LV supposed to generate the block of observed variables. We consider each path model separately. In the ith path model, let $\mathbf{X}_i$ denote a collection of pivotal variables in the blocks of variables whose corresponding LVs are used as predictor variables. Let $\mathbf{v}_i$ denote the vector of pivotal variable in the block of variables corresponding to the criterion LV. We regress $\mathbf{v}_i$ onto $\mathbf{X}_i$. However, the use of ordinary least squares (OLS) method is not suitable in this context because the use of observed variables ($\mathbf{X}_i$) as predictors in place of true LVs incurs correlations between predictors and disturbance terms, which is against the basic OLS (ordinary least squares) assumptions. To avoid this difficulty, Bollen's 2SLS method employs instrumental variable (IV) estimation. Let $\mathbf{Z}_i$ denote the matrix of IVs, which typically consist of non-pivotal variables in the blocks whose LVs are used as predictor variables. The unconstrained 2SLS estimates of path coefficients in the ith path model are then given by

$$\hat{\boldsymbol{\beta}}_u^{(2SLS_i)} = (\mathbf{X}_i'\mathbf{P}_{Z_i}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{P}_{Z_i}\mathbf{v}_i = (\mathbf{S}_{Z_iX_i}'\mathbf{S}_{Z_iZ_i}^{-1}\mathbf{S}_{Z_iX_i})^{-1}\mathbf{S}_{Z_iX_i}'\mathbf{S}_{Z_iZ_i}^{-1}\mathbf{S}_{Z_iv_i}, \tag{97}$$

where $\mathbf{P}_{Z_i} = \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'$ is the orthogonal projector onto the column space of $\mathbf{Z}_i$. The 2SLS estimates of unconstrained path coefficients for other path models are similarly obtained. The constrained estimates of path coefficients will be given in Section 7.2 in the specific context of fitted constrained path models.

7 An Empirical Study

The foregoing discussions on methods of initial LV extraction and bias correction provide little information as to which combinations work best in parameter recovery. To investigate the problem, we conduct a Monte Carlo study, in which a simple SEM is assumed as a population model. Replicated data sets are generated from the assumed model, to which the methods discussed in the previous sections are applied, and estimates of parameters in the assumed model are obtained. Parameter recovery is assessed by how close the derived estimates are to the population parameters used to generate the data sets.

In the study, we also investigate how to obtain good estimates of standard errors (SEs) of parameter estimates from a single data set. This information is essential for assessing the reliability of estimated parameters, and for testing their significance (Devlieger et al., 2015). We use replicated samples of data (called the Monte Carlo samples) to obtain benchmark estimates of SEs, against which estimates of SEs obtained by the bootstrap method (Efron, 1979), the method based on OLS regression, and the method based on the inverse Hessian, are compared.

In what follows, we present the basic set-up of the Monte Carlo study: Assumed population parameters, the data generation procedure, exact implementations of the methods to be compared, definitions of the performance measure and other statistics of interest, and results of the Monte Carlo study.

7.1 Assumed population parameters

We assume model (1) and (2) as the basic measurement model, and model (3) as the basic structural model in our simulation study. This is a confirmatory factor analysis model in which only correlations (covariances) among the LVs are specified. We further assume that there are five LVs ($m = 5$), each generating five indicator variables ($p_i = 5$, $i = 1, \cdots, m$) according to (1) through (3). Assumed values of parameters are: $\mathbf{a}_i = .7 \times \mathbf{1}_5$ for all $i = 1, \cdots, 5$, where $\mathbf{1}_5$ is the five-component vector of ones, $\boldsymbol{\Delta}_i = .51 \times \mathbf{I}_5$ (this leads to a covariance matrix between observed variables, in which the variances due to systematic variations and those due to random variations (the unique variances) are about even), and

$$\boldsymbol{\Phi} = \left[\begin{array}{ccc|c|c} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} \\ \hline \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} \\ \hline \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} \end{array}\right] = \left[\begin{array}{ccc} \boldsymbol{\Phi}_{1:3} & \boldsymbol{\phi}_4 & \boldsymbol{\phi}_5 \\ \boldsymbol{\phi}_4' & \phi_{44} & \phi_{45} \\ \boldsymbol{\phi}_5' & \phi_{54} & \phi_{55} \end{array}\right] = \left[\begin{array}{ccccc} 1 & .2 & .2 & .6 & .6 \\ .2 & 1 & .2 & .6 & .6 \\ .2 & .2 & 1 & .6 & .6 \\ .6 & .6 & .6 & 1 & .6 \\ .6 & .6 & .6 & .6 & 1 \end{array}\right]. \quad (98)$$

(The correlations among the first three LVs are .2 and all other correlations are .6. Later (see Eqs. (99) and (100)) we postulate two path models in which LVs 4 and 5 serve as the criterion variables, and the first three LVs as common predictor variables. The correlations between predictor LVs are low, and

correlations between predictor and criterion LVs, and those between criterion LVs are somewhat higher, as may be expected. These values ensure positive-definiteness of the population covariance matrix between LVs as well as that between observed variables.) From these parameter vectors and matrices, the 25 by 25 population correlation (covariance) matrix among the observed variables is derived. One hundred data sets are then generated, for each of three sample sizes, $n = 200$ (small sample size), $n = 400$ (medium sample size), and $n = 800$ (large sample size), in such a way that each case (observation unit) in a sample follows a 25-variate normal distribution with 0 means and the prescribed correlation (covariance) matrix.

We actually tried a few other sets of parameter values (e.g., $\mathbf{a}_i = .8 \times \mathbf{1}_5$ and $\rho_{ij} = .7$) informally before we settled on the above specification, although the basic model adopted remained the same. We found, in all cases, results very similar to the one to be reported below. We thus consider that the result we report is relatively unaffected by the actual parameter values used in the study.

We also consider the second structural model, in which we postulate explicit path models among the LVs, namely

$$\theta_4 = \theta_1\beta_1 + \theta_2\beta_2 + \theta_3\beta_3 + e_1, \tag{99}$$

and

$$\theta_5 = \theta_1\beta_4 + \theta_2\beta_5 + \theta_3\beta_6 + e_2. \tag{100}$$

It is convenient to write these path models in one equation, namely

$$\begin{pmatrix} \theta_4 \\ \theta_5 \end{pmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \boldsymbol{\beta}_u + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \tag{101}$$

where

$$\boldsymbol{\beta}_u = \big(\, \beta_1,\, \beta_2,\, \beta_3,\, \beta_4,\, \beta_5,\, \beta_6 \,\big)'. \tag{102}$$

The population path coefficients are then given by

$$\boldsymbol{\beta}_u = \boldsymbol{\Phi}_{(1:3)}^{-1} \begin{pmatrix} \boldsymbol{\phi}_4 \\ \boldsymbol{\phi}_5 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_{1:3}^{-1}\boldsymbol{\phi}_4 \\ \boldsymbol{\Phi}_{1:3}^{-1}\boldsymbol{\phi}_5 \end{pmatrix}, \tag{103}$$

where

$$\boldsymbol{\Phi}_{(1:3)} = \begin{bmatrix} \boldsymbol{\Phi}_{1:3} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Phi}_{1:3} \end{bmatrix}. \tag{104}$$

The population path coefficients thus calculated happen to be all equal to .4286.

We also consider a constrained path model in which we assume $\beta_3 = \beta_4 \equiv \beta_{c1}$, $\beta_1 = \beta_5 \equiv \beta_{c2}$, and $\beta_2 = \beta_6 \equiv \beta_{c3}$ in (101). The constrained population path coefficients are derived by

$$\boldsymbol{\beta}_c = \begin{pmatrix} \beta_{c1} \\ \beta_{c2} \\ \beta_{c3} \end{pmatrix} = (\mathbf{L}'\boldsymbol{\Phi}_{(1:3)}\mathbf{L})^{-1}\mathbf{L}' \begin{pmatrix} \boldsymbol{\phi}_4 \\ \boldsymbol{\phi}_5 \end{pmatrix}, \tag{105}$$

where $\boldsymbol{\Phi}_{(1:3)}$ is as defined in (104), and

$$\mathbf{L} = \begin{bmatrix} 0\,0\,1\,1\,0\,0 \\ 1\,0\,0\,0\,1\,0 \\ 0\,1\,0\,0\,0\,1 \end{bmatrix}'. \tag{106}$$

The constrained path coefficients thus calculated happen to be all identical to the unconstrained path coefficients obtained above.

7.2 Implementations of the estimation methods to be compared

The generated data are analyzed by sixteen different methods, designated as (1) Bentler(ULS)/Croon, (2) Bentler(GLS)/Croon, (3) Bentler(ULS)/SL, (4) Bentler(GLS)/SL, (5) Hägglund's FABIN2/Croon, (6) FABIN3/Croon, (7) Ihara-Kano/Croon, (8) PLSc, (9) PLSc1, (10) PLSc*, (11) blockwise-ULS, (12) blockwise-ML, (13) Bollen's 2SLS, (14) full-ULS, (15) full-GLS, and (16) full-ML. The first twelve methods pertain to multi-step methods, requiring calculations of LV scores and possible bias corrections. The general constituents of these methods have already been described in some detail in Sections 3, 4, and 5, so that we discuss here only specific ones. The remaining four methods ((13) through (16)) requiring no bias corrections, on the other hand, are described in Section 6. These one-step methods are included in the comparison as benchmark methods, against which the performance of the multi-step methods is calibrated.

The first four methods use Bentler's method discussed in Section 3.1 for initial LV extraction, the first two of which apply Bentler's method to each block of indicators to extract one LV at a time. In each case, the first variable in each block is regarded as the pivotal variable, while the remaining ones are taken as non-pivotal variables. The two methods are distinguished by the criterion used to obtain estimates of parameters for non-pivotal variables, ULS ($\mathbf{W} = \mathbf{I}_p$ in (17)) in the first method and GLS ($\mathbf{W} = \mathbf{S}_{ii}^{-1}$ in (17)) in the second method. Bentler(ULS)/Croon then applies the LS formula, while Bentler(GLS)/Croon the Bartlett formula, to construct LV scores. The third and fourth methods pertaining to the Skrondal-Laake's corrections fit unconstrained and constrained path models directly, so that no estimates other than those of path coefficients are available. These methods apply Bentler's method to extract three predictor LVs simultaneously corresponding to the first three blocks of observed variables, followed by the Bartlett formula for score construction. Criterion LVs, on the other hand, are extracted in the same way as in the first two methods. The third and fourth methods are again distinguished by the criterion used to obtain estimates of parameters in non-pivotal variables.

The next two methods use Hägglund's IV estimation method described in Section 3.2 to extract initial LVs. These methods are applied to each block of observed variables to extract one LV each. FABIN2 uses (41) to estimate loading vectors followed by the LS formula for LV score construction, while

FABIN3 uses (42) to estimate loadings followed by the Bartlett method for score construction. As in Bentler's method, the first variable in each block is taken as the pivotal variable.

Ihara-Kano's method is applied to each block of observed variables to extract one LV at a time. The LS formula is then used for constructing LV scores. This method is fully described in Section 3.3, and no further implementation details are needed.

The next three methods are the consistent PLS methods described in Section 5.3. PLSc iterates on (80) until convergence, while PLSc1 only once, both starting from homogeneous weights. Both PLSc and PLSc1 use (90) for $\hat{c}_i$, while PLSc* the $\hat{c}_i$ that minimizes (93).

The next two methods, blockwise-ULS and blockwise-ML, are described in Section 3.4. These methods are applied to each block of indicators to extract one LV at a time so as to minimize (59) and (60). As noted earlier, the minimizations require iterative algorithms, and initial estimates of unique variances. We use 1 - smc (the squared multiple correlations) as initial estimates of unique variances. Once parameters in the measurement models are estimated, the blockwise-ULS uses the LS formula to derive LV scores, while the blockwise-ML uses the Bartlett formula.

The full-ULS, full-GLS, and full-ML methods have been described in Section 6.1. The only thing we should note here is that they typically require good initial estimates of unique variances to start the iterations. While we could have used Ihara-Kano's method for this purpose, we simply used population unique variances that are available in the present context.

Bollen's (1993) 2SLS method has been described in Section 6.2. To obtain constrained estimates of path coefficients, the unconstrained estimates given in (97) for the ith path model should be written in a single equation for all path models bound by the constraints. For the constrained path models given toward the end of the previous subsection, let $\mathbf{K}$ represent a block diagonal matrix of $\mathbf{X}_i'\mathbf{P}_{Z_i}\mathbf{X}_i$, and let $\mathbf{v}^*$ represent a super vector of $\mathbf{X}_i'\mathbf{P}_{Z_i}\mathbf{v}_i$. Then, the unconstrained 2SLS estimates of path coefficients can be written collectively as

$$\hat{\beta}_u^{(2SLS)} = \mathbf{K}^{-1}\mathbf{v}^*. \tag{107}$$

The constrained 2SLS estimates of path coefficients, on the other hand, are given by

$$\hat{\beta}_c^{(2SLS)} = (\mathbf{L}'\mathbf{K}\mathbf{L})^{-1}\mathbf{L}'\mathbf{v}^*, \tag{108}$$

where $\mathbf{L}$ is defined in (106).

7.3 Performance and other measures of interest

One of our main interests in the present study lies in comparing the performance of various methods of estimation in parameter recovery. The recovery is measured by root mean square errors (rmse's). There are five groups of parameters in assumed SEMs: loadings ($\mathbf{a}_i$, $i = 1, \cdots, m$), unique variances

(diagonal elements of $\boldsymbol{\Delta}_i$, $i = 1, \cdots, m$), covariances among LVs ($\boldsymbol{\Phi}$), unconstrained path coefficients ($\boldsymbol{\beta}_u$), and constrained path coefficients ($\boldsymbol{\beta}_c$). The rmse is defined for each group of parameters separately. Let $\mathbf{a}$ denote the super vector consisting of $\mathbf{a}_i$'s placed on top of each other. Then, the rmse of loadings is defined by

$$\text{rmse}_a = (\sum_{q=1}^{n_b} [\hat{\mathbf{a}}^{(q)} - \mathbf{a}^{(pop)})'(\hat{\mathbf{a}}^{(q)} - \mathbf{a}^{(pop)})/25n_b]^{1/2}, \tag{109}$$

where q indexes a replicated sample (called a Monte Carlo sample in this paper), n_b indicates the total number of replications, and $\mathbf{a}^{(pop)}$ is the population loading vector whose elements are arranged in the same way as $\mathbf{a}$. Squared differences between estimates and parameters are summed over all loadings and replications, and divided by the number of loadings times the number of parameters, before a square-root is taken. The rmse's of other groups of parameters are defined similarly, except for $\boldsymbol{\Phi}$, which is a symmetric matrix. To avoid redundancy in the symmetric matrix, only non-redundant portions are taken into account.

The rmse defined above is a function of both bias and average standard error (SE). Here, the word bias is used in the conventional sense of the word in statistics, namely the difference between true parameters and expected values of their estimates. To assess the degree of bias and to obtain a purer measure of variability in the estimates of parameters, we also calculate average SEs, defined by

$$\text{SE}_a = [\sum_{q=1}^{n_b} (\hat{\mathbf{a}}^{(q)} - \bar{\mathbf{a}})'(\hat{\mathbf{a}}^{(q)} - \bar{\mathbf{a}})/25n_b]^{1/2} \tag{110}$$

for loadings, where $\bar{\mathbf{a}}$ is the mean estimate of loadings over replications. This is similar to (109) except that $\mathbf{a}^{(pop)}$ in (109) is replaced by $\bar{\mathbf{a}}$. Average SEs for the other groups of parameters are similarly defined. When the SE's are not much different from the corresponding rmse's, the biases in estimates are considered small.

The average SEs can be calculated over replicated samples under the present circumstances because the replicated samples of data are generated from a population model. However, in practical data analytic situations, there are no such replicated samples, and estimates of SEs must be obtained from a single set of observed data. There are several techniques used to obtain estimates of SEs in such situations, among which the bootstrap method (Efron, 1979) is perhaps most widely applicable. It can be used as far as cases (sampling units) are statistically independent. In this method, many so-called bootstrap samples are generated by repeatedly resampling cases from an original data set. Each bootstrap sample is analyzed by the methods of interest. Different parameter estimates are produced from different bootstrap samples, but their variabilities provide estimates of SEs. The bootstrap estimates of SEs can be obtained for any groups of parameters in SEMs. We take the first

Monte Carlo data set as the original data set from which bootstrap samples are generated.

Another possible estimate of SEs is based on the ordinary least squares (OLS) regression, in which all assumptions underlying the OLS estimation are supposed to hold. Those assumptions include: Predictor variables are fixed (not random) and measured without errors, and unpredictable portions of observations on the criterion variable vary independently from each other with constant variance σ^2. Then, the OLS estimates of SEs for the unconstrained path coefficients are given by

$$\hat{\boldsymbol{\sigma}}^2_{\hat{\beta}_u} = \hat{\sigma}^2_u \mathrm{diag}(\hat{\boldsymbol{\Phi}}^{-1}_{1:3}), \tag{111}$$

where $\hat{\boldsymbol{\Phi}}_{1:3}$ is an estimate of $\boldsymbol{\Phi}_{1:3}$ introduced in (98), and

$$\hat{\sigma}^2_u = \hat{\phi}_{44} + \hat{\phi}_{55} - \hat{\boldsymbol{\phi}}'_{4,5}\hat{\boldsymbol{\beta}}_u, \tag{112}$$

while those for the constrained path coefficients by

$$\hat{\boldsymbol{\sigma}}^2_{\hat{\beta}_c} = \hat{\sigma}^2_c \mathrm{diag}((\mathbf{L}'\hat{\boldsymbol{\Phi}}_{1:3}\mathbf{L})^{-1})\mathbf{1}, \tag{113}$$

where $\mathbf{1}$ is a vector of ones, and

$$\hat{\sigma}^2_c = \hat{\phi}_{44} + \hat{\phi}_{55} - \hat{\boldsymbol{\phi}}'_{4,5}\hat{\boldsymbol{\beta}}_c. \tag{114}$$

The $\hat{\boldsymbol{\sigma}}^2_{\hat{\beta}_u}$ and $\hat{\boldsymbol{\sigma}}^2_{\hat{\beta}_c}$ are averaged over the regression coefficients to obtain mean SEs. As is clear from the above, the OLS estimates of SEs can only be obtained when specific path models are postulated as the structural model.

As noted earlier, in the asymptotically efficient estimators, such as the full-GLS and full-ML methods, square roots of the diagonal elements of the inverse of the Hessian matrix at the convergence point provide estimates of SEs. We are interested in comparing SEs obtained by these methods against each other and against those obtained from replicated samples of data.

Parameters in SEMs are not entirely free to vary, but restricted. The loadings should all be between -1 and 1 if the data are standardized. The unique variances should be non-negative. A covariance (correlation) matrix among LVs must be positive definite, that is, their eigenvalues must all be positive. Solutions that violate these restrictions are called improper solutions. As far as we know, no methods for estimating SEMs are equipped with proper devices for preventing improper solutions. In our simulation study, frequencies of improper solutions are counted and compared across different methods.

7.4 The results

The results of the Monte Carlo study are summarized in Tables 2 and 3. Table 2 reports rmse's as functions of sample sizes, ($n = 200, 400$, and 800), groups of parameters (loadings, unique variances, covariances among LVs, and unconstrained and constrained path coefficients), and sixteen methods of parameter

estimation. The smaller the rmse, the better is the parameter recovery. Table 2 also reports frequencies of improper solutions out of 100 Monte Carlo samples in the last column. These frequencies represent the number of data sets that produced improper solutions of some kind, although a vast majority of improper solutions were due to non-positive definite estimates of the covariance matrix among LVs.

As expected, rmse decreases as the sample size increases. It is observed that there are only small differences in rmse between Bentler(ULS) and Bentler(GLS). Between Croon's bias correction method and Skrondal-Laake's (SL) method, the former seems to have a clear edge over the latter. There are also only small differences between FABIN2 and FABIN3. Both methods work remarkably well in terms of both parameter recovery and frequencies of improper solutions. Ihara-Kano's method, on the other hand, does not seem to work as well in parameter recovery or in the frequency of improper solutions. In fact, it produces excessive numbers of improper solutions across all sample sizes. It may also be observed that there are only small differences among three PLS methods. This indicates that iterations beyond the first are indeed unnecessary (although this may be due to the fact that the population weights are homogeneous), and that a more "elaborate" choice of $\hat{c}^i$ in PLSc* is not of much help in improving parameter recovery, while increasing the number of improper solutions. All three versions of PLSc tend to produce larger rmse's for parameters in measurement models (loadings and unique variances) than the other methods examined in this study, while producing approximately equal rmse's for parameters in structural models (correlations between LVs and path coefficients), which is probably more important in SEMs. Both blockwise-ULS and blockwise-ML with Croon's correction work very well. They work as well as the full versions of these methods (full-ULS and full-ML) in most cases, and even better in some cases. Somehow Bollen's 2SLS method performs rather poorly across all sample sizes. The overall winner in parameter recovery seems to be Hägglundd's methods (FABIN2 and FABIN3). They work as well as blockwise-ML and full-ML in both rmse's and frequencies of improper solutions with much less computation time.

One important observation to be made is that what is evaluated in the simulation study is the combinations (=interactions) of the initial LV extraction methods, the methods of LV scores construction methods, and the bias correction methods. From the simulation study, it seems clear now that the initial LV extraction methods are the most important factor. The effects of the LV score construction methods are relatively minor, provided that proper bias corrections are applied.

Table 3 reports four kinds of estimates of SEs (Monte Carlo, bootstrap, OLS, and inverse Hessian) of estimated parameters obtained by four selected methods of parameter estimation (FABIN3, PLSc1, blockwise-ML, and full-ML) for varying sample sizes ($n = 200, 400$, and 800). It should be noted first that, as expected, estimates of SEs invariably get smaller as the sample size increases. We next look at the rows labeled "Monte Carlo" more closely. These are the SEs obtained from replicated data sets generated according to

Table 2 Parameter recovery: Root mean square error (rmse) per parameter obtained by the various methods

Sample Size	Methods	L.V. Load.	Unique Vari.	L.V. Corr.	Path Coeff. Unconst.	Path Coeff. Const.	Improper Solu.
	Bentler(ULS)/Croon	558	772	662	591	425	11
	Bentler(GLS)/Croon	570	642	674	595	429	5
	Bentler(ULS)/SL				617	443	
	Bentler(GLS)/SL				614	455	
	FABIN2/Croon	456	644	661	592	426	12
	FABIN3/Croon	463	644	663	596	429	7
	Ihara-Kano/Croon	458	780	675	607	439	39
200	PLSc	799	1092	656	590	423	11
	PLSc1	774	1061	656	589	423	10
	PLSc*	772	1041	662	596	420	27
	blockwise-ULS/Croon	455	631	661	592	426	12
	blockwise-ML/Croon	458	638	662	596	428	9
	Bollen(2SLS)				1008	710	
	full-ULS	512	711	662	682		18
	full-GLS	533	1234	764	700		11
	full-ML	432	601	666	632		19
	Bentler(ULS)/Croon	401	555	486	399	301	6
	Bentler(GLS)/Croon	408	445	494	399	301	3
	Bentler(ULS)/SL				413	319	
	Bentler(GLS)/SL				416	323	
	FABIN2/Croon	320	446	486	399	300	7
	FABIN3/Croon	324	450	487	398	300	4
	Ihara-Kano/Croon	326	586	493	406	307	24
400	PLSc	560	770	481	400	301	8
	PLSc1	541	744	481	400	301	9
	PLSc*	541	739	481	401	302	14
	blockwise-ULS/Croon	320	445	486	399	300	11
	blockwise-ML/Croon	321	446	487	398	300	9
	Bollen(2SLS)				714	522	
	full-ULS	361	500	485	530		9
	full-GLS	340	694	518	491		10
	full-ML	305	423	486	446		12
	Bentler(ULS)/Croon	278	389	336	294	214	1
	Bentler(GLS)/Croon	279	322	338	294	214	1
	Bentler(ULS)/SL				293	217	
	Bentler(GLS)/SL				295	220	
	FABIN2/Croon	230	321	337	294	215	1
	FABIN3/Croon	230	322	337	297	216	1
	Ihara-Kano/Croon	234	408	342	406	307	13
800	PLSc	395	549	337	294	214	1
	PLSc1	380	529	337	294	214	1
	PLSc*	380	527	338	295	214	1
	blockwise-ULS/Croon	230	321	337	294	215	1
	blockwise-ML/Croon	230	322	337	293	214	1
	Bollen(2SLS)				523	360	
	full-ULS	258	360	337	451		0
	full-GLS	229	415	356	338		1
	full-ML	218	305	338	352		0

Tabled numbers should be multiplied by 10^{-4}.

the assumed population model. Observe that the values of the Monte Carlo SEs are all similar to the corresponding values of rmse's in the previous table. This means that in all cases there is little bias in the estimates of parameters. The Monte Carlo SEs also serve as benchmark SEs, against which other estimates of SEs obtained from single samples are compared.

Following the Monte Carlo SEs, the bootstrap estimates of SEs are reported in the table. It may be observed that they are all fairly close to the Monte Carlo SEs. This indicates that the bootstrap SEs are reasonably good approximations to the Monte Carlo SEs, so that they may be safely used in significance testing of estimated parameters. An exception is the bootstrap SEe for the loadings by PLSc1, which are appreciably larger than the corresponding Monte Carlo SEs. This indicates that the tests of loading parameters in PLSc1 tend to be more conservative than they are supposed to.

The OLS estimates of SEs are available only for path coefficients. It should be pointed out that they tend to be smaller than the corresponding Monte Carlo SEs across all methods and sample sizes. This implies that the tests of significance of path coefficients based on the OLS estimates of SEs tend to be too liberal, leading to too many rejections of the null hypotheses stating that the corresponding population path coefficients are zero. Perhaps, it is too optimistic to expect that the assumptions underlying OLS hold for LVs in SEMs.

The estimates of SEs obtained by the inverse Hessian are available only for full-ML. They consistently overestimate the Monte Carlo SEs across all parameter types and sample sizes. Consequently, the significance tests of parameters based on these estimates tend to be more conservative than they should. It may be that numerical approximations to the Hessian matrix is accurate enough for optimization, but not for obtaining the SE estimates.

8 Discussion

In this paper, we compared several estimators of parameters in structural equation models in terms of their parameter recovery capability (as measured by rmse's) and efficiency (as measured by SEs), which are almost synonymous to each other in near absence of bias, as in the preset situation. The estimators included four versions of Bentler's non-iterative confirmatory factor analysis, two versions of Hägglund's IV estimation method, Ihara-Kano's non-iterative uniqueness estimation method, three versions of PLSc (Consistent PLS), and two blockwise estimation methods. These methods were used for initial LV extractions to estimate parameters in measurement models (factor loadings and unique variances), followed by calculations of LV scores and bias corrections to estimate parameters in structural models (covariances among LVs and path coefficients). The performance of these methods are also compared with one-step estimation methods, such as the full-ULS, -GLS, and -ML methods, and Bollen's 2SLS method, that estimate all relevant parameters simultaneously. It may be concluded that Hägglund's (1982) FABIN3 combined with Croon's

Table 3 Estimates of standard errors for selected methods

Sample Size	Methods	Kinds of SE	L.V. Load.	Unique Vari.	L.V. Corr.	Path Coeff. Unconst.	Path Coeff. Const.
200	FABIN3	Monte Carlo	458	636	653	592	426
		Bootstrap	468	644	678	601	432
		OLS				506	357
	PLSc1	Monte Carlo	746	1027	693	586	420
		Bootstrap	893	1010	693	582	420
		OLS				487	419
	blockwise-ML	Monte Carlo	455	633	653	593	426
		Bootstrap	461	634	678	601	435
		OLS				506	357
	full-ML	Monte Carlo	429	597	656	604	
		Inv. Hess.	928	851	946	1078	
400	FABIN3	Monte Carlo	318	443	479	395	296
		Bootstrap	327	453	473	418	302
		OLS				355	250
	PLSc1	Monte Carlo	522	799	473	396	296
		Bootstrap	630	720	464	412	297
		OLS				352	248
	blockwise-ML	Monte Carlo	317	442	479	394	296
		Bootstrap	325	450	476	421	307
		OLS				355	250
	full-ML	Monte Carlo	301	419	479	401	
		Inv. Hess.	657	605	673	759	
800	FABIN3	Monte Carlo	229	320	330	290	211
		Bootstrap	228	317	335	294	215
		OLS				250	176
	PLSc1	Monte Carlo	367	571	329	291	210
		Bootstrap	442	509	331	291	213
		OLS				249	176
	blockwise-ML	Monte Carlo	229	320	330	290	212
		Bootstrap	227	316	336	295	216
		OLS				249	176
	full-ML	Monte Carlo	217	302	331	295	
		Inv. Hess.	464	427	475	536	

Tabled numbers should be multiplied by 10^{-4}.

bias correction methods worked best among all the methods compared in this paper. It is non-iterative, easy to implement, and resistant to improper solutions.

We also examined methods for obtaining estimates of SEs of parameter estimates from a single data set, the bootstrap method, OLS estimates of SEs for path coefficients, and SEs based on the inverse Hessian in full-ML, against those obtained from replicated samples (Monte Carlo estimates). It was found that the bootstrap method could obtain estimates of SEs close to the Monte Carlo estimates of SEs, while the OLS estimates tended to be systematically smaller, and the inverse Hessian estimates systematically larger, than the Monte Carlo estimates.

It should be pointed out, however, that our empirical study is somewhat limited in scope. In particular, only very simple SEMs were assumed and no

structural errors were incorporated in the data generation process. The performance of the methods we examined may well be affected by these factors. More comprehensive studies are in order so as to draw more general conclusions on the performance of the methods. A study is currently underway that specifically addresses potential impacts of structural errors such as cross loadings (i.e., existence of indicators that significantly load on more than one LV) on the quality of parameter estimates obtained by the methods we investigated in the present paper.

Finally, it should be noted that MATLAB routines implementing the estimation methods discussed in this paper are available from Online Resource.

References

1. Bentler, P. M., Confirmatory factor analysis via noniterative estimation: A fast, inexpensive method, Journal of Marketing Research, 19, 417–424 (1982)
2. Bentler, P. M., Lee, S. Y., Some extensions of matrix calculus, General Systems, 20, 145–150, (1975)
3. Bollen, K. A., An alternative two stage least squares (2SLS) estimator for latent variable equations, Psychometrika, 61, 109–121 (1996)
4. Croon, M., Using predicted latent scores in general latent structure models. In G. A. Marcoulides, I. Moustaki (eds) Latent variable and latent structure models, 195–223, Lawrence Earlbaum Associates, Mahwah, NJ (2002)
5. Devlieger, I., Mayer, A., Rosseel, Y., Hypothesis testing using factor score regression: A comparison of four methods, Educational and Psychological Measurement, 75, 1–30 (2015)
6. Cudeck, R., Noniterative factor analysis estimators, with algorithms for subset and instrumental variable selection, Journal of Educational Statistics, 16, 35–52 (1991)
7. Cudeck, R., Klebe, K.J., Henly, S.J., A simple Gauss-Newton procedure for covariance structure analysis with high-level computer languages, Psychometrika, 58, 211–232 (1993)
8. Dijkstra, T. K., The simplest possible factor model estimator, Unpublished note, University of Groningen (2013)
9. Dijkstra, T. K., A note on how to make partial least squares consistent, Unpublished note, University of Groningn (2016)
10. Dijkstra, T. K., Schermelleh-Engel, K., Consistent partial least squares for nonlinear structural equation models, Psychometrika, 79, 585–604 (2014)
11. Efron, B., Bootstrap methods: another look at the Jackknife, Annals of Statistics, 7, 1–26 (1979)
12. Hägglund, G., Factor analysis by instrumental variables methods, Psychometrika, 47, 209–222 (1982)
13. Ihara, M., Kano, Y., A new estimator of the uniqueness in factor analysis, Psychometrika, 51, 563–566 (1986)
14. Jöreskog, K. G., Factor analysis by least-squares and maximum likelihood methods, in K. Enslein, A. Ralston, H. S. Wilf (eds.) Statistical Methods for Digital Computers, 125–153, Wiley-Interscience, New york, (1977)
15. Jung, S., Takane, Y., Regularized common factor analysis, In K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (Eds.), New Trends in Psychometrics, 141–149, University Academic Press, Tokyo (2008)
16. Lu, I. R. R., Kwan, E., Thomas, D. R., Cedzynski, M., Two new methods for estimating structural equation models: an illustration and a comparison with two established methods, International Journal of Research in Marketing, 28, 258–268 (2011).
17. McDonald, R. P., Burr, E. J., A comparison of four methods of constructing factor scores, Psychometrika, 32, 381–401 (1967)
18. Skrondal, A., Laake, P., Regression among factor scores, Psychometrika, 66, 563–576 (2001)
19. Skrondal, A., Rabe-Hesketh, S., Generalized Latent Variables Modeling, Chapman-Hall/CRC Press, Boca Raton (2004)